A Mathematical Introduction to Signals and Systems

Volume III. Advanced analysis

Andrew D. Lewis

This version: 2022/03/07

Preface for series

The subject of signals and systems, particularly linear systems, is by now an entrenched part of the curriculum in many engineering disciplines, particularly electrical engineering. Furthermore, the offshoots of signals and systems theory—e.g., control theory, signal processing, and communications theory—are themselves well-developed and equally basic to many engineering disciplines. As many a student will agree, the subject of signals and systems is one with a reliance on tools from many areas of mathematics. However, much of this mathematics is not revealed to undergraduates, and necessarily so. Indeed, a complete accounting of what is involved in signals and systems theory would take one, at times quite deeply, into the fields of linear algebra (and to a lesser extent, algebra in general), real and complex analysis, measure and probability theory, and functional analysis. Indeed, in signals and systems theory, many of these topics are woven together in surprising and often spectacular ways. The existing texts on signals and systems theory, and there is a true abundance of them, all share the virtue of presenting the material in such a way that it is comprehensible with the bare minimum background.

Should I bother reading these volumes?

This virtue comes at a cost, as it must, and the reader must decide whether this cost is worth paying. Let us consider a concrete example of this, so that the reader can get an idea of the sorts of matters the volumes in this text are intended to wrestle with. Consider the function of time

$$f(t) = \begin{cases} e^{-t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

In the text (Example IV-6.1.3–2) we shall show that, were one to represent this function in the frequency domain with frequency represented by ν, we would get

$$\hat{f}(\nu) = \int_{\mathbb{R}} f(t) e^{-2i\pi\nu t}\, dt = \frac{1}{1 + 2i\pi\nu}.$$

The idea, as discussed in Chapter IV-2, is that $\hat{f}(\nu)$ gives a representation of the "amount" of the signal present at the frequency ν. Now, it is desirable to be able to reconstruct f from $\hat{f}$, and we shall see in Section IV-6.2 that this is done via the formula

$$f(t) \text{“=”} \int_{\mathbb{R}} \hat{f}(\nu) e^{2i\pi\nu t}\, d\nu. \tag{FT}$$

The easiest way to do the integral is, of course, using a symbolic manipulation program. I just tried this with MATHEMATICA®, and I was told it could not do the computation. Indeed, the integral *does not converge*! Nonetheless, in many tables of

Fourier transforms (that is what the preceding computations are about), we are told that the integral in (FT) does indeed produce $f(t)$. Are the tables wrong? Well, no. But they are only correct when one understands exactly what the right-hand side of (FT) means. What it means is that the integral converges, *in* $\mathsf{L}^2(\mathbb{R};\mathbb{C})$ to f. Let us say some things about the story behind this that are of a general nature, and apply to many ideas in signal and system theory, and indeed to applied mathematics as a whole.

1. The story—it is the story of the L^2-Fourier transform—is not completely trivial. It requires *some* delving into functional analysis at least, and some background in integration theory, if one wishes to understand that "L" stands for "Lebesgue," as in "Lebesgue integration." At its most simple-minded level, the theory is certainly understandable by many undergraduates. Also, at its most simple-minded level, it raises more questions than it answers.
2. The story, even at the most simple-minded level alluded to above, takes some time to deliver. The full story takes *a lot* of time to deliver.
3. It is not necessary to fully understand the story, perhaps even the most simple-minded version of it, to be a user of the technology that results.
4. By understanding the story well, one is led to new ideas, otherwise completely hidden, that are practically useful. In control theory, quadratic regulator theory, and in signal processing, the Kalman filter, are examples of this.
5. The full story of the L^2-Fourier transform, and the issues stemming from it, directly or otherwise, is beautiful.

The nature of the points above, as they relate to this series, are as follows. Points 1 and 2 indicate why the story cannot be told to all undergraduates, or even most graduate students. Point 3 indicates why it is okay that the story not be told to everyone. Point 4 indicates why it is important that the story be told to someone. Point 5 should be thought of as a sort of benchmark as to whether the reader should bother with understanding what is in this series. Here is how to apply it. If one reads the assertion that this is a beautiful story, and their reaction is, "Okay, but there better be a payoff," or, "So what?" or, "Beautiful to who?" then perhaps they should steer clear of this series. If they read the assertion that this is a beautiful story, and respond with, "Really? Tell me more," then I hope they enjoy these books. They were written for such readers. Of course, most readers' reactions will fall somewhere in between the above extremes. Such readers will have to sort out for themselves whether the volumes in this series lie on the right side, for them, of being worth reading. For these readers I will say that this series is *heavily* biased towards readers who react in an unreservedly positive manner to the assertions of intrinsic beauty.

For readers skeptical of assertions of the usefulness of mathematics, an interesting pair of articles concerning this is [Wigner 1960] and [Hamming 1980].

What is the best way of getting through this material?

Now that a reader has decided to go through with understanding what is in these volumes, they are confronted with actually doing so: a possibly nontrivial matter, depending on their starting point. Let us break down our advice according to the background of the reader.

I look at the tables of contents, and very little seems familiar. Clearly if nothing seems familiar at all, then a reader should not bother reading on until they have acquired an at least passing familiarity with some of the topics in the book. This can be done by obtaining an undergraduate degree in electrical engineering (or similar), or pure or applied mathematics.

If a reader already possess an undergraduate degree in mathematics or engineering, then certainly some of the following topics will appear to be familiar: linear algebra, differential equations, some transform analysis, Fourier series, system theory, real and/or complex analysis. However, it is possible that they have not been taught in a manner that is sufficiently broad or deep to quickly penetrate the texts in this series. That is to say, relatively inexperienced readers will find they have some work to do, even to get into topics with which they have some familiarity. The best way to proceed in these cases depends, to some extent, on the nature of one's background.

I am familiar with some or all of the applied topics, but not with the mathematics. For readers with an engineering background, even at the graduate level, the depth with which topics are covered in these books is perhaps a little daunting. The best approach for such readers is to select the applied topic they wish to learn more about, and then use the text as a guide. When a new topic is initiated, it is clearly stated what parts of the book the reader is expected to be familiar with. The reader with a more applied background will find that they will not be able to get far without having to unravel the mathematical background almost to the beginning. Indeed, readers with a typical applied background will normally be lacking a good background in linear algebra and real analysis. Therefore, they will need to invest a good deal of effort acquiring some quite basic background. At this time, they will quickly be able to ascertain whether it is worth proceeding with reading the books in this series.

I am familiar with some or all of the mathematics, but not with the applied topics. Readers with an undergraduate degree in mathematics will fall into this camp, and probably also some readers with a graduate education in engineering, depending on their discipline. They may want to skim the relevant background material, just to see what they know and what they don't know, and then proceed directly to the applied topics of interest.

I am familiar with most of the contents. For these readers, the series is one of reference books.

Comments on organisation

In the current practise of teaching areas of science and engineering connected with mathematics, there is much emphasis on "just in time" delivery of mathematical ideas and techniques. Certainly I have employed this idea myself in the classroom, without thinking much about it, and so apparently I think it a good thing. However, the merits of the "just in time" approach in written work are, in my opinion, debatable. The most glaring difficulty is that the same mathematical ideas can be "just in time" for multiple non-mathematical topics. This can even happen in a single one semester course. For example—to stick to something germane to this series—are differential equations "just in time" for general system theory? for modelling? for feedback control theory? The answer is, "For all of them," of course. However, were one to choose one of these topics for a "just in time" written delivery of the material, the presentation would immediately become awkward, especially in the case where that topic were one that an instructor did not wish to cover in class.

Another drawback to a "just in time" approach in written work is that, when combined with the corresponding approach in the classroom, a connection, perhaps unsuitably strong, is drawn between an area of mathematics and an area of application of mathematics. Given that one of the strengths of mathematics is to facilitate the connecting of seemingly disparate topics, inside and outside of mathematics proper, this is perhaps an overly simplifying way of delivering mathematical material. In the "just simple enough, but not too simple" spectrum, we fall on the side of "not too simple."

For these reasons and others, the material in this series is generally organised according to its mathematical structure. That is to say, mathematical topics are treated independently and thoroughly, reflecting the fact that they have life independent of any specific area of application. We do not, however, slavishly follow the Bourbaki[1] ideals of logical structure. That is to say, we do allow ourselves the occasional forward reference when convenient. However, we are certainly careful to maintain the standards of deductive logic that currently pervade the subject of "mainstream" mathematics. We also do not slavishly follow the Bourbaki dictum of starting with the most general ideas, and proceeding to the more specific. While there is something to be said for this, we feel that for the subject and intended readership of this series, such an approach would be unnecessarily off-putting.

Andrew D. Lewis Kingston, ON, Canada

[1] Bourbaki refers to "Nicolas Bourbaki," a pseudonym given (by themselves) to a group of French mathematicians who, beginning in mid-1930's, undertook to rewrite the subject of mathematics. Their dictums include presenting material in a completely logical order, where no concept is referred to before being defined, and starting developments from the most general, and proceeding to the more specific. The original members include Henri Cartan, André Weil, Jean Delsarte, Jean Dieudonné, and Claude Chevalley, and the group later counted such mathematicians as Roger Godement, Jean-Pierre Serre, Laurent Schwartz, Emile Borel, and Alexander Grothendieck among its members. They have produced eight books on fundamental subjects of mathematics.

Preface for Volume 3

In this third volume of this five volume series, we depart decidedly from the coverage of material that is typically covered in an undergraduate programme in either mathematics or in engineering and the physical sciences. Most of the material in this volume is covered only at the graduate level, with some material being basic at the graduate level and some material only being covered in specialised courses of the graduate level.

We begin in Chapter 1 by presenting material at a general level that touches upon topics touched on in Chapters I-2, I-3, and II-1. The subject of this chapter, "general topology," is one that is often not taught at all in an undergraduate curriculum, and also often not in a dedicated course at the graduate level, although some of the material inevitably finds its way into the corpus through other courses, where some basic knowledge of topology is essential. Our view is that a systematic development of topological concepts is important to have at one's disposal eventually, as it is often most convenient and clear to introduce ideas in their most general context. That being said, in Chapter 1 we cover the topics of openness, closedness, compactness, convergence, and continuity in the general setting of topological spaces, where they are most naturally defined. We also work with other topics in general topology, as these provide important context for much material to follow. Our presentation here is a more or less standard one; in topics covered at the graduate level in mathematics, the style of treatment typically merges with our default style, which is to be focussed on structure and proving useful theorems related to this structure.

The next topic covered is measure theory, in Chapter 2. This subject *is* almost always part of the core of graduate mathematics education, as it is difficult to do anything resembling serious analysis without having the tools of measure theory at hand to understand basic parts of the subject. We present a comprehensive treatment of measure theory and integration in the general setting, and as well develop carefully and independently, both the single and multivariable Lebesgue integral. With the understanding that the Lebesgue measure and the Lebesgue integral are among the most challenging concepts to learn at a first encounter, substantial effort is devoted to understanding the motivations for the Lebesgue integral.

One of the results of a systematic development of measure theory is a collection of useful Banach spaces. In Chapter 3 we introduce the notion of a Banach space, and study some basic properties. We shall make essential use of the structure of Banach spaces subsequently in this volume and also in subsequent volumes. Banach spaces comprise a basic ingredient of basic functional analysis, and a development of their elementary properties is a part of the graduate level curriculum in mathematics. Moreover, again the graduate level, Banach spaces are a commonly used tool in applications of mathematics in engineering and the physical sciences. We develop the theory a little beyond the introductory level at which it is often presented, either to mathematics or to engineering graduate students. Our motivation for this is to have at hand the fuller context for Banach spaces and some of the subtleties involved once one explores their structure beyond what is on the surface.

A special sort of Banach space is the Hilbert space, to which we turn our attention in Chapter 4. Hilbert spaces are important in mathematics and in applications of mathematics since they offer a surprisingly rich extra structure beyond that of a general Banach space. (Indeed, one might say that the interesting thing about the structural theory of Banach spaces is the way in which they can generally differ from Hilbert spaces.)

In Chapter 5 we study convexity in a general way. The material we develop here has already been used in Chapter 3 to explain some important structural properties of Banach spaces; see Theorems 3.4.8 and 3.7.5. It will also be used in Chapter 6 to develop some tools in functional analysis. Moreover, basic ideas in convex analysis arise in essential ways in applied problems in optimisation theory. Indeed, a common approach in optimisation theory is to convert an optimisation problem into a convex problem, then declare victory.

One of the topics we cover that is not normally a part of the background of even the most mathematically inclined engineer or scientist is that of topological vector spaces, beyond the standard theory of Banach and Hilbert spaces. This material is covered in Chapter 6 of this volume. The need to cover this material is a consequence of the way we formulate our theory of differential and difference equations in Chapters V-4 and V-5, and the way we formulate system theory in Chapter V-6. Apart from this, the framework of topological vector spaces provides perhaps the clearest framework for understanding the theory of distributions which we present in Chapter IV-3.

The final topic of this volume is another mathematically specialised subject, but one that is important in the theory of linear systems. The subject is that of the special classes of holomorphic functions, known as "Hardy spaces." These spaces serve as a place where concepts from measure theory, functional analysis, complex analysis, and Fourier theory overlap.

Andrew D. Lewis Kingston, ON, Canada

This version: 2022/03/07

Table of Contents

1 Topology **1**
1.1 Metric spaces . . . 3
1.1.1 Definitions and simple examples . . . 3
1.1.2 Subsets of metric spaces . . . 5
1.1.3 Sequences in metric spaces . . . 7
1.1.4 Maps between metric spaces . . . 8
1.1.5 Semimetric spaces . . . 11
1.1.6 Completeness . . . 13
1.1.7 Completions of metric spaces . . . 15
1.1.8 Compact subsets of metric spaces . . . 20
Exercises . . . 23
1.2 Topological spaces . . . 26
1.2.1 Definitions and simple examples . . . 26
1.2.2 Bases and subbases . . . 29
1.2.3 Neighbourhoods and neighbourhood bases . . . 33
1.2.4 Interior, closure, boundary, etc. . . . 36
Exercises . . . 42
1.3 Continuity . . . 44
Exercises . . . 44
1.4 Subspaces, products, and quotients . . . 45
1.4.1 Subspaces of topological spaces . . . 45
1.5 Convergence in topological spaces . . . 46
1.5.1 The most general setting for Landau symbols . . . 46
1.6 Compactness . . . 47
Exercises . . . 47
1.7 Connectedness . . . 48
1.8 Countability and separability conditions . . . 49
1.8.1 Countability conditions . . . 49
1.8.2 Separability definitions . . . 49
1.9 The topologies of convergence of maps . . . 50
1.9.1 The pointwise convergence topology . . . 50
1.9.2 The uniform convergence topology . . . 50
1.9.3 The compact-open topology . . . 50
1.9.4 Maps between metric spaces . . . 50
Exercises . . . 51
1.10 Metrisable spaces . . . 52
1.11 Locally compact topological spaces . . . 53

1.11.1 Locally compact topological groups 53

2 Measure theory and integration **55**
2.1 Some motivation for abstract measure theory and integration 59
2.1.1 The Jordan measure and its limitations 59
2.1.2 Some limitations of the Riemann integral 69
2.1.3 An heuristic introduction to the Lebesgue integral 72
2.1.4 Notes 73
Exercises 73
2.2 Measurable sets 74
2.2.1 Algebras and σ-algebras 74
2.2.2 Algebras and σ-algebras generated by families of subsets . . 77
2.2.3 Products of measurable spaces 82
Exercises 84
2.3 Measures 86
2.3.1 Functions on families of subsets 86
2.3.2 Outer measures, measures, and their relationship 92
2.3.3 Complete measures and completions of measures 99
2.3.4 Outer and inner measures associated to a measure 101
2.3.5 Probability measures 109
2.3.6 Product measures 111
2.3.7 Signed measures 116
2.3.8 Complex measures 125
2.3.9 Vector measures 128
2.3.10 Spaces of positive, signed, complex, and vector measures . . 129
2.3.11 Notes 131
Exercises 131
2.4 Lebesgue measure on $\mathbb{R}$ 133
2.4.1 The Lebesgue outer measure and the Lebesgue measure on $\mathbb{R}$ 133
2.4.2 Borel sets in $\mathbb{R}$ as examples of Lebesgue measurable sets . . . 138
2.4.3 Further properties of the Lebesgue measure on $\mathbb{R}$ 143
2.4.4 Notes 149
Exercises 149
2.5 Lebesgue measure on $\mathbb{R}^n$ 151
2.5.1 The Lebesgue outer measure and the Lebesgue measure on $\mathbb{R}^n$ 151
2.5.2 Borel sets in $\mathbb{R}^n$ as examples of Lebesgue measurable sets . . 153
2.5.3 Further properties of the Lebesgue measure on $\mathbb{R}^n$ 158
2.5.4 Lebesgue measure on $\mathbb{R}^n$ as a product measure 165
2.5.5 Coverings of subsets of $\mathbb{R}^n$ 167
2.5.6 The Banach–Tarski Paradox 171
2.5.7 Notes 183
Exercises 183
2.6 Measurable functions 184

2.6.1 General measurable maps and functions . . . 184
2.6.2 Measurability and operations on functions . . . 190
2.6.3 Sequences of measurable functions . . . 196
2.6.4 $\mathbb{C}$- and vector-valued measurable functions . . . 203
2.6.5 Simple functions and approximations of measurable functions 206
2.6.6 Topological characterisations of convergence for sequences of measurable functions . . . 210
Exercises . . . 216
2.7 Integration on measure spaces . . . 217
2.7.1 Definition of the integral . . . 217
2.7.2 The integral and operations on functions . . . 227
2.7.3 Limit theorems . . . 232
2.7.4 Integration with respect to probability measures . . . 237
2.7.5 Topological characterisations of limit theorems . . . 241
2.7.6 Image measure and integration by image measure . . . 243
2.7.7 The integral for $\mathbb{C}$- and vector-valued functions . . . 244
2.7.8 Integration with respect to signed, complex, and vector measures . . . 249
2.7.9 Notes . . . 260
Exercises . . . 260
2.8 Integration on products . . . 262
2.8.1 The product measure by integration . . . 262
2.8.2 The integral on product spaces . . . 265
2.8.3 Fubini's Theorem . . . 267
Exercises . . . 275
2.9 The single-variable Lebesgue integral . . . 276
2.9.1 Lebesgue measurable functions . . . 276
2.9.2 The (conditional) Lebesgue integral . . . 281
2.9.3 Properties of the Lebesgue integral . . . 284
2.9.4 Swapping operations with the Lebesgue integral . . . 290
2.9.5 Locally Lebesgue integrable functions . . . 293
2.9.6 Absolute continuity . . . 296
2.9.7 The Fundamental Theorem of Calculus for the Lebesgue integral . . . 303
2.9.8 Lebesgue points . . . 315
2.9.9 Maximal functions . . . 318
2.9.10 The change of variables formula . . . 318
2.9.11 Topological characterisations of the deficiencies of the Riemann integral . . . 320
Exercises . . . 322
2.10 The multivariable Lebesgue integral . . . 324
2.10.1 Lebesgue measurable functions . . . 324
2.10.2 The (conditional) Lebesgue integral . . . 327

2.10.3 Properties of the Lebesgue integral 327
2.10.4 Swapping operations with the Lebesgue integral 327
2.10.5 The change of variables formula 327
2.10.6 Locally integrable functions 327
2.11 Differentiation of measures 329
2.11.1 Absolutely continuous measures 329
2.11.2 Singular measures 340
2.11.3 Notes 340
2.12 Measures on locally compact topological spaces 341
2.12.1 Haar measure on a locally compact topological group 341
2.12.2 The case of $\mathbb{T}$ 341
2.12.3 The case of $\mathbb{R}$ 341
2.12.4 The case of $\mathbb{R}^n$ 348

3 Banach spaces **349**
3.1 Definitions and properties of normed vector spaces 352
3.1.1 Norms and seminorms 352
3.1.2 Open and closed subsets of normed vector spaces 358
3.1.3 Subspaces, direct sums, and quotients 363
Exercises 365
3.2 Sequences in normed vector spaces 367
3.2.1 Definitions and properties of sequences 367
3.2.2 Algebraic operations on sequences 369
3.2.3 Multiple sequences 370
Exercises 370
3.3 Completeness and completions 371
3.3.1 Completeness (Banach spaces) 371
3.3.2 Why completeness is important 375
3.3.3 Completeness and direct sums and quotients 377
3.3.4 Completions 378
Exercises 381
3.4 Series in normed vector spaces 383
3.4.1 Definitions and properties of series 383
3.4.2 Absolute and unconditional convergence 385
3.4.3 Algebraic operations on series 395
3.4.4 Multiple series 397
3.4.5 Cesàro convergence of sequences and series 398
3.4.6 Series in normed vector spaces with arbitrary index sets . . . 400
3.4.7 Notes 405
Exercises 405
3.5 Continuous maps between normed vector spaces 407
3.5.1 General continuous maps between normed vector spaces . . 407
3.5.2 Continuous linear maps between normed vector spaces . . . 410

3.5.3 Induced topologies on continuous linear maps 416
3.5.4 Invertibility of continuous linear maps 423
3.5.5 Spectral properties for continuous operators on Banach spaces 425
3.5.6 The Open Mapping Theorem and Closed Graph Theorem . . 428
Exercises 429
3.6 Topology of normed vector spaces 430
3.6.1 Properties of balls in normed vector spaces 430
3.6.2 Interior, closure, boundary, etc. 431
3.6.3 Compactness 436
3.6.4 Closed subspaces 439
3.6.5 Bases for normed vector spaces 448
3.6.6 Notes 450
Exercises 451
3.7 Split subspaces? 452
3.7.1 Split subspaces 452
3.7.2 Questions and answers about closed and split subspaces . . . 453
3.7.3 Projection constants 454
3.7.4 The proof of the Split Subspace Theorem 454
3.7.5 Notes 454
3.8 Examples of Banach spaces 455
3.8.1 The p-norms on $\mathbb{F}^n$ 455
3.8.2 Banach spaces of sequences 463
3.8.3 Banach spaces of direct sums 472
3.8.4 Banach spaces of continuous functions on $\mathbb{R}$ 477
3.8.5 Banach spaces of continuous functions on metric spaces . . . 485
3.8.6 Banach spaces of continuous functions on locally compact topological spaces 485
3.8.7 Banach spaces of integrable functions on $\mathbb{R}$ 485
3.8.8 Banach spaces of integrable functions on measure spaces . . 497
3.8.9 Banach spaces of measures 497
3.8.10 Notes 500
Exercises 500
3.9 Topological duals of Banach spaces 501
3.9.1 Reflexivity in Banach spaces 502
3.10 Examples of duals of Banach spaces 503
3.10.1 The dual of $\mathsf{L}^p(I;\mathbb{F})$ 503

4 Hilbert spaces **507**
4.1 Definitions and properties of inner product spaces 509
4.1.1 Inner products and semi-inner products 509
4.1.2 Inner product spaces as normed vector spaces 512
4.1.3 Orthogonality 517
4.1.4 Hilbert spaces and their subspaces 520

4.1.5 Minimising distance to a set . . . 525
4.1.6 Norms . . . 528
Exercises . . . 528
4.2 Continuous maps between inner product spaces . . . 531
4.2.1 The dual of an inner product space . . . 531
4.2.2 Particular aspects of continuity for inner product spaces . . . 534
4.2.3 The adjoint of a continuous linear map . . . 535
4.2.4 Spectral properties for operators on Hilbert spaces . . . 536
4.2.5 Notes . . . 539
Exercises . . . 539
4.3 Examples of Hilbert spaces . . . 542
Exercises . . . 542
4.4 Orthonormal bases in Hilbert spaces . . . 543
4.4.1 General definitions and results . . . 543
4.4.2 Finite orthonormal sets and finite Hilbert bases . . . 554
4.4.3 Enumerable orthonormal sets and enumerable Hilbert bases 558
4.4.4 Generalised Fourier series . . . 565
4.4.5 Classification of Hilbert spaces . . . 567
4.4.6 Notes . . . 572
Exercises . . . 572

5 Convexity **575**
5.1 Convex bodies . . . 576
5.1.1 Definitions . . . 576
5.1.2 Maximal and minimal ellipsoids . . . 579
5.1.3 Notes . . . 592
5.2 Dvoretzky's Theorem . . . 593
5.2.1 Concentration of Gaussian measure . . . 593
5.2.2 Dvoretzky's Theorem . . . 600
5.2.3 Consequences of Dvoretzky's Theorem . . . 610
5.2.4 Notes . . . 617

6 Topological vector spaces **619**
6.1 General topological vector spaces . . . 621
6.1.1 Why go beyond norms? . . . 621
6.1.2 Definitions and basic properties . . . 623
6.1.3 Boundedness in topological vector spaces . . . 626
6.1.4 Attributes of subsets of a topological vector space . . . 628
6.1.5 Completeness of topological vector spaces . . . 629
Exercises . . . 629
6.2 Locally convex topological vector spaces . . . 631
6.2.1 Seminorms and Minkowski functionals . . . 631
6.2.2 Seminorms and locally convex topologies . . . 634

6.2.3 Properties of locally convex topological vector spaces 638
Exercises 641
6.3 Special classes of topological vector spaces 642
6.3.1 Finite-dimensional spaces 642
6.3.2 Normed spaces 643
6.3.3 Metrisable spaces 643
6.3.4 Countable union spaces 643
6.4 Linear maps between locally convex topological vector spaces 645
6.4.1 Duals of locally convex topological vector spaces 645
6.4.2 Reflexivity in locally convex topological vector spaces 645
6.4.3 Duals of Fréchet spaces 645
6.4.4 Duals of countable union spaces 646
6.5 Examples of locally convex spaces 648
6.5.1 Locally convex spaces of sequences 648
6.5.2 Locally convex spaces of continuous functions on $\mathbb{R}$ 650
6.5.3 Locally convex spaces of continuous functions on metric spaces 652
6.5.4 Locally convex spaces of locally integrable functions on $\mathbb{R}$ . . 652
6.5.5 Distributions as elements of locally convex topological vector spaces 653
Exercises 655

7 Hardy spaces **659**
7.1 Preliminary constructions and results 661
7.1.1 The Hardy–Littlewood maximal function 661
7.1.2 Nontangential limits 668
7.1.3 Constructions with subharmonic functions 670
7.2 Hardy spaces of harmonic functions defined on vertical strips 671
7.2.1 Definitions 671
7.2.2 Poisson integral representations of harmonic functions on vertical strips 672
7.2.3 Nontangential limits for the harmonic Hardy spaces on vertical strips 686
Exercises 693
7.3 Hardy spaces of harmonic functions defined on annuli 694
7.3.1 Definitions 694
7.3.2 Poisson integral representations of harmonic functions on annuli 696
7.3.3 Nontangential limits for the harmonic Hardy spaces on annuli 696
7.4 Hardy spaces of holomorphic functions defined on vertical strips . . 697
7.4.1 Definitions 697
7.4.2 Poisson integral representations of holomorphic functions on vertical strips 698

7.4.3 Banach spaces of holomorphic functions on vertical strips . . 698
7.4.4 Locally convex topological vector spaces of holomorphic functions on vertical strips 701
7.5 Hardy spaces of holomorphic functions defined on annuli 704
7.5.1 Definitions 704
7.5.2 Poisson integral representations of holomorphic functions on annuli 705
7.5.3 Banach spaces of holomorphic functions on annuli 705
7.5.4 Locally convex topological vector spaces of holomorphic functions on annuli 705
7.5.5 Hardy spaces of vector-valued holomorphic functions defined on annuli 705

Chapter 1

Topology

It is possible to say much of what we have to say in this series in ways that avoid the subject of topology in its full generality. However, as one progresses to more and more abstract notions, topology is helpful in connecting seemingly disparate ideas, and so allowing one's intuition to extend into areas where it may have difficulty going otherwise. Moreover, at some point the effort required to *not* use topology overtakes the effort of using it.

The subject of topology has as its basis the fact that many concepts in analysis have at their centre the idea of an open set. For example, the usual "ϵ–δ" notion of continuity is equivalent to a requirement that the preimage of open sets be themselves open. Thus, in topology one introduces axioms on what constitutes a collection of open sets. This forms the starting point for notions like closedness, compactness, connectedness, etc. While the abstraction of topology is unnecessary to understand basic material such as we encountered in Chapters I-2 and II-3, it is very useful to understand topology when we get to the more abstract material such as in Chapter 3 and particularly Chapter 6.

Our presentation of topology is a little biased because of the use we make of topology in these volumes. Indeed, our excursions into topology exclusively come by way of analysis. For this reason, it is notions such as continuity, compactness, and completeness that are of most interest to us. But there is another important side to topology and that is as a tool for understanding the geometric properties of a space. This typically ends up in algebraic topology and we refer the reader to [Munkres 1984] as an example of a text in this area.

Do I need to read this chapter? At just what point a reader decides to engage this material is something of a matter of choice. An attempt has been made to ensure that the presentation builds on the material of a topological nature that has already been presented in simpler settings in Chapters I-2, I-3, and II-3. Also, the material in Chapter 3 will provide a useful context for some of the abstraction encountered here. Some readers may be happier reading Chapter 3 before getting to the material here. However, this chapter on topology will be assumed in our presentation in Chapter 6, and also in parts of Chapter 2. •

Contents

1.1 Metric spaces . . . 3
1.1.1 Definitions and simple examples . . . 3
1.1.2 Subsets of metric spaces . . . 5
1.1.3 Sequences in metric spaces . . . 7
1.1.4 Maps between metric spaces . . . 8
1.1.5 Semimetric spaces . . . 11
1.1.6 Completeness . . . 13
1.1.7 Completions of metric spaces . . . 15
1.1.8 Compact subsets of metric spaces . . . 20
Exercises . . . 23
1.2 Topological spaces . . . 26
1.2.1 Definitions and simple examples . . . 26
1.2.2 Bases and subbases . . . 29
1.2.3 Neighbourhoods and neighbourhood bases . . . 33
1.2.4 Interior, closure, boundary, etc. . . . 36
Exercises . . . 42
1.3 Continuity . . . 44
Exercises . . . 44
1.4 Subspaces, products, and quotients . . . 45
1.4.1 Subspaces of topological spaces . . . 45
1.5 Convergence in topological spaces . . . 46
1.5.1 The most general setting for Landau symbols . . . 46
1.6 Compactness . . . 47
Exercises . . . 47
1.7 Connectedness . . . 48
1.8 Countability and separability conditions . . . 49
1.8.1 Countability conditions . . . 49
1.8.2 Separability definitions . . . 49
1.9 The topologies of convergence of maps . . . 50
1.9.1 The pointwise convergence topology . . . 50
1.9.2 The uniform convergence topology . . . 50
1.9.3 The compact-open topology . . . 50
1.9.4 Maps between metric spaces . . . 50
Exercises . . . 51
1.10 Metrisable spaces . . . 52
1.11 Locally compact topological spaces . . . 53
1.11.1 Locally compact topological groups . . . 53

Section 1.1

Metric spaces

The simplest, but still reasonably general, form of a what we shall later dub a topological space is a metric space. Almost all, but not all, topological spaces we encounter in these volumes will be metric spaces, at least in the sense that their topology is equivalent (in a sense we shall make precise) to that of a metric space. Metric spaces, as we shall see, generalise the ideas with which we are familiar from real and complex analysis. In this section we will not present a complete picture of metric space theory, or even that part of it that we will use. Instead, we concentrate on some basic ideas connected with metric spaces—open and closed sets, continuous maps, sequences, completeness, and compactness—and how these lead naturally to generalisations of metric spaces to topological spaces. Some other properties of metric spaces will be presented in the course of the remainder of the chapter.

Do I need to read this section? If one is encountering the material in this chapter for the first time, this is the place to start. There are not too many concepts in topology that are not best introduced via metric spaces. •

1.1.1 Definitions and simple examples

We begin by defining the notion of a metric space. Intuitively, a metric gives the structure of "distance" between two points in a set. As we shall see, this generalises our usual perception of distance in Euclidean space.

1.1.1 Definition (Metric space) A ***metric*** on a set S is a map $d\colon S \times S \to \mathbb{R}$ with the following properties:

(i) $d(x_1, x_2) = d(x_2, x_1)$ for all $x_1, x_2 \in S$ (***symmetry***);

(ii) $d(x_1, x_2) = 0$ if and only if $x_1 = x_2$ (***definiteness***);

(iii) $d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3)$ for all $x_1, x_2, x_3 \in S$ (***triangle inequality***).

A ***metric space*** is a pair (S, d) where d is a metric on the set S. •

Note that the triangle inequality gives $d(x_1, x_1) \leq 2d(x_1, x_2)$ for all $x_1, x_2 \in S$, and therefore $d(x_1, x_2) \in \mathbb{R}_{\geq 0}$ for all $x_1, x_2 \in S$.

Let us give some simple examples of metric spaces, some that we have encountered before, and some that are new.

1.1.2 Examples (Metric spaces)

1. Let $S = \mathbb{R}$ and take $d_{\mathbb{R}}(x_1, x_2) = |x_1 - x_2|$. The properties of the absolute value given in Proposition I-2.2.12 allow us to verify the properties that ensure that $d_{\mathbb{R}}$ is a metric. Unless we explicitly state otherwise, we shall always assume we are using the metric $d_{\mathbb{R}}$ on $\mathbb{R}$.

what

2. Let $S = \mathbb{C}$ and take $d_{\mathbb{C}}(z_1, z_2) = |z_1 - z_2|$. From the properties of the complex magnitude function given in it follows that $d_{\mathbb{C}}$ is a metric. Unless we explicitly state otherwise, we shall always assume we are using the metric $d_{\mathbb{C}}$ on $\mathbb{C}$.
3. Let $S = \mathbb{R}^n$ and define

$$d_{\mathbb{R}^n}(\boldsymbol{x}, \boldsymbol{y}) = \Big(\sum_{j=1}^{n}(x_j - y_j)^2\Big)^{1/2}.$$

   That this defines a metric on $\mathbb{R}^n$ follows from Proposition II-1.1.8.
4. For any set S, define

$$d_{\mathrm{disc}}(x_1, x_2) = \begin{cases} 0, & x_1 = x_2, \\ 1, & x_1 \neq x_2. \end{cases}$$

   In this case, it is straightforward to check the metric properties for d. While this shows that every set can be made into a metric space, the metric is not that useful in general. The subscript "disc" stands for ***discrete metric***, and refers to the fact that, as we shall see, d_{disc} defines the so-called "discrete topology" on S.
5. On $\overline{\mathbb{R}}$ define a metric d by

$$d(x, y) = \begin{cases} \left|\frac{x}{1+|x|} - \frac{y}{1+|y|}\right|, & x, y \in \mathbb{R}, \\ \left|\frac{x}{1+|x|} - 1\right|, & x \in \mathbb{R},\ y = \infty, \\ \left|1 - \frac{y}{1+|y|}\right|, & x = \infty,\ y \in \mathbb{R}, \\ \left|\frac{x}{1+|x|} + 1\right|, & x \in \mathbb{R},\ y = -\infty, \\ \left|1 + \frac{y}{1+|y|}\right|, & x = -\infty,\ y \in \mathbb{R}, \\ 0, & x = y = -\infty \text{ or } x = y = \infty, \\ 2, & x = \infty,\ y = -\infty \text{ or } x = -\infty,\ y = \infty. \end{cases}$$

   We leave the verification that this is a metric space to the reader as Exercise 1.1.1.
6. If (S, d) is a metric space and $A \subseteq S$, then the map $d_A \colon A \times A \to \mathbb{R}$ defined by $d_A(x_1, x_2) = d(x_1, x_2)$ is a metric on A. •

We shall see many more examples of metric spaces in Chapter 6.
The following is a useful inequality concerning the metric.

1.1.3 Proposition (Two useful metric inequalities) *If* (S, d) *is a metric space, then*

(i) $|\mathrm{d}(\mathrm{x}_1, \mathrm{x}_0) - \mathrm{d}(\mathrm{x}_2, \mathrm{x}_0)| \leq \mathrm{d}(\mathrm{x}_1, \mathrm{x}_2)$ *for all* $\mathrm{x}_0, \mathrm{x}_1, \mathrm{x}_2 \in \mathrm{S}$*, and*

(ii) $|\mathrm{d}(\mathrm{x}_1, \mathrm{x}_3) - \mathrm{d}(\mathrm{x}_2, \mathrm{x}_4)| \leq \mathrm{d}(\mathrm{x}_1, \mathrm{x}_2) + \mathrm{d}(\mathrm{x}_3, \mathrm{x}_4)$ *for all* $\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3, \mathrm{x}_4 \in \mathrm{S}$.

Proof (i) By the triangle inequality we have

$$d(x_2, x_0) \leq d(x_1, x_0) + d(x_1, x_2), \quad d(x_1, x_0) \leq d(x_2, x_0) + d(x_1, x_2).$$

Therefore,

$$d(x_2, x_0) - d(x_1, x_0) \leq d(x_1, x_2), \quad d(x_1, x_0) - d(x_2, x_0) \leq d(x_1, x_2),$$

implying that $|d(x_2, x_0) - d(x_1, x_0)| \leq d(x_1, x_2)$, as desired.

(ii) We have

$$\begin{aligned} |d(x_1, x_3) - d(x_2, x_4)| &= |d(x_1, x_3) - d(x_2, x_3) + d(x_2, x_3) - d(x_2, x_4)| \\ &\leq |d(x_1, x_3) - d(x_2, x_3)| + |d(x_2, x_3) - d(x_2, x_4)| \\ &\leq d(x_1, x_2) + d(x_3, x_4), \end{aligned}$$

where we have used the triangle inequality for $d_{\mathbb{R}}$ and part (i). ■

1.1.2 Subsets of metric spaces

We have already seen that, for $\mathbb{R}$ and $\mathbb{C}$, there exists the notion of what it means for a set to be open. These notions are easily adapted to general metric spaces.

1.1.4 Definition (Open, closed, and bounded sets in metric spaces) Let (S, d) be a metric space.

(i) The ***open* d*-ball*** of radius r about $x \in S$ is the set

$$\mathsf{B}_d(r, x) = \{y \in S \mid d(x, y) < r\}.$$

(ii) The ***closed* d*-ball*** of radius r about $x \in S$ is the set

$$\overline{\mathsf{B}}_d(r, x) = \{y \in S \mid d(x, y) \leq r\}.$$

(iii) A subset $U \subseteq S$ is ***open*** if, for each $x \in U$, there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}_d(\epsilon, x) \subseteq U$. (The empty set is also open, by declaration.)

(iv) A subset $A \subseteq S$ is ***closed*** if $S \setminus A$ is open.

(v) A subset $A \subseteq S$ is ***bounded*** if there exists $x_0 \in S$ and $R \in \mathbb{R}_{>0}$ such that $S \subseteq \mathsf{B}_d(R, x_0)$.

(vi) A subset $A \subseteq S$ is ***totally bounded*** if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists a finite set $\{x_1, \dots, x_k\} \subseteq S$ such that $A \subseteq \mathsf{B}_d(\epsilon, x_1) \cup \cdots \cup \mathsf{B}_d(\epsilon, x_k)$.

(vii) The ***diameter*** of a subset $A \subseteq S$ is

$$\operatorname{diam}(A) = \sup\{d(x_1, x_2) \mid x_1, x_2 \in A\}$$ •

Let us see what open sets look like in our examples.

1.1.5 Examples (Example 1.1.2 cont'd)

1. The open subsets of the metric space $(\mathbb{R}, d_{\mathbb{R}})$ are, recalling their definition from Definition I-2.5.2, exactly the open sets as we usually think of them.
2. This is similarly true for the metric space $(\mathbb{C}, d_{\mathbb{C}})$, recalling from Definition II-3.1.4 the notion of an open subset of $\mathbb{C}$.
3. Since this is the first occasion we have talked systematically about open subsets of $\mathbb{R}^n$, we take Definition 1.1.4 as the *definition* of what we mean by an open subset of $\mathbb{R}^n$, noting that this agrees with the usual situation when $n = 1$.

4. In the case of the metric space (S, d_{disc}), *every* subset is open. Indeed, let $A \subseteq S$ and let $x \in A$. Then clearly $\mathsf{B}_{d_{\text{disc}}}(x, \frac{1}{2}) = \{x\} \subseteq A$.
5. For a metric space (S, d) and a subset $A \subseteq S$, we let d_A be the metric on A inherited from S. We then have $\mathsf{B}_{d_A}(r, x) = A \cap \mathsf{B}_d(r, x)$ for each $x \in A$. Using this fact, one can easily show that a subset $U \subseteq A$ is open if and only if there exists an open subset $\tilde{U} \subseteq S$ for which $U = A \cap \tilde{U}$. We leave the details of this to the reader as Exercise 1.1.3. •

Let us discuss the properties of open sets that we will generalise when we discuss topological spaces in Section 1.2.

1.1.6 Proposition (Properties of open subsets of metric spaces) *If* (S, d) *is a metric space, then the following statements hold:*

(i) for $(\mathrm{U}_\mathrm{a})_{\mathrm{a} \in \mathrm{A}}$ *an arbitrary family of open sets,* $\cup_{\mathrm{a} \in \mathrm{A}} \mathrm{U}_\mathrm{a}$ *is open;*

(ii) for $(\mathrm{U}_1, \dots, \mathrm{U}_\mathrm{n})$ *a finite family of open sets,* $\cap_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{U}_\mathrm{j}$ *is open.*

Proof (i) Let $x \in \cup_{a \in A} U_a$. Then, since $x \in U_{a_0}$ for some $a_0 \in A$, there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}_d(\epsilon, x) \subseteq U_{a_0} \subseteq \cup_{a \in A} U_a$.

(ii) Let $x \in \cap_{j=1}^n U_j$. For each $j \in \{1, \dots, n\}$, choose $\epsilon_j \in \mathbb{R}_{>0}$ such that $\mathsf{B}_d(\epsilon_j, x) \subseteq U_j$, and let $\epsilon = \min\{\epsilon_1, \dots, \epsilon_n\}$. Then $\mathsf{B}_d(\epsilon, x) \subseteq U_j$, $j \in \{1, \dots, n\}$, and so $\mathsf{B}_d(\epsilon, x) \subseteq \cap_{j=1}^n U_j$. ■

As with subsets of $\mathbb{R}$ we can define the interior, closure, and boundary of a subset of a metric space. Here we just give the definitions; general definitions and properties of interior, closure, and boundary in this general setup, will be given in .

where?

1.1.7 Definition (Interior, closure, boundary) Let (S, d) be a metric space and let $A \subseteq S$.

(i) The ***interior*** of A is the set

$$\operatorname{int}(A) = \cup\{U \mid U \subseteq A, \ U \text{ open}\}.$$

(ii) The ***closure*** of A is the set

$$\operatorname{cl}(A) = \cap\{C \mid A \subseteq C, \ C \text{ closed}\}.$$

(iii) The ***boundary*** of A is the set $\operatorname{bd}(A) = \operatorname{cl}(A) \cap \operatorname{cl}(\mathbb{R} \setminus A)$. •

The following example illustrates that things are not always as intuitive for general metric spaces as they are for the metric spaces with which you are familiar.

1.1.8 Example (The closure of an open ball may not be the closed ball) Let $S = \mathbb{R} \setminus (1, 2)$ and equip S with the metric $d(x, y) = |x - y|$ coming from the fact that $S \subseteq \mathbb{R}$. Then

$$\mathsf{B}_d(2, 0) = (-2, 1), \quad \operatorname{cl}(\mathsf{B}_d(2, 0)) = [-2, 1), \quad \overline{\mathsf{B}}_d(2, 0) = [-2, 1) \cup \{2\}.$$

Thus the closure of the open ball of radius 2 about 0 is strictly contained in closed ball of radius 2 about 0.

Note that it will always be the case, for any metric space, that $\operatorname{cl}(\mathsf{B}_d(r, x)) \subseteq \overline{\mathsf{B}}_d(r, x)$ (why?). •

1.1.3 Sequences in metric spaces

In this section we continue our project of defining a concept for general metric spaces that extends, in a fairly obvious way, notions we have already seen. Then we provide an alternative characterisation that will lead to the more general setting of topological space in subsequent sections.

In this section, the topic is convergence of sequences.

1.1.9 Definition (Convergent sequences in metric spaces) Let (S, d) be a metric space and let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in S.

(i) The sequence ***converges*** to $x_0 \in S$ if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $d(x_j, x_0) < \epsilon$ for all $j \geq N$.

(ii) If the sequence converges to x_0, then x_0 is the ***limit*** of the sequence.

(iii) The sequence is ***convergent*** if it converges to some $x_0 \in S$.

(iv) The sequence is ***divergent*** if it is not convergent.

(v) The sequence is ***bounded*** if there exists $x_0 \in S$ and $R \in \mathbb{R}_{>0}$ such that $x_j \in \mathsf{B}_d(R, x_0)$ for all $j \in \mathbb{Z}_{>0}$. •

1.1.10 Remark The definition reflects the intuitive idea that the tail of a sequence converging to x_0 lies near x_0. One can state this as follows. Given $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $x_j \in \mathsf{B}_d(\epsilon, x_0)$ for all $j \geq N$. •

The notions of convergent sequences in $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{R}^n$ are simple, so we forgo examples of these. Let us merely look at some more general situations.

1.1.11 Examples (Convergent sequences)

1. If (S, d_{disc}) is a metric space with the discrete topology, then a sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ that converges to $x_0 \in S$ must have the property that there exists $N \in \mathbb{Z}_{>0}$ such that $x_j = x_0$ for all $j \geq N$.
2. If (S, d) is a metric space and if $A \subseteq S$ is equipped with the metric d_A defined by restricting d to A, then a sequence $(x_j)_{j\in\mathbb{Z}_{>0}} \subseteq A$ converges to $x_0 \in A$ if and only if it converges to x_0, thinking of $(x_j)_{j\in\mathbb{Z}_{>0}}$ as a sequence in S. •

Now let us give a notion of convergence of sequences in metric spaces that will suggest how this can be done in topological spaces.

1.1.12 Proposition (Equivalent notion of convergence) *Let* (S, d) *be a metric space. For a sequence* $(\mathrm{x_j})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *in* S*, the following statements are equivalent:*

(i) the sequence converges to $\mathrm{x}_0 \in \mathrm{S}$*;*

(ii) for each open subset $\mathrm{U} \subseteq \mathrm{S}$ *containing* x_0*, there exists* $\mathrm{N} \in \mathbb{Z}_{>0}$ *such that* $\mathrm{x_j} \in \mathrm{U}$ *for each* $\mathrm{j} \geq \mathrm{N}$*.*

Proof (i) $\implies$ (ii) Let $U \subseteq S$ be an open subset containing x_0 and let $\epsilon \in \mathbb{R}_{>0}$ have the property that $\mathsf{B}_d(\epsilon, x_0) \subseteq U$. Then there exists $N \in \mathbb{Z}_{>0}$ such that $x_j \in \mathsf{B}_d(\epsilon, x_0) \subseteq U$ for each $j \geq N$.

(ii) $\implies$ (i) Let $\epsilon \in \mathbb{R}_{>0}$ and note that $\mathsf{B}_d(\epsilon, x_0)$ is an open set containing x_0. Thus there exists $N \in \mathbb{Z}_{>0}$ such that $x_j \in \mathsf{B}_d(\epsilon, x_0)$ for $j \geq N$, so completing the proof. ■

As with our investigation of sequences in $\mathbb{Q}$ and $\mathbb{R}$ in Section I-2.1, the notion of a Cauchy sequence is an important one in understanding the properties of a metric space.

1.1.13 Definition (Cauchy sequence) Let (S, d) be a metric space. A sequence $(x_j)_{j \in \mathbb{Z}_{>0}}$ is a ***Cauchy sequence*** if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $d(x_j, x_k) < \epsilon$ for all $j, k \geq N$. •

Cauchy sequences have the useful property of being bounded.

1.1.14 Proposition (Cauchy sequences are bounded) *If* (S, d) *is a metric space and if* $(\mathrm{x}_j)_{j \in \mathbb{Z}_{>0}}$ *is a Cauchy sequence then the sequence is also bounded.*

Proof Let $x_0 \in S$. Choose $N \in \mathbb{Z}_{>0}$ such that $d(v_j, v_k) < 1$ for $j, k \in \mathbb{Z}_{>0}$. Then take R_N to be the largest of the nonnegative real numbers $d(x_1, x_0), \dots, d(x_N, x_0)$. Then, for $j \geq N$ we have, using the triangle inequality,

$$d(x_j, x_0) = d(x_j, x_N) + d(x_N, x_0) \leq 1 + R_N,$$

giving the result by taking $R = R_N + 1$. ■

The notion of a Cauchy sequence is more subtle than that of a convergent sequence in that elements do not get close to a limit, but only to one another. Nonetheless, one intuitively expects there to be some connections between Cauchy sequences and convergent sequences. One implication is true.

1.1.15 Proposition (Convergent sequences are Cauchy) *If* (S, d) *is a metric space, and if* $(\mathrm{x}_j)_{j \in \mathbb{Z}_{>0}}$ *is a sequence converging to* $\mathrm{x}_0 \in \mathrm{S}$*, then* $(\mathrm{x}_j)_{j \in \mathbb{Z}_{>0}}$ *is a Cauchy sequence.*

Proof Let $\epsilon \in \mathbb{R}_{>0}$ and choose $N \in \mathbb{Z}_{>0}$ such that $d(x_j, x_0) \leq \frac{\epsilon}{2}$ for all $j \geq N$. Then, for $j, k \geq N$, we have $d(x_j, x_k) \leq d(x_j, x_0) + d(x_0, x_j) < \epsilon$, using the triangle inequality. ■

The matter of whether Cauchy sequences converge is the matter of completeness which we discuss in detail in Section 1.1.6.

1.1.4 Maps between metric spaces

In the preceding section, we discussed open subsets of metric spaces. The properties of these, as given in Proposition 1.1.6, shall form the basis for our definition of a topological space. In this section, we carry out this objective for the notion of continuity. That is to say, we give a definition of a continuous map between metric spaces, and then characterise this notion of continuity in a way that is easily generalised to topological spaces.

First the definition which generalises the notions of continuity we have seen in Definitions I-3.1.1 and II-3.2.4.

1.1.16 Definition (Continuity of maps between metric spaces) Let (S_1, d_1) and (S_2, d_2) be metric spaces, and let $f\colon S_1 \to S_2$ be a map. Then

(i) f is ***continuous at*** $\mathbf{x_0}$ if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that $d_1(x, x_0) < \delta$ implies that $d_2(f(x), f(x_0)) < \epsilon$,

(ii) f is ***continuous*** if it is continuous at each $x \in S_1$,

(iii) f is ***discontinuous at*** $\mathbf{x_0}$ if it is not continuous at x_0, and

(iv) f is ***discontinuous*** if it is not continuous.

(v) f is ***uniformly continuous*** if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that $d_1(x_1, x_2) < \delta$ implies that $d_2(f(x_1), f(x_2)) < \epsilon$. •

1.1.17 Remark (Alternative characterisation of continuity) It is easy to give the following useful alternate characterisation of continuity: a map $f\colon S_1 \to S_2$ is continuous at x_0 if and only if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that $f(\mathsf{B}_{d_1}(\delta, x_0)) \subseteq \mathsf{B}_{d_2}(\epsilon, f(x_0))$. In words, f is continuous at x_0 if and only if every ball around $f(x_0)$ contains the image under f of some ball around x_0. •

Let us give some examples of maps that are continuous or not, in order to illustrate the concept. We refer the reader to Section I-3.1 for simpler examples. In particular, we refer the reader to Example I-3.1.7 for an example of a map that is continuous but not uniformly continuous.

1.1.18 Examples (Continuous and discontinuous maps)

1. Let (S, d_{disc}) be a metric space with the discrete metric, and let (T, d) be any other metric space. We claim that *every* map $f\colon S \to T$ is continuous. Indeed, let $x_0 \in S$ and let $\epsilon \in \mathbb{R}_{>0}$. If $\delta = \frac{1}{2}$ then $d(x, x_0) < \delta$ implies that $x = x_0$ and so $d(f(x), f(x_0)) = 0 < \epsilon$. •

Now we give an equivalent characterisation of continuity for metric spaces, one that will be key in motivating our general notion of continuity for topological spaces.

1.1.19 Theorem (Equivalent notions of continuity) *For metric spaces* (S_1, d_1) *and* (S_2, d_2)*, the following two statements concerning a map* $f\colon S_1 \to S_2$ *are equivalent:*

(i) f *is continuous at* x_0*;*

(ii) for every open set $V \subseteq S_2$ *containing* $f(x_0)$*, there exists an open subset* $U \subseteq S_1$ *containing* x_0 *and such that* $f(U) \subseteq V$*;*

(iii) if the sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ *in* S_1 *converges to* x_0 *then the sequence* $(f(x_j))_{j\in\mathbb{Z}_{>0}}$ *converges to* $f(x_0)$ *in* S_2.

In particular, f *is continuous if and only if* $f^{-1}(V) \subseteq S_1$ *is open for every open subset* $V \subseteq S_2$.

Proof (i) $\Longrightarrow$ (ii) Let $V \subseteq S_2$ be an open subset containing $f(x_0)$ and let $\epsilon \in \mathbb{R}_{>0}$ be such that $\mathsf{B}_{d_2}(\epsilon, f(x_0)) \subseteq V$. Then, since f is continuous, there exists $\delta \in \mathbb{R}_{>0}$ such that $f(\mathsf{B}_{d_1}(\delta, x_0)) \subseteq \mathsf{B}_{d_2}(\epsilon, f(x_0)) \subseteq V$. The result follows by taking $U = \mathsf{B}_{d_1}(\delta, x_0)$.

(ii) $\implies$ (iii) Let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in S_1 converging to x_0 and let $\epsilon \in \mathbb{R}_{>0}$. By hypothesis there exists a neighbourhood U of x_0 in S_1 such that $f(U) \subseteq \mathsf{B}_{d_2}(\epsilon, f(x_0))$. Thus there exists $\delta \in \mathbb{R}_{>0}$ such that $f(\mathsf{B}_{d_1}(\delta, x_0) \subseteq \mathsf{B}_{d_2}(\epsilon, f(x_0))$ since U is open in S_1. Now choose $N \in \mathbb{Z}_{>0}$ sufficiently large that $d_1(x_j, x_0) < \delta$ for $j \geq N$. It then follows that $d_2(f(x_j), f(x_0)) < \epsilon$ for $j \geq N$, so giving convergence of $(f(x_j))_{j\in\mathbb{Z}_{>0}}$ to $f(x_0)$, as desired.

(iii) $\implies$ (i) Let $\epsilon \in \mathbb{R}_{>0}$. Then, by definition of convergence, there exists $\delta \in \mathbb{R}_{>0}$ such that, for $x \in \mathsf{B}_{d_1}(\delta, x_0)$, $d(f(x), f(x_0)) < \epsilon$, which is exactly the definition of continuity of f at x_0.

For the final statement in the proposition, let V be an open subset of S_2 and note that, for each $y \in V$, there exists $\epsilon_y \in \mathbb{R}_{>0}$ such that $\mathsf{B}_{d_2}(\epsilon_y, y) \subseteq U$. Therefore, $V = \cup_{y\in V}\mathsf{B}_{d_2}(\epsilon_y, y)$, so that $f^{-1}(V) = \cup_{y\in V} f^{-1}(\mathsf{B}_{d_2}(\epsilon_y, y))$. Since each of the subsets $f^{-1}(\mathsf{B}_{d_2}(\epsilon_y, y))$ is open by the first part of the proposition, it follows from Proposition 1.1.6(i) that $f^{-1}(V)$ is open if f is continuous. That f is continuous if $f^{-1}(V)$ is open for every open V follows from the first part of the proposition. ■

A useful fact about metric spaces is that the metric is itself continuous, in the following sense.

1.1.20 Proposition (Continuity of the metric) *If* (S, d) *is a metric space and if* $x_0 \in S$*, then the map* $x \mapsto d(x, x_0)$ *is a continuous map from* (S, d) *to* $(\mathbb{R}, d_{\mathbb{R}})$*.*

Proof Let $x \in S$, let $\epsilon \in \mathbb{R}_{>0}$, and choose $\delta = \epsilon$. If $d(y, x) < \delta$ then we have

$$|d(y, x_0) - d(x, x_0)| \leq d(y, x) < \delta = \epsilon,$$

using Proposition 1.1.3(i), showing that the map of the proposition is continuous at x. ■

Another important feature of maps between metric spaces is one that concerns preservation of the metric.

1.1.21 Definition (Isometry) Let (S_1, d_1) and (S_2, d_2) be metric spaces. A map $f: S_1 \to S_2$ is an ***isometry*** if $d_2(f(x_1), f(x_2)) = d_1(x_1, x_2)$ for each $x_1, x_2 \in S_1$. •

Clearly, isometries are continuous. However, just as clearly, there are many continuous maps that are not isometries.

An important class of maps from a metric space to itself are those that reduce distance.

1.1.22 Definition (Contraction map) If (S, d) is a metric space, a map $f: S \to S$ is a ***contraction map*** if there exists $\lambda \in [0, 1)$ such that $d(f(x_1), f(x_2)) \leq \lambda d(x_1, x_2)$ for every $x_1, x_2 \in S$. •

Contraction maps have the following important property.

1.1.23 Theorem (Contraction Mapping Theorem) *If* (S, d) *is a complete metric space and if* $f: S \to S$ *is a contraction map then there exists a unique point* $x_0 \in S$ *with the property that* $f(x_0) = x_0$*.*

Proof Let $\lambda \in [0,1)$ have the property that $d(f(x_1), f(x_2)) \leq \lambda d(x_1, x_2)$ for every $x_1, x_2 \in S$. Let $y_0 \in S$ and define a sequence $(y_j)_{j\in\mathbb{Z}_{>0}}$ by asking that $y_1 = f(y_0)$ and then inductively by defining $y_{j+1} = f(y_j)$, $j \in \mathbb{Z}_{>0}$. We claim that $(y_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence. First of all we compute

$$\begin{aligned}
& d(y_1, y_2) = d(f(y_0), f(y_1)) \leq \lambda d(y_0, y_1) \\
\implies\quad & d(y_2, y_3) = d(f(y_1), f(y_2)) \leq \lambda d(y_1, y_2) \leq \lambda^2 d(y_0, y_1) \\
& \qquad\vdots \\
\implies\quad & d(y_j, y_{j+1}) = d(f(y_{j-1}), f(y_j)) \leq \lambda^j d(y_0, y_1), \qquad j \in \mathbb{Z}_{>0}.
\end{aligned}$$

Therefore, using the triangle inequality, for $k, l \in \mathbb{Z}_{>0}$ with $l > k$,

$$d(y_k, y_l) \leq d(y_k, y_{k+1}) + \cdots + d(y_{l-1}, y_l) \leq (\lambda^k + \cdots + \lambda^{l-1}) d(y_0, y_1).$$

Now, by Example I-2.4.2–1 the series $\sum_{j=1}^{\infty} \lambda^j$ converges. Thus the corresponding sequence of partial sums is Cauchy and so there exists $N \in \mathbb{Z}_{>0}$ sufficiently large that

$$d(y_0, y_1) \sum_{j=k}^{l-1} \lambda^j < \epsilon, \qquad k, l \geq N,\ l > k.$$

Then, for $k, l \geq N$ with $l > k$ we have $d(y_k, y_l) < \epsilon$, giving the sequence $(y_j)_{j\in\mathbb{Z}_{>0}}$ as a Cauchy sequence, as desired. Since S is complete there exists $x_0 \in S$ such that $(y_j)_{j\in\mathbb{Z}_{>0}}$ converges to x_0. We claim that $f(x_0) = x_0$. For $\epsilon \in \mathbb{R}_{>0}$ let $j \in \mathbb{Z}_{>0}$ be sufficiently large that $d(x_0, y_j) < \frac{\epsilon}{2(1+\lambda)}$ and such that $\lambda^j d(y_0, y_1) < \frac{\epsilon}{2}$. Then

$$\begin{aligned}
d(x_0, f(x_0)) &\leq d(x_0, y_j) + d(y_j, f(y_j)) + d(f(y_j), f(x_0)) \\
&\leq (1 + \lambda) d(x_0, y_j) + \lambda^j d(y_0, y_1) < \epsilon.
\end{aligned}$$

Thus $d(x_0, f(x_0)) = 0$ and so $f(x_0) = x_0$. This gives the existence part of the theorem.

For uniqueness, suppose that $\tilde{x}_0$ has the property that $f(\tilde{x}_0) = \tilde{x}_0$. Then

$$d(x_0, \tilde{x}_0) = d(f(x_0), f(\tilde{x}_0)) \leq \lambda d(x_0, \tilde{x}_0) < d(x_0, \tilde{x}_0).$$

Therefore, $d(x_0, \tilde{x}_0) = 0$ and so $x_0 = \tilde{x}_0$. ■

1.1.5 Semimetric spaces

In this section we define a concept that is weaker in terms of its structure than a metric. It is not immediately clear why this extra flexibility might be useful, but we shall see in Section 6.2.2 that it is essential in characterising locally convex topological vector spaces.

As we see, a semimetrics is a metric, *sans* the requirement of definiteness.

1.1.24 Definition (Semimetric space) A ***semimetric*** on a set S is a map $d\colon S \times S \to \mathbb{R}$ with the following properties:

(i) $d(x_1, x_2) = d(x_2, x_1)$ for all $x_1, x_2 \in S$ (***symmetry***);

(ii) $d(x_1, x_2) = 0$ if $x_1 = x_2$ (***semidefiniteness***);

(iii) $d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3)$ for all $x_1, x_2, x_3 \in S$ (***triangle inequality***).

A ***semimetric space*** is a pair (S, d) where d is a semimetric on the set S. •

Similarly as we did with metrics, the triangle inequality gives $d(x_1, x_2) \leq 2d(x_1, x_2)$ for all $x_1, x_2 \in S$, so that $d(x_1, x_2) \in \mathbb{R}_{\geq 0}$ for all $x_1, x_2 \in S$.

Let us give some examples of semimetric spaces.

1.1.25 Examples (Semimetric spaces)

1. Clearly, every metric space is also a semimetric space.
2. On $\mathbb{C}$ define a semimetric d by $d(z_1, z_2) = |\mathrm{Re}(z_1 - z_2)|$. It is evident that d is a semimetric. However, it is not a metric since if z_1 and z_2 are any two complex numbers with equal real part, then $d(z_1, z_2) = 0$. In a metric space, the distance between two distinct points should be nonzero.
3. If S is a set, we define the semimetric d_{triv} by $d_{\mathrm{triv}}(x_1, x_2) = 0$ for all $x_1, x_2 \in S$. This is clearly a semimetric, and, equally clearly, it is not a very useful one. The notation suggests that the topology defined by the semimetric is the so-called "trivial topology." We call d_{triv} the ***trivial semimetric***. •

In a semimetric space, it is still possible to define notions of open and closed sets.

1.1.26 Definition (Open and closed sets in semimetric spaces) Let (S, d) be a semimetric space.

(i) The ***open* d*-ball*** of radius r about $x \in S$ is the set

$$\mathsf{B}_d(r, x) = \{y \in S \mid d(x, y) < r\}.$$

(ii) A subset $U \subseteq S$ is ***open*** if, for each $x \in U$, there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}_d(\epsilon, x)$ such that $\mathsf{B}_d(\epsilon, x) \subseteq U$. (The empty set is also open, by declaration.)

(iii) A subset $A \subseteq S$ is ***closed*** if $S \setminus A$ is closed. •

These definitions read exactly like those for metric spaces. However, the interpretations are slightly different, as the following examples show.

1.1.27 Examples (Example 1.1.25 cont'd)

1. For $\mathbb{C}$ equipped with the semimetric $d(z_1, z_2) = |\mathrm{Re}(z_1 - z_2)|$, an open d-ball is a vertical strip of the form $\{z \in \mathbb{C} \mid \mathrm{Re}(z) \in (a, b)\}$.
2. For the semimetric space (S, d_{triv}), there are two open sets, the empty set (by definition) and S itself. Indeed, note that, for any $x \in S$ and $r \in \mathbb{R}_{>0}$, we have $\mathsf{B}_d(r, x) = S$ by definition. This then implies that S is the only nonempty open subset of S. •

For semimetric spaces, one has notions of continuity and convergence of sequences. Indeed, if (S_1, d_1) and (S_2, d_2) are semimetric spaces, then a map $f: S_1 \to S_2$ is ***continuous at $\mathbf{x_0 \in S_1}$*** if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that $d_1(x, x_0) < \delta$ implies that $d_2(f(x), f(x_0)) < \epsilon$. The map f is ***continuous*** if it is continuous at each point $x \in S_1$. A sequence $(x_j)_{j \in \mathbb{Z}_{>0}}$ in a semimetric space (S, d) ***converges*** to $x_0 \in S$ if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $d(x_j, x_0) < \epsilon$ for all $j \geq N$. The reader is invited to explore some of the consequences of these definitions in Exercises 1.1.11 and 1.1.12.

The main result we wish to record in this section is the following one, which associates a metric space to every semimetric space.

1.1.28 Theorem (Metric spaces from semimetric spaces) *Let* (S, d) *be a semimetric space and define an equivalence relation in* S *by* $\mathrm{x}_1 \sim \mathrm{x}_2$ *if* $\mathrm{d}(\mathrm{x}_1, \mathrm{x}_2) = 0$ *(the reader can show this is an equivalence relation in Exercise 1.1.13). Let* $([\mathrm{S}], [\mathrm{d}])$ *be defined by letting* $[\mathrm{S}]$ *be the set of equivalence classes in* S*, and by letting* $[\mathrm{d}]([\mathrm{x}_1], [\mathrm{x}_2]) = \mathrm{d}(\mathrm{x}_1, \mathrm{x}_2)$*. Then* $([\mathrm{S}], [\mathrm{d}])$ *is a metric space.*

Proof First we show that $[d]$ is well-defined, meaning that it is independent of choices of representatives. Let $x_1 \sim y_1$ and $x_2 \sim y_2$. Then

$$d(x_1, x_2) \leq d(x_1, y_1) + d(y_1, y_2) + d(y_2, x_2) \leq d(y_1, y_2),$$

using the triangle inequality. Similarly one shows that $d(y_1, y_2) \leq d(x_1, x_2)$, so that $d(x_1, x_2) = d(y_1, y_2)$. Therefore, $[d]$ is indeed well-defined. Now we show that $[d]$ is a metric on $[S]$. It is clear that $[d]$ is symmetric and satisfies the triangle inequality, since d has these properties. We need only show that $[d]$ is definite. Suppose that $[d]([x_1], [x_2]) = 0$. Then $d(x_1, x_2) = 0$ so that $[x_1] = [x_2]$, as desired. ■

Let us give some examples of this construction of a metric space from a semimetric space.

1.1.29 Examples (Example 1.1.25 cont'd)

1. For $\mathbb{C}$ with the semimetric $d(z_1, z_2) = |\mathrm{Re}(z_1 - z_2)|$, we note that two points z_1 and z_2 are equivalent under the relation of Theorem 1.1.28 exactly if $\mathrm{Re}(z_1) = \mathrm{Re}(z_2)$. Therefore, points in $[\mathbb{C}]$ are simply subsets of points in $\mathbb{C}$ with equal real parts, i.e., vertical lines. Then $[d]([z_1], [z_2])$ is simply the horizontal distance between the vertical lines through z_1 and z_2.
2. In the case of (S, d_{triv}), $[S]$ consists of one point since all points in S are equivalent under the equivalence relation of Theorem 1.1.28. In *any* metric space consisting of only one point, the metric must be zero (why?). That is, $[d]([x_1], [x_2]) = 0$. •

1.1.6 Completeness

In the development of Section I-2.1 we saw that the property of completeness of the real numbers was an important one, giving us such familiar conclusions as "every bounded monotonically increasing sequences converges." The notion of completeness is valid for general metric spaces, and in this section we introduce the notion and some ideas attached to it. Since we do not yet have at hand many profound examples of metric spaces (other than $\mathbb{R}$, which we studied in detail in Chapter I-2), it is difficult to explain the importance of completeness here. However, the concept is extremely important, and we say a few words about this in the context of normed vector spaces in Section 3.3.2.

The following definition gets things started, and is a natural one given the constructions of Section I-2.1.

1.1.30 Definition (Complete metric space) A metric space (S, d) is

(i) ***complete*** if every Cauchy sequence in S converges, and is

(ii) ***incomplete*** if it is not complete. •

In Proposition 1.1.15 we showed that convergent sequences are Cauchy. The converse of Proposition 1.1.15 is not generally true.

1.1.31 Examples (Cauchy sequences may not converge)

1. Let $S = \mathbb{Q}$ and let $d_{\mathbb{Q}}$ be the metric on S inherited from the metric $d_{\mathbb{R}}$ on $\mathbb{R}$. In Example I-2.1.15 we shows that there are Cauchy sequences in $\mathbb{Q}$ that do not converge. Actually, we saw that there are many such nonconvergent Cauchy sequences. Thus $(\mathbb{Q}, d_{\mathbb{Q}})$ is not a complete metric space.
2. Let $S = (0, 1]$ and define d_S to be the metric on S inherited from $\mathbb{R}$. In S consider the sequence $(x_j = \frac{1}{j})_{j\in\mathbb{Z}_{>0}}$. It is straightforward to check that (a) this is a Cauchy sequence in S and that (b) it converges, as a sequence in $\mathbb{R}$ to $x_0 = 0$. Therefore, this is a Cauchy sequence in S that does not converge in S.
3. Let $S = [0, 1]$ and define d_S to be the metric on S inherited from $\mathbb{R}$. If $(x_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence, we claim that this sequence is, in fact, convergent in S. To prove this, one can reproduce the steps used above in proving that $\mathbb{R}$ is complete, since there we began by first showing that a Cauchy sequence in $\mathbb{R}$ is contained in some closed interval. (The reader would benefit from understanding why this same idea does not work in the case where $S = (0, 1]$.) •

1.1.32 Remark (Convergence to points not in the metric space) In Examples 1.1.31–1 and 2, we showed the nonconvergence of Cauchy sequences by showing that they converge, but to a point not in the metric space. In these examples, this is sensible, since the metric spaces were themselves subsets of metric spaces. However, such examples can oversimplify the situation somewhat, and it is not uncommon to hear students, when giving the definition of a complete metric space, say, "it is complete when every Cauchy sequence in S converges in S," in cases where S is not *a priori* a subset of anything. This language is not recommended in such cases, since all elements in S know is "S-ness." Indeed, a sequence in S would likely be gravely offended at even the inference that convergence elsewhere might be possible. (As we shall see in Theorem 1.1.34, perhaps these sequences are being a little over-sensitive.) •

In Examples 1.1.31–1 and 2, the reason why the subsets of $\mathbb{R}$ were not complete when equipped with the metric inherited from $\mathbb{R}$ is strongly related to their not being closed. Let us illustrate that this is the case in general. The following result relies on some properties of closed sets that we shall prove in a more general setting in Section 1.2.4.

1.1.33 Proposition (Relationship between completeness and closedness) *Let* (S, d) *be a metric space and let* $\mathrm{A} \subseteq \mathrm{S}$ *possess the metric* d_{A} *inherited from* d*. Then the following statements hold:*

(i) if (S, d) *is complete and* A *is closed, then* $(\mathrm{A}, \mathrm{d}_{\mathrm{A}})$ *is complete;*

(ii) if $(\mathrm{A}, \mathrm{d}_{\mathrm{A}})$ *is complete then* A *is closed.*

Proof (i) If $(x_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in A, then this is also a Cauchy sequence S. Thus the sequence converges to some $x \in S$. As we shall see in Corollary 1.5.3, a closed subset of a metric space contains all limits of sequences in the set; therefore $x \in A$, and so (A, d_A) is complete.

(ii) By Corollary 1.5.3, it suffices to show that if $(x_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence in A converging to $x \in S$, then $x \in A$. However, since such a sequence in a Cauchy sequence in S, it is also a Cauchy sequence in A, by definition of d_A. Therefore, it indeed converges to a point $x \in A$. ■

1.1.7 Completions of metric spaces

Examples 1.1.31–1 and 2 might lead one to think that, for a metric space that is not complete, it should be a subset of a complete metric space. The idea, intuitively, is that one "adds" to the space the points corresponding to Cauchy sequences that do not converge. Less intuitively, but more precisely, we have the proof of the following theorem.

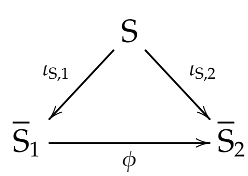
1.1.34 Theorem (Completion of metric space) *If* (S, d) *is a metric space, then there exists a complete metric space* $(\overline{\mathrm{S}}, \overline{\mathrm{d}})$ *with the following properties:*

(i) there exists an isometry $\iota_{\mathrm{S}} \colon \mathrm{S} \to \overline{\mathrm{S}}$*;*

(ii) for each $\overline{\mathrm{x}} \in \overline{\mathrm{S}}$*, there exists a sequence* $(\mathrm{x}_j)_{j\in\mathbb{Z}_{>0}}$ *for which* $(\iota_{\mathrm{S}}(\mathrm{x}_j))_{j\in\mathbb{Z}_{>0}}$ *converges to* $\overline{\mathrm{x}}$*.*

Such a metric space $(\overline{\mathrm{S}}, \overline{\mathrm{d}})$ *is a* ***completion*** *of* (S, d)*.*

Furthermore, if $(\overline{\mathrm{S}}_1, \overline{\mathrm{d}}_1)$ *and* $(\overline{\mathrm{S}}_2, \overline{\mathrm{d}}_2)$ *are two completions of* (S, d) *with* $\iota_{\mathrm{S},1} \colon \mathrm{S} \to \overline{\mathrm{S}}_1$ *and* $\iota_{\mathrm{S},2} \colon \mathrm{S} \to \overline{\mathrm{S}}_2$ *being the corresponding isometries, then there exists a bijective isometry* ϕ *from* $\overline{\mathrm{S}}_1$ *to* $\overline{\mathrm{S}}_2$ *such that the following diagram of commutes:*



Proof Let CS(S) be the collection of Cauchy sequences in S. Define a map $\tilde{d}\colon \mathrm{CS}(S) \times \mathrm{CS}(S) \to \mathbb{R}$ by

$$\tilde{d}((x_j)_{j\in\mathbb{Z}_{>0}}, (y_j)_{j\in\mathbb{Z}_{>0}}) = \lim_{j\to\infty} d(x_j, y_j).$$

To show that this definition makes sense, we must show that $\lim_{j\to\infty} d(x_j, y_j)$ exists. We do this by showing that $(d(x_j, y_j))_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence for Cauchy sequences $(x_j)_{j\in\mathbb{Z}_{>0}}$ and $(y_j)_{j\in\mathbb{Z}_{>0}}$. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be chosen such that $d(x_j, x_k) \le \frac{\epsilon}{2}$ and $d(y_j, y_k) < \frac{\epsilon}{2}$ for $j, k \ge N$. Then, by Proposition 1.1.3(ii), we have

$$|d(x_j, y_j) - d(x_k, y_k)| \le d(x_j, x_k) + d(y_j, y_k) < \epsilon$$

for all $j, k \ge N$. Thus $(d(x_j, y_j))_{j\in\mathbb{Z}_{>0}}$ is indeed a Cauchy sequence, and so convergent.

We claim that $(\mathrm{CS}(S), \tilde{d})$ is a semimetric space. The only not entirely obvious property to verify is the triangle inequality. To verify this, let $(x_j)_{j\in\mathbb{Z}_{>0}}$, $(y_j)_{j\in\mathbb{Z}_{>0}}$, and $(z_j)_{j\in\mathbb{Z}_{>0}}$ be Cauchy sequences in S. By the triangle inequality in S we have

$$d(x_j, y_j) \le d(x_j, z_j) + d(z_j, y_j), \qquad j \in \mathbb{Z}_{>0}.$$

Therefore,

$$\lim_{j\to\infty} d(x_j, y_j) \le \lim_{j\to\infty} d(x_j, z_j) + \lim_{j\to\infty} d(z_j, y_j),$$

which is the triangle inequality for $\tilde{d}$, so showing that $(\mathrm{CS}(S), \tilde{d})$ is a semimetric space.

We now let $(\overline{S}, \overline{d})$ be the metric space associated with $(\mathrm{CS}(S), \tilde{d})$ as in Theorem 1.1.28. We let $\pi_S\colon \mathrm{CS}(S) \to \overline{S}$ be the map assigning to a Cauchy sequence its equivalence class in $\overline{S}$. We then define a map $\iota_S\colon S \to \overline{S}$ by $\iota_S(x) = \pi_S((s_j(x))_{j\in\mathbb{Z}_{>0}})$, where $(s_j(x))_{j\in\mathbb{Z}_{>0}}$ is the constant sequence $s_j(x) = x$, $j \in \mathbb{Z}_{>0}$. It is evident that ι_S is an isometry since, for $x, y \in S$, we have

$$\overline{d}(\iota_S(x), \iota_S(y)) = \tilde{d}((s_j(x))_{j\in\mathbb{Z}_{>0}}, (s_j(y))_{j\in\mathbb{Z}_{>0}}) = \lim_{j\to\infty} d(s_j(x), s_j(y)) = d(x, y).$$

Next we show that, for each $\overline{x} \in \overline{S}$, there exists a sequence of the form $(\iota_S(x_j))_{j\in\mathbb{Z}_{>0}}$ converging to $\overline{x}$. Let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in S for which $\pi_S((x_j)_{j\in\mathbb{Z}_{>0}}) = \overline{x}$. We claim that $(\iota_S(x_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\overline{x}$. Indeed, let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ satisfy $d(x_j, x_k) < \epsilon$ for all $j, k \ge N$. Then

$$\overline{d}(\iota_S(x_j), \overline{x}) = \lim_{k\to\infty} d(x_j, x_k) < \epsilon$$

for $j \ge N$. Now we show that $(\overline{S}, \overline{d})$ is complete. Let $(\overline{x}_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $(\overline{S}, \overline{d})$. For each $j \in \mathbb{Z}_{>0}$, define $x_j \in S$ such that $\overline{d}(\iota_S(x_j), \overline{x}_j) < \frac{1}{j}$, this being possible since for every point in $\overline{S}$ there is a sequence in image(ι_S) converging to that point. We claim that $(x_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in S and that $(\overline{x}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $\pi_S((x_j)_{j\in\mathbb{Z}_{>0}})$. For the first claim, let $\epsilon \in \mathbb{R}_{>0}$ and choose $N \in \mathbb{Z}_{>0}$ such that $\overline{d}(\overline{x}_j, \overline{x}_k) < \epsilon$ for $j, k \ge N$, and such that $\frac{1}{N} < \epsilon$. Then

$$\overline{d}(\iota_S(x_j), \iota_S(x_k)) \le \overline{d}(\iota_S(x_j), \overline{x}_j) + \overline{d}(\overline{x}_j, \overline{x}_k) + \overline{d}(\overline{x}_k, \iota_S(x_k)) < \epsilon,$$

for $j,k \geq N$. Thus $(x_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence. Now we show that $(\overline{x}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $\pi_S((x_j)_{j\in\mathbb{Z}_{>0}})$. Let $\epsilon \in \mathbb{R}_{>0}$ and choose $N \in \mathbb{Z}_{>0}$ such that $d(x_j, x_k) < \frac{\epsilon}{3}$ for $j,k \geq N$ and such that $\frac{1}{N} < \frac{\epsilon}{3}$. Then

$$\begin{aligned}\overline{d}(\overline{x}_j, \pi_S((x_j)_{j\in\mathbb{Z}_{>0}})) &\leq \overline{d}(\overline{x}_j, \iota_S(x_j)) + \overline{d}(\iota_S(x_j), \pi_S((x_k)_{k\in\mathbb{Z}_{>0}})) \\ &\leq \tfrac{1}{j} + \lim_{k\to\infty} d(x_j, x_k) < \epsilon,\end{aligned}$$

provided that $j \geq N$. This gives completeness of $(\overline{S}, \overline{d})$.

Let us now prove the final assertion in the theorem. Let $\overline{x}_1 \in \overline{S}_1$ and let $(x_j)_{j\in\mathbb{Z}_{>0}} \in \mathrm{CS}(S)$ have the property that $(\iota_{S,1}(x_j))_{j\in\mathbb{Z}_{>0}})$ converges to $\overline{x}_1$. Note that $(\iota_{S,2}(x_j))_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\overline{S}_2$ since $\iota_{S,2}$ is an isometry. Therefore, since $(\overline{S}_2, \overline{d}_2)$ is complete, this sequence converges to some point $\overline{x}_2 \in \overline{S}_2$. Define $\phi(\overline{x}_1) = \overline{x}_2$, so defining a map $\phi\colon \overline{S}_1 \to \overline{S}_2$. To show that ϕ is well-defined, we need to show that $\overline{x}_2$ is independent of the choice of Cauchy sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ for which $(\iota_{S,1}(x_j))_{j\in\mathbb{Z}_{>0}})$ converges to $\overline{x}_1$. So let $(y_j)_{j\in\mathbb{Z}_{>0}}$ be another Cauchy sequence having this property. Let $\epsilon \in \mathbb{R}_{>0}$ and choose $N \in \mathbb{Z}_{>0}$ such that $\overline{d}_1(\iota_{S,1}(x_j), \overline{x}_1) < \frac{\epsilon}{2}$ and $\overline{d}_1(\overline{x}_1, \iota_{S,1}(y_j)) < \frac{\epsilon}{2}$. Then, since $\iota_{S,1}$ is an isometry,

$$d(x_j, y_j) = \overline{d}_1(\iota_{S,1}(x_j), \iota_{S,1}(y_j)) \leq \overline{d}_1(\iota_{S,1}(x_j), \overline{x}_1) + \overline{d}_1(\overline{x}_1, \iota_{S,1}(y_j)) < \epsilon$$

for $j \geq N$. Therefore, $\lim_{j\to\infty} d(x_j, y_j) = 0$. Now, again let $\epsilon \in \mathbb{R}_{>0}$ and choose $N \in \mathbb{Z}_{>0}$ such that $\overline{d}_2(\iota_{S,2}(y_j), \iota_{S,2}(x_j)) < \frac{\epsilon}{2}$ and $\overline{d}_2(\iota_{S,2}(x_j), \overline{x}_2) < \frac{\epsilon}{2}$ for $j \geq N$, this being possible since $\iota_{S,2}$ is an isometry. Then

$$\overline{d}_2(\iota_{S,2}(y_j), \overline{x}_2) \leq \overline{d}_2(\iota_{S,2}(y_j), \iota_{S,2}(x_j)) + \overline{d}_2(\iota_{S,2}(x_j), \overline{x}_2) < \epsilon,$$

provided that $j \geq N$. Thus $(\iota_{S,2}(y_j))_{j\in\mathbb{Z}_{>0}}$ also converges to $\overline{x}_2$, and so ϕ is well-defined.

We next need to show that ϕ has the properties of the map asserted in the theorem statement. It is clear by construction that $\iota_{S,2} = \phi \circ \iota_{S,1}$. We now show that ϕ is an isometry. Let $\overline{x}_1, \overline{y}_1 \in \overline{S}_1$ and let $(x_j)_{j\in\mathbb{Z}_{>0}}$ and $(y_j)_{j\in\mathbb{Z}_{>0}}$ be sequences in S for which the sequences $(\iota_{S,1}(x_j))_{j\in\mathbb{Z}_{>0}}$ and $(\iota_{S,1}(y_j))_{j\in\mathbb{Z}_{>0}}$ converge to $\overline{x}_1$ and $\overline{y}_1$, respectively. Since $\iota_{S,1}$ is an isometry, the sequences $(x_j)_{j\in\mathbb{Z}_{>0}}$ and $(y_j)_{j\in\mathbb{Z}_{>0}}$ are Cauchy. Since $\iota_{S,2}$ is an isometry, the sequences $(\iota_{S,2}(x_j))_{j\in\mathbb{Z}_{>0}}$ and $(\iota_{S,2}(y_j))_{j\in\mathbb{Z}_{>0}}$ are Cauchy, and so converge to points $\overline{x}_2$ and $\overline{y}_2$ in $\overline{S}_2$, respectively. We claim that

$$\lim_{j\to\infty} \overline{d}_1(\iota_{S,1}(x_j), \iota_{S,1}(y_j)) = \overline{d}_1(\overline{x}_1, \overline{y}_1), \quad \lim_{j\to\infty} \overline{d}_2(\iota_{S,2}(x_j), \iota_{S,2}(y_j)) = \overline{d}_2(\overline{x}_2, \overline{y}_2). \tag{1.1}$$

Indeed, let $\epsilon \in \mathbb{R}_{>0}$ and choose $N \in \mathbb{Z}_{>0}$ such that $\overline{d}(\iota_{S,1}(x_j), \overline{x}_1) < \frac{\epsilon}{2}$ and $\overline{d}_1(\iota_{S,1}(y_j), \overline{y}_1) < \frac{\epsilon}{2}$ for all $j \geq N$. Then

$$\left|\overline{d}_1(\iota_{S,1}(x_j), \iota_{S,1}(y_j)) - \overline{d}_1(\overline{x}_1, \overline{y}_1)\right| \leq \overline{d}(\iota_{S,1}(x_j), \overline{x}_1) + \overline{d}_1(\iota_{S,1}(y_j), \overline{y}_1) < \epsilon,$$

using Proposition 1.1.3(ii). Thus the first of equations (1.1) holds, and the second holds in exactly the same manner. Now we note that, since $\iota_{S,1}$ and $\iota_{S,2}$ are isometries,

$$\overline{d}_1(\iota_{S,1}(x_j), \iota_{S,1}(y_j)) = \overline{d}_2(\iota_{S,2}(x_j), \iota_{S,2}(y_j)) = d(x_j, y_j),$$

and therefore, from (1.1), $\overline{d}_1(\overline{x}_1, \overline{y}_1) = \overline{d}_2(\overline{x}_2, \overline{y}_2)$. By construction, we note that $\overline{x}_2 = \phi(\overline{x}_1)$ and $\overline{y}_2 = \phi(\overline{y}_1)$, and so it follows that ϕ is an isometry. Finally, ϕ we note that ϕ is injective as in Exercise 1.1.6. It is also clear that ϕ is surjective. Indeed, let $\overline{x}_2 \in \overline{S}_2$. Then there exists a sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ such that $(\iota_{S,2}(x_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\overline{x}_2$. Since $\iota_{S,2}$ is an isometry, $(x_j)_{j\in\mathbb{Z}_{>0}}$ is Cauchy. Thus, since $\iota_{S,1}$ is an isometry, $(\iota_{S,1}(x_j))_{j\in\mathbb{Z}_{>0}}$ converges to a point $\overline{x}_1 \in \overline{S}_1$. By definition of ϕ, $\phi(\overline{x}_1) = \overline{x}_2$. ■

1.1.35 Remarks (Properties of the completion)

more specific?

1. As we shall see in Section 1.2.4, the condition (ii) of Theorem 1.1.34 is tantamount to saying that $\iota_S(S)$ is a dense subset of $\overline{S}$. The final assertion in Theorem 1.1.34 concerning the uniqueness of the completion can be interpreted, in the language of Section 1.3 as saying that the completion of a metric space is unique up to an isometric homeomorphism.

more precise?

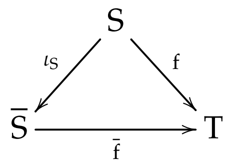
2. The construction given for the completion in the proof of Theorem 1.1.34 realises the completion using the collection of Cauchy sequences. This is certainly a natural construction, given the definition of completeness. However, there is another common construction of the completion that we ask the reader to give in Exercise 1.1.14. •

1.1.36 Examples (Completions of metric spaces)

1. *Example 1.1.31–1 cont'd:* We have seen that $(\mathbb{Q}, d_{\mathbb{Q}})$ is not a complete metric space. Moreover, our exertions of Section I-2.1 show that $(\mathbb{R}, d_{\mathbb{R}})$ is the completion of $(\mathbb{Q}, d_{\mathbb{Q}})$. Indeed, the constructions of Section I-2.1 are a special case of the general proof of Theorem 1.1.34. In Section I-2.1 we showed more than just that $\mathbb{R}$ is the completion of $\mathbb{Q}$, but also that the arithmetic properties of $\mathbb{Q}$ carry over to $\mathbb{R}$.
2. *Example 1.1.31–2 cont'd:* The completion of $(0, 1]$, with the metric inherited from $\mathbb{R}$, is clearly $[0, 1]$, with it too being equipped with the metric inherited from $\mathbb{R}$. To see this, one need only show that $[0, 1]$ has the properties of a completion, and this is easily done. •

The following property of the completion of a metric space is often useful, and also sometimes is used as a means of proving the existence of the completion. The following description of the completion gives what is in mathematics known as a "universal property" of the completion. This means that the completion is *defined* by the abstract property of the existence and uniqueness of a map into another metric space.

1.1.37 Theorem (Property of completion) *Let* (S, d) *be a metric space with* $(\overline{\mathsf{S}}, \overline{\mathsf{d}})$ *a completion of* (S, d) *and* $\iota_{\mathsf{S}} \colon \mathsf{S} \to \overline{\mathsf{S}}$ *the isometry of Theorem 1.1.34. Then, for each complete metric space* $(\mathsf{T}, \mathsf{d}_{\mathsf{T}})$ *and for each uniformly continuous map* $\mathsf{f} \colon \mathsf{S} \to \mathsf{T}$*, there exists a unique* *uniformly continuous map* $\bar{\mathrm{f}}\colon \bar{\mathrm{S}} \to \mathrm{T}$ *for which the diagram*



commutes.

Proof Let $f\colon S \to T$ be uniformly continuous. Let $\overline{x} \in \overline{S}$ and let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in S for which $(\iota_S(x_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\overline{x}$. Thus $(\iota_S(x_j))_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence. We claim that $(f(x_j))_{j\in\mathbb{Z}_{>0}}$ is also a Cauchy sequence. Let $\epsilon \in \mathbb{R}_{>0}$, choose $\delta \in \mathbb{R}_{>0}$ such that if $d(x,y) < \delta$ then $d_T(f(x), f(y)) < \epsilon$, and choose $N \in \mathbb{Z}_{>0}$ such that $d(x_j, x_k) < \delta$ for $j, k \geq N$. Then

$$d(x_j, x_k) < \delta \quad \Longrightarrow \quad d_T(f(x_j), f(x_k)) < \epsilon$$

for $j, k \geq N$. Since T is complete, the sequence $(f(x_j))_{j\in\mathbb{Z}_{>0}}$ converges to a point in T that we denote by $\overline{f}(\overline{x})$. We claim that $\overline{f}(\overline{x})$ is independent of the Cauchy sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$. To see this, let $(y_j)_{j\in\mathbb{Z}_{>0}}$ be another Cauchy sequence for which $(\iota_S(y_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\overline{x}$. Let $\epsilon \in \mathbb{R}_{>0}$. Then choose $\delta \in \mathbb{R}_{>0}$ such that if $d(x,y) < \delta$ then $d_T(f(x), f(y)) < \frac{\epsilon}{2}$, and choose $N \in \mathbb{Z}_{>0}$ such that $d(y_j, x_k) < \delta$ for $j, k \geq N$ and such that $d_T(f(x_k), \overline{f}(\overline{x})) < \frac{\epsilon}{2}$ for $k \geq N$. Then

$$d_T(f(y_j), \overline{f}(\overline{x})) \leq d_T(f(y_j), f(x_k)) + d_T(f(x_k), \overline{f}(\overline{x})) < \epsilon$$

for $j, k \geq N$, so giving convergence of $(f(y_j))_{j\in\mathbb{Z}_{>0}}$ to $\overline{f}(\overline{x})$. Thus this defines a map $\overline{f}\colon \overline{S} \to T$.

We claim that $\overline{f}$ is uniformly continuous. Let $\epsilon \in \mathbb{R}_{>0}$ and choose $\delta \in \mathbb{R}_{>0}$ such that $d(x,y) < \delta$ implies that $d_T(f(x), f(y)) < \frac{\epsilon}{2}$. Now let $\overline{x}, \overline{y} \in \overline{S}$ satisfy $\overline{d}(\overline{x}, \overline{y}) < \delta$ and let $(x_j)_{j\in\mathbb{Z}_{>0}}$ and $(y_j)_{j\in\mathbb{Z}_{>0}}$ be sequences in S for which $(\iota_S(x_j))_{j\in\mathbb{Z}_{>0}}$ and $(\iota_S(y_j))_{j\in\mathbb{Z}_{>0}}$ converge to $\overline{x}$ and $\overline{y}$, respectively. By Exercise 1.1.10 $\overline{d}(\overline{x}, \overline{y}) = \lim_{j\to\infty} d(x_j, y_j)$. We claim that

$$d_T(\overline{f}(\overline{x}), \overline{f}(\overline{y})) = \lim_{j\to\infty} d_T(f(x_j), f(y_j)).$$

Indeed, let $\epsilon' \in \mathbb{R}_{>0}$ and choose $N \in \mathbb{Z}_{>0}$ such that $d_T(\overline{f}(\overline{x}), f(x_j)) < \frac{\epsilon'}{2}$ and $d_T(\overline{f}(\overline{y}), f(x_k)) < \frac{\epsilon'}{2}$ for $j \in \mathbb{Z}_{>0}$.

$$\left|d_T(\overline{f}(\overline{x}), \overline{f}(\overline{y})) - d_T(f(x_j), f(y_j))\right| \leq d_T(\overline{f}(\overline{x}), f(x_j)) + d_T(\overline{f}(\overline{y}), f(x_k)) < \epsilon'$$

for $j \geq N$. Now we have

$$\lim_{j\to\infty} d(x_j, y_j) < \delta \quad \Longrightarrow \quad \lim_{j\to\infty} d_T(f(x_j), f(y_j)) \leq \tfrac{\epsilon}{2} < \epsilon,$$

or $\overline{d}(\overline{x}, \overline{y}) < \delta$ implies $d_T(\overline{f}(\overline{x}), \overline{f}(\overline{y}))$, as desired.

Finally, for this part of the proof, we need to show that $\overline{f}$ is the only map having the properties given in the theorem statement. Suppose that $\tilde{f}\colon \overline{S} \to T$ is a uniformly continuous map for which $\tilde{f} \circ \iota_S = f$. Let $\overline{x} \in \overline{S}$ and let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in S for which $(\iota_S(x_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\overline{x}$. Note that $f(x_j) = \overline{f}(\iota_S(x_j)) = \tilde{f}(\iota_S(x_j))$ for all $j \in \mathbb{Z}_{>0}$. Now let $\epsilon \in \mathbb{R}_{>0}$. Since $\overline{f}$ and $\tilde{f}$ are uniformly continuous, choose $\delta \in \mathbb{R}_{>0}$ such that, if $\overline{d}(x, y) < \delta$, then $d_T(\overline{f}(x), \overline{f}(\overline{y})) < \frac{\epsilon}{2}$ and $d_T(\tilde{f}(x), \tilde{f}(\overline{y})) < \frac{\epsilon}{2}$. Now let $N \in \mathbb{Z}_{>0}$ satisfy $\overline{d}(\overline{x}, x_j) < \delta$ for $j \geq N$. Then

$$d_T(\overline{f}(\overline{x}), \tilde{f}(\overline{x})) \leq d_T(\overline{f}(\overline{x}), f(x_j)) + d_T(f(x_j), \tilde{f}(\overline{x})) < \epsilon$$

for $j \geq N$. Thus $\tilde{f}(\overline{x}) = \overline{f}(\overline{x})$, as desired. ■

The requirement that f be uniformly continuous is necessary.

1.1.38 Example (A continuous map not extending to the completion) Let $S = (0, 1]$ with d the standard metric induced from $\mathbb{R}$. The completion of S is $\overline{S} = [0, 1]$ and the induced metric on $\operatorname{cl} S$ is the standard metric induced from $\mathbb{R}$. The map $f: (0, 1] \to \mathbb{R}$ defined by $f(x) = \frac{1}{x}$ is continuous but not uniformly continuous. Note that there exists no continuous map $\overline{f}: \overline{S} \to \mathbb{R}$ extending f. •

1.1.8 Compact subsets of metric spaces

In Section I-2.5 we studied compact subsets of $\mathbb{R}$, giving a simple characterisation of such sets in terms of the Heine–Borel Theorem. In Section I-3.1.4 we showed that compact sets possessed many nice properties in terms of continuous functions. In this section we carry out this same programme for compact subsets of metric spaces. As we shall see, some of the properties of compact subsets of $\mathbb{R}$ carry over to metric spaces, but some do not. Thus the reader should learn to exercise the proper care when extending their intuition about compactness from $\mathbb{R}$ to metric spaces.

We begin with the definition of a cover and subcover.

1.1.39 Definition (Open cover of a subset of a metric space) Let (S, d) be a metric space and let $A \subseteq S$.

(i) An ***open cover*** for A is a family $(U_i)_{i\in I}$ of open subsets of S having the property that $A \subseteq \cup_{i\in I} U_i$.

(ii) A ***subcover*** of an open cover $(U_i)_{i\in I}$ of A is an open cover $(V_j)_{j\in J}$ of A having the property that $(V_j)_{j\in J} \subseteq (U_i)_{i\in I}$. •

With these, one has the same definitions associated with compactness as one has for subsets of $\mathbb{R}$.

1.1.40 Definition (Compact, precompact) Let (S, d) be a metric space.

(i) A subset $K \subseteq S$ is compact if every open cover $(U_i)_{i\in I}$ of A possesses a finite subcover.

(ii) A subset $A \subseteq S$ is ***precompact***[1] if $\mathrm{cl}(A)$ is compact. •

Let us explore some equivalent characterisations of compact subsets of metric spaces. We shall see in Section 1.6 that these characterisations are not generally valid for topological spaces.

1.1.41 Theorem (Characterisations of compact subsets of metric spaces) *If* (S, d) *is a metric space then the following statements are equivalent for a subset* $\mathsf{K} \subseteq \mathsf{S}$*:*

(i) K *is compact;*

(ii) if $\mathsf{A} \subseteq \mathsf{K}$ *is infinite then there exists a point* $\mathsf{x}_0 \in \mathsf{K}$ *such that* $(\mathsf{B}_{\mathrm{d}}(\epsilon, \mathsf{x}_0) \setminus \{\mathsf{x}_0\}) \cap \mathsf{A} \neq \varnothing$ *for every* $\epsilon \in \mathbb{R}_{>0}$*;*

(iii) every sequence in S *possesses a convergent subsequence;*

(iv) K *is complete and totally bounded.*

Proof (i) $\Longrightarrow$ (ii) Suppose that A is an infinite subset of K that does not have the stated property. Then for every $x \in K$ there exists $\epsilon_x \in \mathbb{R}_{>0}$ such that $(\mathsf{B}_d(\epsilon_x, x) \setminus \{x\}) \cap A = \varnothing$. Thus $(\mathsf{B}_d(\epsilon_x, x))_{x \in K}$ is an open cover of K. We claim that it has no finite subcover. Indeed, let $F \subseteq K$ be finite. For each $x \in F$ we have

$$(\mathsf{B}_d(\epsilon_x, x) \setminus \{x\}) \cap A = \varnothing \quad \Longrightarrow \quad \mathsf{B}_d(\epsilon_x, x) \cap A \subseteq \{x\}.$$

Therefore, $\cup_{x \in F} \mathsf{B}_d(\epsilon_x, x)$ can contain at most finitely many points from A, and so cannot be an open cover. Thus K is not compact.

(ii) $\Longrightarrow$ (iii) Let $(x_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence and define $A = \{x_j \mid j \in \mathbb{Z}_{>0}\}$. If $\mathrm{card}\, A < \infty$ then denote $A = \{y_1, \dots, y_k\}$ with the points $y_1, \dots, y_k$ being distinct. Thus there exists $l \in \{1, \dots, k\}$ and an increasing sequence $(k_j)_{j \in \mathbb{Z}_{>0}}$ such that $x_{k_j} = y_l$ for all $j \in \mathbb{Z}_{>0}$. The subsequence $(x_{k_j})_{j \in \mathbb{Z}_{>0}}$ is obviously convergent. On the other hand, if $\mathrm{card}(A) = \infty$ then we claim that it still has a convergent subsequence. Indeed, by assumption on A there exists $x_0 \in K$ such that $(\mathsf{B}_d(\epsilon, x_0) \setminus \{x_0\}) \cap A \neq \varnothing$ for every $\epsilon \in \mathbb{R}_{>0}$. We claim that this implies that $\mathsf{B}_d(\epsilon, x_0) \cap A$ is infinite for all $\epsilon \in \mathbb{R}_{>0}$. Indeed, suppose that $\mathsf{B}_d(\epsilon, x_0) \cap A$ is finite for some $\epsilon \in \mathbb{R}_{>0}$, and denote

$$\mathsf{B}_d(\epsilon, x_0) \cap A = \{y_1, \dots, y_k\}.$$

Let

$$\epsilon_0 = \tfrac{1}{2} \min\{d(x_0, y_l) \mid l \in \{1, \dots, k\}\}.$$

Then $\mathsf{B}_d(\epsilon_0, x_0) \cap A \subseteq \{x_0\}$ which contradicts the assumptions on A. Thus $\mathsf{B}_d(\epsilon, x_0) \cap A$ is infinite for all $\epsilon \in \mathbb{R}_{>0}$. Thus there exists a subsequence $(x_{k_j})_{j \in \mathbb{Z}_{>0}}$ of $(x_j)_{j \in \mathbb{Z}_{>0}}$ such that $d(x_0, x_{k_j}) < \frac{1}{2^j}$. This subsequence then converges to x_0.

(iii) $\Longrightarrow$ (iv) First we show that (iii) implies that K is complete. Let $(x_j)_{j \in \mathbb{Z}_{>0}}$ be a Cauchy sequence. By assumption this sequence has a convergent subsequence; let us denote this subsequence by $(x_{k_j})_{j \in \mathbb{Z}_{>0}}$ and its limit by x_0. We claim that $(x_j)_{j \in \mathbb{Z}_{>0}}$ converges to x_0. Indeed, for $\epsilon \in \mathbb{R}_{>0}$ let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $d(x_j, x_k) < \frac{\epsilon}{2}$

[1] What we call "precompact" is very often called "relatively compact." However, we shall use the term "relatively compact" for something different.

for $j, k \geq N$. Let $l \in \mathbb{Z}_{>0}$ be sufficiently large that $k_l \geq N$ and such that $d(x_{k_l}, x_0) < \frac{\epsilon}{2}$. Then

$$d(x_j, x_0) \leq d(x_j, x_{k_l}) + d(x_{k_l}, x_0) < \epsilon,$$

giving convergence, and so completeness of K.

Now suppose that K is not totally bounded. Then there exists $\epsilon \in \mathbb{R}_{>0}$ such that any covering of K by balls of radius ϵ must be infinite. Let $x_1 \in K$ and choose $x_2 \in K$ such that $d(x_2, x_1) \geq \epsilon$. This is possible since otherwise $(\mathsf{B}_d(\epsilon, x_1))$ is a finite cover of K by balls of radius ϵ. Now let $x_3 \in K$ satisfy $d(x_3, x_1), d(x_3, x_1) \geq \epsilon$. Again, this is possible since otherwise $(\mathsf{B}_d(\epsilon, x_1), \mathsf{B}_d(\epsilon, x_2))$ is a finite covering of K by balls of radius ϵ. Proceeding in this way we define a sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ for which $d(x_j, x_k) \geq \epsilon$ for $j \neq k$. This sequence cannot possibly possess a convergent subsequence and so (iii) cannot hold.

(iv) $\Longrightarrow$ (i) Suppose that K is not compact so that there exists an open cover $(U_i)_{i\in I}$ having no finite subcover. Also suppose that K is complete and totally bounded. By total boundedness, for each $k \in \mathbb{Z}_{>0}$ there exists a finite set $F_k \subseteq K$ such that $K \subseteq \cup_{x\in F_k} \mathsf{B}_d(\frac{1}{n}, x)$. Let $x_1 \in F_1$ be such that no finite subcover of $(U_i)_{i\in I}$ covers $C_1 \triangleq K \cap \overline{\mathsf{B}}_d(1, x_1)$. Note that

$$C_1 = \cup_{x\in F_2}(C_1 \cap \overline{\mathsf{B}}_d(\tfrac{1}{2}, x)$$

so that there exists $x_2 \in F_2$ for which no finite subcover of $(U_i)_{i\in I}$ covers $C_2 \triangleq C_1 \cap \overline{\mathsf{B}}_d(\frac{1}{2}, x_2)$. Continuing in this way we define a sequence $(C_j)_{j\in\mathbb{Z}_{>0}}$ of subsets defined by $C_j = C_{j-1} \cap \overline{\mathsf{B}}_d(\frac{1}{j}, x_j)$ and having the property that C_j, $j \in \mathbb{Z}_{>0}$, cannot be covered by any finite subcover of $(U_i)_{i\in I}$. Let $(y_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in K such that $y_j \in C_j$, $j \in \mathbb{Z}_{>0}$. The sequence $(y_j)_{j\in\mathbb{Z}_{>0}}$ is Cauchy. Indeed, for $\epsilon \in \mathbb{R}_{>0}$ let N be sufficiently large that $\frac{2}{N} < \epsilon$. Then, since all points in C_j are within distance $\frac{1}{j}$ of x_j and since $C_j \subseteq C_k$ for $j \geq k$, it follows that $d(x_j, x_k) < \epsilon$ for $j, k \geq N$. Since K is complete the sequence $(y_j)_{j\in\mathbb{Z}_{>0}}$ converges to $y_0 \in K$. Let $i_0 \in I$ be such that $y_0 \in U_{i_0}$. Then, since $y_0 \in C_j$ for each $j \in \mathbb{Z}_{>0}$ and since the sets C_j become arbitrarily small, there exists some $N \in \mathbb{Z}_{>0}$ for which $C_N \subseteq U_{i_0}$. This is a contradiction. ■

1.1.42 Corollary (Precompact equals totally bounded in complete metric spaces) *If* (S, d) *is a complete metric space, a subset* $\mathrm{A} \subseteq \mathrm{S}$ *is precompact if and only if it is totally bounded.*

Proof Suppose that A is precompact so that $\mathrm{cl}(A)$ is compact. By Theorem 1.1.41 it follows that $\mathrm{cl}(A)$ is totally bounded. Since a subset of a totally bounded set is totally bounded (why?), it follows that A is totally bounded. Now suppose that A is totally bounded so that $\mathrm{cl}(A)$ is complete by Proposition 1.1.33. We claim that $\mathrm{cl}(A)$ is also totally bounded. Indeed, since A is totally bounded, let $\epsilon \in \mathbb{R}_{>0}$ and let $x_1, \dots, x_k \in S$ have the property that $A \subseteq \cup_{j=1}^k \mathsf{B}_d(\epsilon, x_j)$. Let $y \in \mathrm{cl}(A)$ so that by there exists a sequence $(y_j)_{j\in\mathbb{Z}_{>0}}$ in A converging to y. ■

what

The reader can see in Exercise 1.1.19 that the corollary is not true for metric spaces that are not complete.

In Theorem I-2.5.27 we indicated that the notion of compactness is concretely characterised for subsets of $\mathbb{R}$ as being equivalent to closedness and boundedness.

Of course, one is inclined to ask whether this is generally true. It is not, although one implication is generally true.

1.1.43 Corollary (Compact sets are closed and bounded) *If* (S, d) *is a metric space and if* $K \subseteq S$ *is compact, then* K *is closed and bounded.*

Proof If K is compact then it is complete by Theorem 1.1.41 and so closed by Proposition 1.1.33. Moreover, K is totally bounded by Theorem 1.1.41 and so bounded by Exercise 1.1.17. ■

The converse of the proposition is not generally true. We shall give a fairly general and interesting setup for this in Theorem 3.6.15. For now let us give a somewhat less interesting example of how closed and bounded sets may not be compact.

1.1.44 Example (Closed and bounded sets may not be compact) Let $S = (0, 1)$ and define a metric on S by $d(x, y) = |x - y|$. Note that (S, d) is obviously closed and bounded. But it is not compact. Indeed, the open cover $((\frac{1}{j+2}, 1 - \frac{1}{j+2}))_{j \in \mathbb{Z}_{>0}}$ does not possess a finite subcover. Indeed, any open cover will be contained in an interval of the form $(\frac{1}{k}, 1 - \frac{1}{k})$ for some $k \geq 3$.

In this case the lack of compactness is directly related to the fact that (S, d) is not complete. The completion of (S, d) is $[0, 1]$ which is compact. More interesting closed and bounded sets that are not compact arise when the metric space is required to be complete. •

Finally, let us indicate that there is a relationship between compactness and continuity.

1.1.45 Proposition (Compact sets are mapped to compact sets by continuous maps) *If* (S_1, d_1) *and* (S_2, d_2) *are metric spaces, if* $f \colon S_1 \to S_2$ *is continuous, and if* $K \subseteq S_1$ *is compact, then* $f(K) \subseteq S_2$ *is compact.*

Proof Let $(U_a)_{a \in A}$ be an open cover of $\mathrm{image}(f)$. Then $(f^{-1}(U_a))_{a \in A}$ is an open cover of S_1, and so there exists a finite subset $ifama_1, \dots, a_k \subseteq A$ such that $\cup_{j=1}^{k} f^{-1}(U_{a_k}) = S_1$. It is then clear that $(f(f^{-1}(U_{a_1})), \dots, f(f^{-1}(U_{a_k})))$ covers $\mathrm{image}(f)$. Moreover, by Proposition I-1.3.5, $f^{-1}(f(U_{a_j})) \subseteq U_{a_j}$, $j \in \{1, \dots, k\}$. Thus $(U_{a_1}, \dots, U_{a_k})$ is a finite subcover of $(U_a)_{a \in A}$. ■

Exercises

1.1.1 Show that the metric of Example 1.1.2–Exercise 1.1.1 is indeed a metric.

1.1.2 If (S, d) is a metric space, show that

$$d'(x_1, x_2) = \frac{d(x_1, x_2)}{1 + d(x_1, x_2)}$$

defines a metric on S, and that $\mathrm{image}(d') \subseteq [0, 1)$.

1.1.3 Let (S, d) be a metric space, let $A \subseteq S$, and define a metric d_A on A by $d_A(x_1, x_2) = d(x_1, x_2)$. Show that a subset $U \subseteq A$ is open if and only if there exists an open subset $\tilde{U} \subseteq S$ such that $U = A \cap \tilde{U}$.

1.1.4 A countable subset $A \subseteq \mathbb{R}$ is ***discrete*** if there exists $r \in \mathbb{R}_{>0}$ such that $x_1, x_2 \in \mathbb{R}$ implies that $|x_1 - x_2| > r$. Show that the subspace topology on a discrete subset of $\mathbb{R}$ is the same as the discrete topology.

1.1.5 Let (S, d) be a metric space. Show that a subset $A \subseteq S$ is bounded if and only if, for each $x \in S$, there exists $R_x \in \mathbb{R}_{>0}$ such that $S \subseteq \mathsf{B}_d(R_x, x)$.

1.1.6 Show that an isometry is injective.

1.1.7 Show that an isometry is uniformly continuous.

1.1.8 Show that the collection of isometric bijections of a metric space is a group.

1.1.9 Show that limits of sequences in metric spaces are unique. That is, show that if the sequence $(x_j)_{j \in \mathbb{Z}_{>0}}$ converges to both x_1 and x_2, then $x_1 = x_2$.

1.1.10 If (S, d) is a metric space, if $x, y \in S$, and if $(x_j)_{j \in \mathbb{Z}_{>0}}$ and $(y_j)_{j \in \mathbb{Z}_{>0}}$ are sequences converging to x and y, respectively, show that

$$d(x, y) = \lim_{j \to \infty} d(x_j, y_j).$$

1.1.11 Limits in semimetric spaces need not be unique. To see this, answer the following two questions.

(a) Show that every sequence converges in the semimetric space (S, d_{triv}).

(b) Come up with an example of a semimetric space (S, d) for which

1. $d \neq d_{\mathrm{triv}}$ and
2. d is not a metric.

Describe the character of the converging sequences.

1.1.12 Let (T, d_{triv}) be a semimetric space with the trivial semimetric and let (S, d) be another semimetric space. Show that every map $f: S \to T$ is continuous.

1.1.13 Let (S, d) be a pseudometric space. Show that the relation "$x_1 \sim x_2$ if $d(x_1, x_2) = 0$" is an equivalence relation.

The following exercise will guide you through a construction of the completion of a metric space. This construction relies on the fact, proven as Theorem 1.9.1, that the set of bounded continuous $\mathbb{R}$-valued functions on a metric space is complete. Note that a consequence of the construction is that every metric space is embedded isometrically in a normed vector space.

finish

1.1.14

1.1.15 Show that the collection of open subsets of a semimetric space forms a topology.

1.1.16 Show that the set of open d-balls in a metric space (S, d) forms a base for the metric topology.

true for semimetric?

1.1.17 Let (S, d) be a metric space. Show that S is totally bounded if and only if there exists $x_0 \in S$ and $R \in \mathbb{R}_{>0}$ such that $d(x, x_0) < R$ for every $x \in S$.

As with any mathematical structure, with metric spaces it is interesting to understand the "invariants" that are associated with the structure. An invariant is typically some property that is independent under some class of transformation. In the following exercises you will see that "total boundedness" and "complete" are not invariants for a metric space under homeomorphism, but that the two properties together are.

move this to homeomorphism section?

1.1.18 Answer the following questions.

(a) Find a complete metric space that is homeomorphic to an incomplete metric space.

(b) Find an totally unbounded metric space that is homeomorphic to a totally bounded metric space.

(c) Show that if metric spaces (S_1, d_1) and (S_2, d_2) are homeomorphic and if (S_1, d_1) is complete and totally bounded, then (S_2, d_2) is complete and totally bounded.

1.1.19 Answer the following questions.

(a) Let (S, d) be a metric space. Show that if $A \subseteq S$ is precompact, then it is totally bounded.

(b) Consider $\mathbb{Q}$ with the metric $d(q_1, q_2) = |q_1 - q_2|$ and let $A = (0, 1) \cap \mathbb{Q}$. Show that A is totally bounded but not precompact.

Section 1.2

Topological spaces

In this section, we introduce the notion of a topological space as a family of subsets that have certain properties relative to unions and intersections. Our motivation for this is derived from Proposition 1.1.6 where properties of open subsets of metric spaces are given; see also Exercise I-2.5.1 concerning open subsets of $\mathbb{R}$. The notion of a topological space is *extremely* unstructured, but it does provide just the right structure to discuss many important ideas in almost all areas of mathematics. The reader is advised to simply take "on faith" the fact that the definition is the correct one. Experience will reveal that it does indeed serve exactly the purpose intended.

Do I need to read this section? If you have decided to read this chapter, and if you found Section 1.1 compelling, then this is indeed the place to start on the path towards learning point set topology. •

1.2.1 Definitions and simple examples

Let us begin outright with the definition.

1.2.1 Definition (Open set, topological space) A ***topology*** on a set S is a family of subsets $\mathscr{O} \subseteq \mathbf{2}^S$ having the following properties:

(i) if $(U_a)_{a\in A}$ is an arbitrary family of subsets from $\mathscr{O}$, then $\cup_{a\in A} U_a \subseteq \mathscr{O}$;

(ii) if $\{U_1, \dots, U_k\}$ is a finite family of subsets from $\mathscr{O}$, then $\cap_{j=1}^k U_j \in \mathscr{O}$;

(iii) $S \in \mathscr{O}$;

(iv) $\varnothing \in \mathscr{O}$.

The sets in $\mathscr{O}$ are called ***open sets*** and the pair $(S, \mathscr{O})$ is called a ***topological space*** if $\mathscr{O}$ is a topology on S. •

1.2.2 Remark (Why are the axioms for a topological space as they are?) It is important to always wonder why the mathematical axioms one is given are the "right ones." Sometimes the axioms seem natural (as with the vector space axioms, perhaps), and so the question does not come up so readily. However, sometimes the rationale for the axioms is not so easy to understand. Perhaps the topological space axioms fall into this class, so let us say a few words about where these axioms come from.

The best way to understand the topological space axioms is via the observation that the axioms (1) hold for open subsets of Euclidean space (see Exercise I-2.5.1 and Proposition II-1.2.19) and (2) are the only properties of open subsets of Euclidean space needed to formulate concepts like continuity (Theorems I-3.1.3

and II-1.3.2), compactness (Definitions I-2.5.26 and II-1.2.34), and connectedness (Definitions I-2.5.33 and II-1.2.44). We shall see as we go along that the basic axioms of a topological space very often need to be supplemented by additional properties—e.g., separability axioms (Section 1.8.2) or countability axioms (Section 1.8.1)—in order to prove useful theorems. But the basic axioms do capture many of the features one wishes to capture in a general setting. •

Note that $\cap_{j=1}^k U_j \in \mathscr{O}$ for every finite family $\{U_1, \dots, U_k\} \subseteq \mathscr{O}$ if and only if $U_1 \cap U_2 \in \mathscr{O}$ for every $U_1, U_2 \in \mathscr{O}$. Thus we need only show that pairwise intersections of open sets are open to show that a family of subsets defines a topology.

Let us consider some examples of topological spaces.

1.2.3 Examples (Topological spaces)

1. The set $\mathbb{R}$ with the open sets as defined in Definition I-2.5.2 is a topological space by Exercise I-2.5.1. This topology on $\mathbb{R}$ is called the ***standard topology***, and will be the topology we use unless we state otherwise.
2. The set $\mathbb{R}^n$ with the family of open sets defined as in Section II-1.2 is a topological sets. Again we call this the ***standard topology*** and use it as the topology of choice, unless otherwise stated.
3. More generally, if (S, d) is a metric or semimetric space, then the family of open subsets of S forms a topology. For metric spaces, this follows from Proposition 1.1.6, and for semimetric spaces, the reader can show this in Exercise 1.1.15. This topology is called the ***metric topology*** or the ***semimetric topology***, depending on whether d is a metric or semimetric, respectively.
4. We define a topology on the extended real numbers $\overline{\mathbb{R}}$ by declaring the open sets to be of the form

$$U, \quad U \cup [-\infty, b), \quad U \cup (a, \infty], \quad U \cup [-\infty, b) \cup (a, \infty], \qquad U \subseteq \mathbb{R} \text{ open}, \ a, b \in \mathbb{R}.$$

It is easy to check that this indeed defines a topology if one notes that the intersection of two sets of the form

$$[-\infty, b), \quad (a, \infty], \quad [-\infty, b) \cup (a, \infty]$$

is again a set of this form, and that unions of sets of this form are again sets of this form. The idea of this topology is that the subsets of $\overline{\mathbb{R}}$ of the form $[-\infty, b)$ and $(a, \infty]$ are neighbourhoods (as we will discuss in Section 1.2.3) of $-\infty$ and ∞, respectively.
5. For any set S, we define the ***trivial topology*** by $\mathscr{O}_{\mathrm{triv}} = \{\emptyset, S\}$. As we saw in Example 1.1.25–3, this topology comes from the semimetric space (S, d_{triv}).
6. For any set S, we define the ***discrete topology*** by $\mathscr{O}_{\mathrm{disc}} = \mathbf{2}^S$. By Example 1.1.2–4, this topology comes from the metric space (S, d_{disc}).

It is most common that we will encounter situations where a set comes equipped with a natural topology, and this is the topology we shall use, e.g., the standard topology for $\mathbb{R}^n$. However, there are also cases we will encounter where the same set has multiple natural topologies, and one wishes to be able to compare these in some way.

1.2.4 Definition (Weaker, stronger, coarser, finer, smaller, larger) Let S be a set and let $\mathscr{O}_1$ and $\mathscr{O}_2$ be topologies on S.

(i) The topology $\mathscr{O}_1$ is ***weaker***, ***coarser***, or ***smaller*** than the topology $\mathscr{O}_2$ if $\mathscr{O}_1 \subseteq \mathscr{O}_2$.

(ii) The topology $\mathscr{O}_1$ is ***stronger***, ***finer***, or ***larger*** than the topology $\mathscr{O}_2$ if $\mathscr{O}_2 \subseteq \mathscr{O}_1$. •

1.2.5 Remark (Understanding the grammar of comparisons of topologies) If you are slightly dyslexic like the author, then you will find it difficult to remember which word goes with which concept.

1. The expressions "smaller" and "larger" are the most easy to use since $\mathscr{O}_1$ is smaller that $\mathscr{O}_2$ (or, equivalently, $\mathscr{O}_2$ is larger than $\mathscr{O}_1$) if $\mathscr{O}_1$ has fewer open sets than $\mathscr{O}_2$.
2. The "coarser" versus "finer" terminology is also fairly intuitive. The idea is that a topology $\mathscr{O}_1$ is coarser than a topology $\mathscr{O}_2$ (or, equivalently, $\mathscr{O}_2$ is finer than $\mathscr{O}_1$) when the nonempty open sets in $\mathscr{O}_1$ are larger than those for $\mathscr{O}_2$.
3. The idea behind the terminology "weaker" and "stronger" is best understood when related to the notion of convergence in a topology which we will discuss in Section 1.5. Let us record this relationship here anyway, since this is the most appropriate time to do so; readers who need to know about convergence in topological spaces can refer ahead. The idea of a topology $\mathscr{O}_1$ being weaker than another topology $\mathscr{O}_2$ (or, equivalently, $\mathscr{O}_2$ being stronger than $\mathscr{O}_1$) is that there are more convergent sequences in the topology $\mathscr{O}_1$ than in $\mathscr{O}_2$. Thus the conditions for convergence are weaker in $\mathscr{O}_1$ than in $\mathscr{O}_2$. •

Let us give some examples that illustrate the language.

1.2.6 Examples (Weakest and strongest topologies on a set) Let S be a set.

1. The weakest, or coarsest, or smallest topology on S is the trivial topology $\mathscr{O}_{\text{triv}} = \{\emptyset, S\}$. Note that this topology obviously has the minimum number of open sets (it is the smallest topology), every sequence in S converges (it is the weakest topology), and the nonempty open sets are large (it is the coarsest topology).
2. The strongest, finest, or largest topology on S is the discrete topology $\mathscr{O}_{\text{disc}} = \mathbf{2}^S$. Note that this topology obviously has the maximum number of open sets (it is the largest topology), only sequences that are eventually constant converge (it is the strongest), and the nonempty open sets can be arbitrarily small (it is the finest topology). •

As with metric spaces, the notion of a closed set can be defined, and is important, for reasons that will become most clear when we talk about topological vector spaces in Chapter 6.

1.2.7 Definition (Closed set) Let $(S, \mathscr{O})$ be a topological space. A subset $A \subseteq S$ is ***closed*** if $S \setminus A$ is open. •

The following result records some of the basic properties of closed sets.

1.2.8 Proposition (Properties of closed sets) *If* $(S, \mathscr{O})$ *is a topological space, then the following statements hold:*

(i) *if* $\{A_1, \dots, A_k\}$ *is a finite family of closed subsets of* S*, then* $\cup_{j=1}^{k} A_j$ *is closed;*

(ii) *if* $(A_b)_{b \in B}$ *is an arbitrary family of closed subsets of* S*, then* $\cap_{b \in B} A_b$ *is closed;*

(iii) S *is closed;*

(iv) $\varnothing$ *is closed.*

Proof Parts (iii) and (iv) are trivial, and parts (i) and (ii) follow from De Morgan's Laws, Proposition I-1.1.5. ■

1.2.9 Remark (Closed sets can define a topology) It is possible to define a topology by defining, not open sets as we have done and as is usually done, but by defining closed sets as having the properties of Proposition 1.2.8. One then can define an open set to be one whose complement is open. Having done this, one can then proceed to prove that open sets have the properties of Definition 1.2.1. •

1.2.10 Example (Topology defined using closed sets) If S is a set, denote by $\mathscr{C}_{\mathrm{fin}}$ the family of finite subsets of S, along with S and the empty set. One can verify that the family $\mathscr{C}_{\mathrm{fin}}$ satisfies the properties of closed sets given in Proposition 1.2.8. Thus, if we define an open set to be one whose complement lies in $\mathscr{C}_{\mathrm{fin}}$, these open sets form a topology. •

1.2.2 Bases and subbases

When specifying the open sets in a topological space, it is often convenient to specify a smaller, more easily describable family of sets. This motivates the following definition.

1.2.11 Definition (Base for a topology) A ***base*** for a topology $\mathscr{O}$ on a set S is a subset $\mathscr{B} \subseteq \mathscr{O}$ with the property that, if $U \in \mathscr{O}$, then $U = \cup_{a \in A} B_a$ for a family of subsets $(B_a)_{a \in A} \subseteq \mathscr{B}$. A topology is said to be ***generated*** by a base for it. •

The following is a useful criterion for recognising when a certain family of open sets forms a base for a topology.

1.2.12 Proposition (Characterisation of bases) *For a topological space* $(S, \mathscr{O})$*, a family of subsets* $\mathscr{B} \subseteq \mathscr{O}$ *is a base for* $\mathscr{O}$ *if and only if, for each* $U \in \mathscr{O}$ *and for each* $x \in U$*, there exists* $B \in \mathscr{B}$ *such that* $x \in B$ *and* $B \subseteq U$*.*

Proof First suppose that $\mathscr{B}$ is a base for $\mathscr{O}$, let $U \in \mathscr{O}$ and let $x \in U$. Since U is a union of set from $\mathscr{B}$, say $U = \cup_{a \in A} B_a$, we have $x \in B_a$ and $B_a \subseteq U$ for some $a \in A$.

Conversely, suppose that for each $U \in \mathscr{O}$ and $x \in U$ there exists $B \in \mathscr{B}$ such that $x \in B$ and $B \subseteq U$. Now choose some $U \in \mathscr{O}$, and for each $x \in U$ let $U_x \in \mathscr{O}$ satisfy $x \in U_x$. By assumption, there exists $B_x \in \mathscr{B}$ such that $x \in B_x$ and $B_x \subseteq U_x$. We the have $U = \cup_{x \in U} B_x$, implying that $\mathscr{B}$ is a base for $\mathscr{O}$. ■

Let us give some examples of bases for some simple topologies.

1.2.13 Examples (Bases for topologies)

1. By Proposition I-2.5.6 the family of open intervals is a base for the standard topology on $\mathbb{R}$.
2. The standard topology on $\mathbb{R}^n$ has the family of open balls as a base for its topology. This is a special case of the next example.
3. Let (S, d) be a metric space. One can easily show that the set of open d-balls, along with the empty set, forms a base for the metric topology. The reader is asked to prove this in Exercise 1.1.16. As a consequence of this, the standard topology of $\mathbb{R}^n$ is generated by open balls, and the topology of $\mathbb{C}$ is generated by open disks.
4. Let $(S, \mathscr{O}_{\text{disc}})$ be a topological space with the discrete topology. Then family of singletons, $\mathscr{B} = \{\{x\} \mid x \in S\}$, along with the empty set, forms a base for the topology. Moreover, it is easy to see that any other base for the discrete topology must contain $\mathscr{B}$.
5. Let $(S, \mathscr{O}_{\text{triv}})$ be a topological space with the trivial topology. Then the only base for the topology is $\mathscr{B} = \{\emptyset, S\}$. •

It is sometimes useful to know when, given a family of subsets of a set S, the family is a base for *some* topology on S.

1.2.14 Proposition (When is a family of subsets a base?) *A family* $\mathscr{B}$ *of subsets of a set* S *is a base for a topology on* S *if and only if*

(i) $\emptyset \in \mathscr{B}$*,*

(ii) $S = \cup_{B \in \mathscr{B}} B$*, and*

(iii) for every finite family $\{B_1, \dots, B_k\}$ *of subsets from* $\mathscr{B}$*, and for every* $x \in \cap_{j=1}^k B_j$*, there exists* $B \in \mathscr{B}$ *such that* $x \in B$ *and* $B \subseteq \cap_{j=1}^k B_j$*.*

Proof It is clear from Proposition 1.2.12 that if $\mathscr{B}$ is a base for some topology then it satisfies conditions (i)–(iii). So suppose that $\mathscr{B}$ satisfies these three conditions. Let $\mathscr{O}$ be the family of subsets of S that are unions of elements of $\mathscr{B}$. That is to say, an element of $U \in \mathscr{O}$ has the form $U = \cup_{a \in A} B_a$ for some family $(B_a)_{a \in A}$ of subsets from $\mathscr{B}$. We claim that $\mathscr{O}$ is a topology, and that $\mathscr{B}$ is a base for $\mathscr{O}$. It is clear that a union of sets from $\mathscr{O}$ is also in $\mathscr{O}$, by definition of $\mathscr{O}$. Now suppose that $U_1, \dots, U_k \in \mathscr{O}$. Then there exists index sets $A_1, \dots, A_k$ such that we can write $U_j = \cup_{a_j \in A_j} B_{a_j}$ for $j \in \{1, \dots, k\}$. By De Morgan's Laws,

$$\left(\cup_{a_1 \in A_1} B_{a_1}\right) \cap \cdots \cap \left(\cup_{a_k \in A_k} B_{a_k}\right) = \cup_{a_1 \in A_1} \cdots \cup_{a_k \in A_k} B_{a_1} \cap \cdots \cap B_{a_k}.$$

Now fix $(a_1, \dots, a_k) \in A_1 \times \cdots \times A_k$. For each $x \in \cap_{j=1}^k B_{a_j}$, there exists $B_x \in \mathscr{B}$ such that $x \in B_x$ and $B_x \subseteq \cap_{j=1}^k B_{a_j}$. Therefore, we have

$$B_{a_1} \cap \cdots \cap B_{a_k} = \cup_{x \in B_{a_1} \cap \cdots \cap B_{a_k}} B_x,$$

meaning that $B_{a_1} \cap \cdots \cap B_{a_k} \in \mathscr{O}$ for each $(a_1, \dots, a_k) \in A_1 \times \cdots \times A_k$. This means that $\cap_{j=1}^k U_j$ is a union of elements of $\mathscr{B}$, and so is in $\mathscr{O}$. Since $\varnothing \in \mathscr{O}$ and $S \in \mathscr{O}$, this shows that $\mathscr{O}$ is a topology. It is obvious that $\mathscr{B}$ is a base for $\mathscr{O}$. ■

Let us look at some sample families of subsets that do and do not form a base for a topology.

1.2.15 Examples (Families of sets that are bases)

1. Let (S, d) be a semimetric space. We already know that the family of open d-balls forms the base for a topology on S. Let us check this by verifying that the conditions of Proposition 1.2.14 are satisfied. Thus let $\mathsf{B}_d(r_1, x_1), \dots, \mathsf{B}_d(r_k, x_k)$ be k open d-balls.

   We claim that $x \in \cap_{j=1}^k \mathsf{B}_d(r_j, x_k)$ if and only if

$$\sum_{j=1}^k d(x, x_j) < \sum_{j=1}^k r_j. \tag{1.2}$$

   First, if $x \in \cap_{j=1}^k \mathsf{B}_d(r_j, x_j)$, then clearly (1.2) holds. Conversely, suppose that

$$\sum_{j=1}^k d(x, x_j) \geq \sum_{j=1}^k r_j.$$

   If $x \in \cap_{j=1}^{k-1} \mathsf{B}_d(r_j, x_j)$ then we have

$$\sum_{j=1}^{k-1} d(x, x_j) < \sum_{j=1}^{k-1} r_j.$$

   This then means that $d(x, x_k) \geq r_k$. Thus x cannot lie in $x \in \cap_{j=1}^k \mathsf{B}_d(r_j, x_j)$, and this gives our claim.

   Now suppose that $x \in \cap_{j=1}^k \mathsf{B}_d(r_j, x_k)$ and let $\epsilon > 0$ have the property that

2. Let $\mathscr{B} \subseteq \mathbf{2}^{\mathbb{R}}$ be the family of closed intervals of $\mathbb{R}$ that have nonzero length, along with the empty set. We claim that there is no topology on $\mathbb{R}$ for which $\mathscr{B}$ is a base. To see this, note that $[-1,0],[0,1] \in \mathscr{B}$, but that there is no element of $\mathscr{B}$ contained in $[-1,0] \cap [0,1] = \{0\}$.

   Note that if we had allowed closed intervals of zero length, i.e., singletons, then the resulting family of sets *would* form a base for a topology, the discrete topology, in fact. •

Now that we understand when a collection of subsets can form the base for a topology, it is sometimes useful to know when two potential bases generate the same topology.

1.2.16 Proposition (When do two bases generate the same topology?) *Let S be a set and let $\mathscr{B}_1$ and $\mathscr{B}_2$ be families of subsets of S that are bases for topologies $\mathscr{O}_1$ and $\mathscr{O}_2$, respectively. Then the following statements are equivalent:*

(i) $\mathscr{O}_1 = \mathscr{O}_2$;

(ii) *it holds that*

  (a) *for any $B_1 \in \mathscr{B}_1$ and for any $x \in B_1$, there exists $B_2 \in \mathscr{B}_2$ such that $x \in B_2$ and $B_2 \subseteq B_1$, and*

  (b) *for any $B_2 \in \mathscr{B}_2$ and for any $x \in B_2$, there exists $B_1 \in \mathscr{B}_1$ such that $x \in B_1$ and $B_1 \subseteq B_2$.*

Proof (i) $\Longrightarrow$ (ii) Let $B_1 \in \mathscr{B}_1$ and let $x \in B_1$. Since $\mathscr{O}_1 = \mathscr{O}_2$ and since B_1 is then open in the topology $\mathscr{O}_2$, it follows from Proposition 1.2.12 that there exists $B_2 \in \mathscr{B}_2$ such that $x \in B_2$ and $B_2 \subseteq B_1$. This gives one part of (ii) and the other part follows by the same argument, swapping "1" and "2."

(ii) $\Longrightarrow$ (i) Let $U_1 \in \mathscr{O}_1$ and let $x \in U_1$. By Proposition 1.2.12 there exists $B_1 \in \mathscr{B}_1$ such that $x \in B_1$ and $B_1 \subseteq U_1$. By the second part of (ii) there exists $B_2 \in \mathscr{B}_2$ such that $x \in B_2$ and $B_2 \subseteq B_1$. Thus, by Proposition 1.2.12 it follows that $U_1 \in \mathscr{O}_2$ and so $\mathscr{O}_1 \subseteq \mathscr{O}_2$. The argument can be repeated, swapping "1" and "2" to give $\mathscr{O}_2 \subseteq \mathscr{O}_1$. ■

Sometimes it is helpful to use a notion that is slightly more flexible than that of a base.

1.2.17 Definition (Subbase for a topology) A ***subbase*** for a topology $\mathscr{O}$ on a set S is a subset $\mathscr{S} \subseteq \mathscr{O}$ with the property that

$$\mathscr{B}_{\mathscr{S}} \triangleq \{\varnothing\} \cup \{S\} \cup \{\cap_{j=1}^{k} S_j \mid S_1, \dots, S_k \in \mathscr{S}\}$$

is a base for $\mathscr{O}$. •

The following characterisation of a subbase gives the essence of the notion.

1.2.18 Proposition (Characterisation of subbases) *Let* S *be a set with a topology* $\mathscr{O}$*. A subset* $\mathscr{S} \subseteq \mathscr{O}$ *is a subbase for* $\mathscr{O}$ *if and only if* $\mathscr{O}$ *is the smallest topology containing* $\mathscr{S}$*.*

Proof First suppose that $\mathscr{S}$ is a subbase for $\mathscr{O}$ and let $\tilde{\mathscr{O}}$ be a topology containing $\mathscr{S}$. Let $U \in \mathscr{O}$. If $U = \varnothing$ or $U = S$ then $U \in \tilde{\mathscr{O}}$. If $U = \cup_{a\in A} B_a$ for a family $(B_a)_{a\in A}$ of subsets from $\mathscr{B}_{\mathscr{S}}$ then it follows that each of the subsets B_a, $a \in A$, is in $\tilde{\mathscr{O}}$ since $\mathscr{S} \subseteq \tilde{\mathscr{O}}$ and since $\tilde{\mathscr{O}}$ is a topology and hence closed under finite intersections. Also, $\cup_{a\in A} B_a \in \tilde{\mathscr{O}}$ since $\mathscr{B}_{\mathscr{S}} \subseteq \tilde{\mathscr{O}}$ and since $\tilde{\mathscr{O}}$ is a topology and hence closed under arbitrary unions. This shows that $U \in \tilde{\mathscr{O}}$. Thus $\mathscr{O} \subseteq \tilde{\mathscr{O}}$ and so $\mathscr{O}$ is the smallest topology containing $\mathscr{S}$.

Now suppose that $\mathscr{O}$ is the smallest topology containing $\mathscr{S}$. Since $\mathscr{O}$ is closed under finite intersections and since $\mathscr{B} \subseteq \mathscr{O}$ it follows that $\mathscr{B}_{\mathscr{S}} \subseteq \mathscr{O}$. Since $\mathscr{O}$ is closed under arbitrary intersections it follows that the topology generated by $\mathscr{B}_{\mathscr{S}}$ is contained in $\mathscr{O}$. Moreover, since $\mathscr{O}$ is the smallest topology containing $\mathscr{S}$ it follows that $\mathscr{O}$ is equal to the topology generated by $\mathscr{B}_{\mathscr{S}}$. From this construction it immediately follows that $\mathscr{S}$ is a subbase for $\mathscr{O}$. ■

The next result explains one of the reasons why the notion of a subbase is so useful. *Any* collection of subsets is a subbase for some topology.

1.2.19 Proposition (When is a family of subsets a subbase?) *Any family* $\mathscr{S}$ *of subsets of a set* S *is a subbase for a topology on* S*.*

Proof We let $\mathscr{B}_{\mathscr{S}}$ be as in Definition 1.2.17 and let $\mathscr{O}$ be the topology generated by $\mathscr{B}_{\mathscr{S}}$. The only thing to show to prove the proposition is that $\mathscr{B}_{\mathscr{S}}$ satisfies the conditions of Proposition 1.2.14 required for it to be a base for a topology. Clearly $\varnothing \in \mathscr{B}_{\mathscr{S}}$ and $S = \cup_{B\in\mathscr{B}_{\mathscr{S}}} B$. Now let $B_1, \dots, B_k \in \mathscr{B}_{\mathscr{S}}$ and let $x \in \cap_{j=1}^k B_j$. Since each of the sets $B_1, \dots, B_k$ is in $\mathscr{B}_{\mathscr{S}}$ it follows from the definition of $\mathscr{B}_{\mathscr{S}}$ that $\cap_{j=1}^k B_j$ is also in $\mathscr{B}_{\mathscr{S}}$. Therefore, if $x \in \cap_{j=1}^k B_j$ then $x \in B$ with $B = \cap_{j=1}^k B_j \in \mathscr{B}_{\mathscr{S}}$, so showing that $\mathscr{B}_{\mathscr{S}}$ indeed generates a topology. ■

1.2.3 Neighbourhoods and neighbourhood bases

The notion of a neighbourhood is a simple one, but is convenient terminology.

1.2.20 Definition (Neighbourhood) If $(S, \mathscr{O})$ is a topological space and if $x \in S$ then a ***neighbourhood*** of x is an open set $U \in \mathscr{O}$ for which $x \in U$. The set of neighbourhoods of x is denoted by $\mathscr{N}_x$. •

Some authors do not ask that a neighbourhood of x be open, but only ask that a neighbourhood contain an open set which itself contains x. This is an inferior definition to the one we give.

Let us give some examples of neighbourhoods.

1.2.21 Examples (Neighbourhoods)

1. If $x \in \mathbb{R}$ then any open interval (a, b) containing x is a neighbourhood of x in the standard topology.

2. Let $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ and let $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}$ have the property that $x_j \in (a_j, b_j)$ for each $j \in \{1, \dots, n\}$. Then the rectangle

$$(a_1, b_1) \times \cdots \times (a_n, b_n) \subseteq \mathbb{R}^n$$

is a neighbourhood of $\boldsymbol{x}$ in the standard topology.

3. If (S, d) is a metric space and if $x \in S$ then any open ball $\mathsf{B}_d(r, x)$ centred at x is a neighbourhood of x in the metric topology. •

The idea of a neighbourhood can profitably be viewed as a generalisation of the notion of an open ball in a metric space. The idea is that the notion of "closeness" quantified by the metric is replaced by a general notion of "closeness." The generalisation, however, can exhibit some features that are simply not seen with metric spaces. The reader is asked to explore ideas along this line in Exercise 1.2.2.

Let us give some properties of the set of neighbourhoods of a point.

1.2.22 Proposition (Properties of neighbourhoods) *Let* $(\mathrm{S}, \mathscr{O})$ *be a topological space and let* $\mathrm{x} \in \mathrm{S}$. *Then the following statements hold:*

(i) $\mathrm{S} \in \mathscr{N}_{\mathrm{x}}$*;*

(ii) if $\mathrm{U} \in \mathscr{N}_{\mathrm{x}}$ *then* $\mathrm{x} \in \mathrm{U}$*;*

(iii) if $\mathrm{U}_1, \mathrm{U}_2 \in \mathscr{N}_{\mathrm{x}}$ *then* $\mathrm{U}_1 \cap \mathrm{U}_2 \in \mathscr{N}_{\mathrm{x}}$*;*

(iv) if $\mathrm{U} \in \mathscr{N}_{\mathrm{x}}$ *and if* $\mathrm{V} \in \mathscr{O}$ *satisfies* $\mathrm{U} \subseteq \mathrm{V}$ *then* $\mathrm{V} \in \mathscr{N}_{\mathrm{x}}$.

Moreover,

(v) $\mathrm{U} \subseteq \mathrm{S}$ *is open if and only if, for each* $\mathrm{x} \in \mathrm{U}$*, there exists* $\mathrm{U}_{\mathrm{x}} \in \mathscr{N}_{\mathrm{x}}$ *such that* $\mathrm{U}_{\mathrm{x}} \subseteq \mathrm{U}$.

Proof (i) and (ii) These are trivial.

(iii) Since U_1 and U_2 are open then $U_1 \cap U_2$ is open. Since $x \in U_1 \cap U_2$ we then have $U_1 \cap U_2 \in \mathscr{N}_x$.

(iv) This follows from the definitions.

(v) First suppose that U is open and let $x \in U$. Then $U \in \mathscr{N}_x$ and so there exists a neighbourhood of x contained in U. Now suppose that for each $x \in U$ there exists a neighbourhood $U_x \in \mathscr{N}_x$ contained in U. Then $U = \cup_{x \in U} U_x$ and so U is a union of open sets, and so it open itself. ■

While the notion of a neighbourhood is elementary, the notion still has value because it *defines* the topology, as the following result shows.

1.2.23 Proposition (Neighbourhoods define a topology) *Let* S *be a set, for each* $\mathrm{x} \in \mathrm{S}$ *let* $\mathscr{N}_{\mathrm{x}}$ *be a family of subsets of* S *satisfying conditions (i)–(iv) of Proposition 1.2.22, and define*

$$\mathscr{O} = \{\varnothing\} \cup \{\mathrm{U} \subseteq \mathrm{S} \mid \textit{for each } \mathrm{x} \in \mathrm{U} \textit{ there exists } \mathrm{U}_{\mathrm{x}} \in \mathscr{N}_{\mathrm{x}} \textit{ such that } \mathrm{U}_{\mathrm{x}} \subseteq \mathrm{U}\}.$$

Then $\mathscr{O}$ *is a topology for* S *and* $\mathscr{N}_{\mathrm{x}}$ *is the set of neighbourhoods of* x *in this topology.*

Proof We obviously have $\varnothing, S \in \mathscr{O}$. Let $(U_a)_{a\in A} \subseteq \mathscr{O}$ and let $x \in \cup_{a\in A} U_a$. For each fixed $a \in A$ there exists $U_x \in \mathscr{N}_x$ such that $U_x \subseteq U_a$. Thus $U_x \subseteq \cup_{a\in A} U_a$ and so $\cup_{a\in A} U_a$ is open. Now let $U_1, U_2 \in \mathscr{O}$ and let $x \in U_1 \cap U_2$. Then there exists $U_{1,x}, U_{2,x} \in \mathscr{N}_x$ such that $U_{1,x} \subseteq U_1$ and $U_{2,x} \subseteq U_2$. By property (iii) of Proposition 1.2.22 we have $U_{1,x} \cap U_{2,x} \in \mathscr{N}_x$ and so $U_1 \cap U_2 \in \mathscr{O}$ since $U_{1,x} \cap U_{2,x} \subseteq U_1 \cap U_2$. ■

For neighbourhoods there is a notion like that of bases for topologies.

1.2.24 Definition (Neighbourhood base) For a topological space $(S, \mathscr{O})$ and for $x \in S$, a ***neighbourhood base*** for x is a family $(B_a)_{a\in A}$ of neighbourhoods of x with the property that, if U is a neighbourhood of x, then there exists $a \in A$ such that $B_a \subseteq U$. •

Neighbourhood bases have the following properties.

1.2.25 Proposition (Properties of neighbourhood bases) *Let* $(S, \mathscr{O})$ *be a topological space, let* $x \in S$, *and let* $\mathscr{B}_x$ *be a neighbourhood base at* x. *Then the following statements hold:*

(i) *if* $B \in \mathscr{B}_x$ *then* $x \in B$;

(ii) *if* $B_1, B_2 \in \mathscr{B}_x$ *then there exists* $B \in \mathscr{B}_x$ *such that* $B \subseteq B_1 \cap B_2$.

Moreover, if we choose a neighbourhood base $\mathscr{B}_x$ *for each* $x \in S$,

(iii) $U \subseteq S$ *is open if and only if, for each* $x \in S$, *there exists* $B_x \in \mathscr{B}_x$ *such that* $B_x \subseteq U$.

Proof (i) This follows since $\mathscr{B}_x \subseteq \mathscr{N}_x$.

(ii) Since $\mathscr{B}_x \subseteq \mathscr{N}_x$ we have $B_1 \cap B_2 \in \mathscr{N}_x$ by Proposition 1.2.22. Thus, by definition of a neighbourhood base, there exists $B \in \mathscr{B}_x$ such that $B \subseteq B_1 \cap B_2$.

(iii) Suppose that U is open and let $x \in U$. Then, by Proposition 1.2.22, there exists $U_x \in \mathscr{N}_x$ such that $U_x \subseteq U$. By the property of a neighbourhood base there exists $B_x \subseteq U_x \subseteq U$. Now suppose that for each $x \in U$ there exists $B_x \in \mathscr{B}_x$ for which $B_x \subseteq U$. Since $\mathscr{B}_x \subseteq \mathscr{N}_x$ for each $x \in U$, it follows from Proposition 1.2.22 that U is open. ■

The great value of neighbourhood bases is that they can be used to define bases for a topology. That is, to define a topology all one need do is specify a neighbourhood base satisfying the following conditions.

1.2.26 Proposition (Neighbourhood bases define bases) *Let* S *be a set, for each* $x \in S$ *let* $\mathscr{B}_x$ *be a family of subsets of* S *satisfying conditions (i) and (ii), and define*

$$\mathscr{O} = \{\varnothing\} \cup \{S\} \cup \{U \subseteq S \mid \textit{for each } x \in U \textit{ there exists } B_x \in \mathscr{B}_x \textit{ such that } B_x \subseteq U\}.$$

Then $\mathscr{O}$ *is a topology for* S, $\cup_{x\in S}\mathscr{B}_x$ *is a base for* S, *and* $\mathscr{B}_x$ *is a neighbourhood base for* x *in this topology.*

Proof By construction $\varnothing, S \in \mathscr{O}$. Let $U_1, U_2 \in \mathscr{O}$ and let $x \in U_1 \cap U_2$. Then there exists $B_{1,x}, B_{2,x} \in \mathscr{B}_x$ such that $B_{1,x} \subseteq U_1$ and $B_{2,x} \subseteq U_2$. There then also exists $B_x \in \mathscr{B}_x$ such that $B_x \subseteq B_{1,x} \cap B_{2,x}$. Then $B_x \subseteq U_1 \cap U_2$ and so $U_1 \cap U_2 \in \mathscr{O}$. Now let $(U_a)_{a\in A}$ be a family of subsets of $\mathscr{O}$ and let $x \in \cup_{a\in A} U_a$. Then, for each fixed $a \in A$, there exists $B_x \in \mathscr{B}_x$ such that $B_x \subseteq U_a$. Therefore, $B_x \subseteq \cup_{a\in A} U_a$ and so $\cup_{a\in A} U_a \in \mathscr{O}$. This $\mathscr{O}$ is a topology.

Now let $U \in \mathscr{O}$ and for each $x \in U$ let $B_x \in \mathscr{B}_x$ have the property that $B_x \subseteq U$. Then $U = \cup_{x \in U} B_x$ and so $\cup_{x \in S} \mathscr{B}_x$ is a base for $\mathscr{O}$.

Finally, let $U \subseteq S$ be a neighbourhood of x in the topology $\mathscr{O}$. By definition of $\mathscr{O}$ there exists $B_x \in \mathscr{B}_x$ such that $B_x \subseteq U$. Thus $\mathscr{B}_x$ is a neighbourhood base for x in the topology $\mathscr{O}$. ■

Let us give some examples of neighbourhood bases.

1.2.27 Examples (Neighbourhood bases) In each of the following examples the reader is asked to verify that the conditions of Proposition 1.2.26 hold for the proposed neighbourhood bases.

1. Let $x \in \mathbb{R}$. Then, referring to Example 1.2.21–1, the set of open intervals containing x is a neighbourhood base for x in the standard topology.
2. Let $\boldsymbol{x} \in \mathbb{R}^n$. Then, referring to Example 1.2.21–2, the set of open rectangles containing $\boldsymbol{x}$ is a neighbourhood base for $\boldsymbol{x}$ in the standard topology.
3. For a metric space (S, d) and for $x \in S$, the set of open balls with centre at x is a neighbourhood base in the metric topology. •

1.2.4 Interior, closure, boundary, etc.

In this section we study some properties of subsets of topological spaces. The reader will observe that most of the ideas here appear already in Section I-2.5. Indeed, the proofs here are almost identical to those for the simple case of the topological space $\mathbb{R}$. We hope that this will serve to make the reader believe in the axioms of topological spaces.

Let us first consider the various flavours of "boundary points" that can arise in a topological space. In Section I-2.5 we saw that these notions were interdependent. For general topological spaces this is not the case. The definition of limit point in the following definition relies on the notion of a convergent sequence. The reader may wish to refer forward to Definition 1.5.1 for this.

1.2.28 Definition (Accumulation point, cluster point, limit point) Let $(S, \mathscr{O}$ be a topological space. For a subset $A \subseteq S$, a point $x \in S$ is:

(i) an ***accumulation point*** for A if, for every neighbourhood U of x, the set $A \cap (U \setminus \{x\})$ is nonempty;

(ii) a ***cluster point*** for A if, for every neighbourhood U of x, the set $A \cap U$ is infinite;

(iii) a ***limit point*** of A if there exists a sequence $(x_j)_{j \in \mathbb{Z}_{>0}}$ in A converging to x.

The set of accumulation points of A is called the ***derived set*** of A, and is denoted by $\mathrm{der}(A)$. •

Let us flesh out the relationships between these concepts. It turns out that the only general implication that one can make is the obvious one.

1.2.29 Proposition (A cluster point is an accumulation point) *If* $(S, \mathscr{O})$ *is a topological space, if* $A \subseteq S$*, and if* $x \in S$ *is a cluster point for* A*, then it is an accumulation point for* A*.*

Proof If $U \cap A$ is infinite then $(U \setminus \{x\}) \cap A$ must be nonempty. That is, if x is a cluster point then it is an accumulation point. ■

The following examples, some of which are somewhat esoteric, show that the implication of the preceding proposition is the only one that generally holds. We shall see that there are "niceness" conditions which rule out some of these examples; see . For more examples we refer the reader to Example I-2.5.14. what?

1.2.30 Examples (Accumulation points, cluster points, limit points)

1. The first example we give is of a point that is an accumulation point, a cluster point, and a limit point for a set. This is easy to do. Let $A = [0, 1] \subseteq \mathbb{R}$. It is easy to check that 0 is an accumulation point, a cluster point, and a limit point.
2. We will give an example of an accumulation point that is not a cluster point. Let $S = \{x_1, x_2\}$ be a set with two points and define $\mathscr{O} = \{\emptyset, \{x_1\}, S\}$. It is a trivial direct verification to see that $\mathscr{O}$ is a topology on S. We claim that x_2 is an accumulation point for the set $A = \{x_1\}$. Indeed, the only neighbourhood of x_2 is S and $A \cap (S \setminus \{x_2\}) = \{x_1\}$ is nonempty. Thus, x_2 is an accumulation point for A as desired. However, the only neighbourhood of x_2 has cardinality 2 and so x_2 is not a cluster point for A.
3. We will give an example of an accumulation point that is not a limit point. Let S be an uncountable set and let $\mathscr{O}$ be the topology whose open sets are complements of countable subsets of S, along with, of course, S and $\emptyset$. That this is a topological space we leave to the reader to verify (this is Exercise 1.2.1). Let $A \subset S$ be uncountable. We claim that if $x \in S \setminus A$ then x is an accumulation point of A. Indeed, let U be a neighbourhood of x. We claim that $(U \setminus \{x\}) \cap A$ is nonempty. Note that $U = S \setminus C$ for a countable set C. Therefore, $(U \setminus \{x\}) \cap A = (S \setminus (\{x\} \cup C)) \cap A$. If $(S \setminus (\{x\} \cup C)) \cap A = \emptyset$ then it must be the case that $A \subseteq (\{x\} \cup C)$ which cannot be the case since A is uncountable and C is countable. Therefore, $(U \setminus \{x\}) \cap A \neq \emptyset$ and so x is indeed an accumulation point.

   In Example 1.5.2–1 we will show that the only sequences in S that converge are those that are eventually constant. Therefore, if $x \in S \setminus A$ then there can be no sequence in A converging to x. Thus x is not a limit point.
4. We will give an example of a cluster point that is not a limit point. As in the preceding example we let S be an uncountable set and let $\mathscr{O}$ be the topology for which open sets are those whose complements are countable, along with $\emptyset$ and S. We again let $A \subset S$ be an uncountable subset and let $x \in S \setminus A$. We take U as a neighbourhood of x and we claim that $U \cap A$ is infinite. Indeed, just as in the preceding example, $U = S \setminus C$ for a countable set C and so $U \cap A = (S \setminus C) \cap A = A - C$. Since C is countable and A is uncountable, $A - C$ is uncountable and so $U \cap A$ is also uncountable. In particular, $U \cap A$ is infinite

and so x is a cluster point of A. But, as we saw in the preceding example, x is not a limit point of A.

5. We will give an example of a limit point that is not an accumulation point. For the set $A = [0,1] \cup \{2\} \subseteq \mathbb{R}$, the point 2 is a limit point of A but not an accumulation point of A.
6. We will give an example of a limit point that is not a cluster point. Consider the set $A = \mathbb{Z} \subseteq \mathbb{R}$. We claim that 0 is a limit point but not a cluster point. Indeed, the constant sequence $(0)_{j\in\mathbb{Z}_{>0}}$ converges to 0 and so 0 is a limit point. However, the neighbourhood $(-\frac{1}{2}, \frac{1}{2})$ of 0 only contains the point 0 from A and so 0 is not a cluster point. •

Let us give some properties of the derived set.

1.2.31 Proposition (Properties of the derived set) *For a topological space* $(\mathrm{S}, \mathscr{O})$ *for* $\mathrm{A}, \mathrm{B} \subseteq \mathrm{S}$*, and for a family of subsets* $(\mathrm{A_i})_{\mathrm{i}\in\mathrm{I}}$ *of* S*, the following statements hold:*

(i) $\mathrm{der}(\varnothing) = \varnothing$*;*

(ii) $\mathrm{der}(\mathrm{S}) = \mathrm{S}$*;*

(iii) $\mathrm{der}(\mathrm{der}(\mathrm{A})) = \mathrm{der}(\mathrm{A})$*;*

(iv) if $\mathrm{A} \subseteq \mathrm{B}$ *then* $\mathrm{der}(\mathrm{A}) \subseteq \mathrm{der}(\mathrm{B})$*;*

(v) $\mathrm{der}(\mathrm{A} \cup \mathrm{B}) = \mathrm{der}(\mathrm{A}) \cup \mathrm{der}(\mathrm{B})$*;*

(vi) $\mathrm{der}(\mathrm{A} \cap \mathrm{B}) \subseteq \mathrm{der}(\mathrm{A}) \cap \mathrm{der}(\mathrm{B})$.

Proof Parts (i) and (ii) follow directly from the definition of the derived set.

True or not? (iii)

(iv) Let $x \in \mathrm{der}(A)$ and let U be a neighbourhood of x. Then the set $A \cap (U \setminus \{x\})$ is nonempty, implying that the set $B \cap (U \setminus \{x\})$ is also nonempty. Thus $x \in \mathrm{der}(B)$.

(v) Let $x \in \mathrm{der}(A \cup B)$ and let U be a neighbourhood of x. Then the set $U \cap ((A \cup B) \setminus \{x\})$ is nonempty. But

$$\begin{aligned} U \cap ((A \cup B) \setminus \{x\}) &= U \cap ((A \setminus \{x\}) \cup (B \setminus \{x\})) \\ &= (U \cap (A \setminus \{x\})) \cup (U \cap (B \setminus \{x\})). \end{aligned} \tag{1.3}$$

Thus it cannot be that both $U \cap (A \setminus \{x\})$ and $U \cap (B \setminus \{x\})$ are empty. Thus x is an element of either $\mathrm{der}(A)$ or $\mathrm{der}(B)$.

Now let $x \in \mathrm{der}(A) \cup \mathrm{der}(A)$. Then, using (1.3), $U \cap ((A \cup B) \setminus \{x\})$ is nonempty, and so $x \in \mathrm{der}(A \cup B)$.

(vi) Let $x \in \mathrm{der}(A \cap B)$ and let U be a neighbourhood of x. Then $U \cap ((A \cap B) \setminus \{x\}) \neq \varnothing$. We have

$$U \cap ((A \cap B) \setminus \{x\}) = U \cap ((A \setminus \{x\}) \cap (B \setminus \{x\}))$$

Thus the sets $U \cap (A \setminus \{x\})$ and $U \cap (B \setminus \{x\})$ are both nonempty, showing that $x \in \mathrm{der}(A) \cap \mathrm{der}(B)$. ■

Now let us turn to the notion of interior, closure, and boundary for general topological spaces.

1.2.32 Definition (Interior, closure, and boundary) Let $(S, \mathscr{O})$ be a topological space and let $A \subseteq S$.

(i) The ***interior*** of A is the set

$$\mathrm{int}(A) = \cup\{U \mid U \subseteq A,\ U \text{ open}\}.$$

(ii) The ***closure*** of A is the set

$$\mathrm{cl}(A) = \cap\{C \mid A \subseteq C,\ C \text{ closed}\}.$$

(iii) The ***boundary*** of A is the set $\mathrm{bd}(A) = \mathrm{cl}(A) \cap \mathrm{cl}(S \setminus A)$. •

The following characterisations of interior, closure, and boundary are helpful.

1.2.33 Proposition (Characterisation of interior, closure, and boundary) *For a topological space* $(S, \mathscr{O})$ *and for* $A \subseteq S$*, the following statements hold:*

(i) $x \in \mathrm{int}(A)$ *if and only if there exists a neighbourhood* U *of* x *such that* $U \subseteq A$*;*

(ii) $x \in \mathrm{cl}(A)$ *if and only if, for each neighbourhood* U *of* x*, the set* $U \cap A$ *is nonempty;*

(iii) $x \in \mathrm{bd}(A)$ *if and only if, for each neighbourhood* U *of* x*, the sets* $U \cap A$ *and* $U \cap (S \setminus A)$ *are nonempty.*

Proof (i) Suppose that $x \in \mathrm{int}(A)$. Since $\mathrm{int}(A)$ is open, there exists a neighbourhood U of x contained in $\mathrm{int}(A)$. Since $\mathrm{int}(A) \subseteq A$, $U \subseteq A$.

Next suppose that $x \notin \mathrm{int}(A)$. Then, by definition of interior, for any open set U for which $U \subseteq A$, $x \notin U$.

(ii) Suppose that there exists a neighbourhood U of x such that $U \cap A = \varnothing$. Then $S \setminus U$ is a closed set containing A. Thus $\mathrm{cl}(A) \subseteq S \setminus U$. Since $x \notin S \setminus U$, it follows that $x \notin \mathrm{cl}(A)$.

Suppose that $x \notin \mathrm{cl}(A)$. Then x is an element of the open set $S \setminus \mathrm{cl}(A)$. Thus there exists a neighbourhood U of x such that $U \subseteq S \setminus \mathrm{cl}(A)$. In particular, $U \cap A = \varnothing$.

(iii) This follows directly from part (ii) and the definition of boundary. ■

The reader should refer to Example I-2.5.17 for some simple examples illustrating the concepts of interior, closure, and boundary. Here we give examples to show what can happen in some atypical cases of topological spaces.

1.2.34 Examples (Interior, closure, and boundary)

1. Let S be a set and consider the trivial topology on S. If $A \subset S$ is nonempty then

$$\mathrm{int}(A) = \varnothing, \quad \mathrm{cl}(A) = S, \quad \mathrm{bd}(S) = S.$$

We also have

$$\mathrm{int}(\varnothing) = \mathrm{cl}(\varnothing) = \mathrm{bd}(\varnothing) = \varnothing, \quad \mathrm{int}(S) = \mathrm{cl}(S) = S, \quad \mathrm{bd}(S) = \varnothing.$$

2. We again let S be an arbitrary set but now consider the discrete topology. If $A \subset S$ is nonempty then we have

$$\operatorname{int}(A) = \operatorname{cl}(A) = A, \quad \operatorname{bd}(A) = \varnothing.$$

It also holds that

$$\operatorname{int}(\varnothing) = \operatorname{cl}(\varnothing) = \operatorname{bd}(\varnothing) = \varnothing, \quad \operatorname{int}(S) = \operatorname{cl}(S) = S, \quad \operatorname{bd}(S) = \varnothing.$$

3. Let S be an uncountable set and following Exercise 1.2.1 we let $\mathscr{O}$ be the collection of subsets of S whose complement is countable, along with $\varnothing$ and S. We then have

$$\operatorname{int}(A) = \begin{cases} \varnothing, & A \text{ countable}, \\ A, & S \setminus A \text{ countable}, \\ ?, & A, S \setminus A \text{ uncountable}, \end{cases} \qquad \operatorname{cl}(A) = \begin{cases} A, & A \text{ countable}, \\ S, & S \setminus A \text{ countable}, \\ ?, & A, S \setminus A \text{ uncountable}. \end{cases}$$

The "?" mean that there is no useful general characterisation. The verification of these statements takes a moments thought, but it is only a matter of checking the definitions. •

Let us give some general properties of the interior operation.

1.2.35 Proposition (Properties of interior) *Let* $(S, \mathscr{O})$ *be a topological space. For* $A, B \subseteq S$ *and for a family of subsets* $(A_i)_{i \in I}$ *of* S*, the following statements hold:*

(i) $\operatorname{int}(\varnothing) = \varnothing$*;*

(ii) $\operatorname{int}(S) = S$*;*

(iii) $\operatorname{int}(\operatorname{int}(A)) = \operatorname{int}(A)$*;*

(iv) if $A \subseteq B$ *then* $\operatorname{int}(A) \subseteq \operatorname{int}(B)$*;*

(v) $\operatorname{int}(A \cup B) \supseteq \operatorname{int}(A) \cup \operatorname{int}(B)$*;*

(vi) $\operatorname{int}(A \cap B) = \operatorname{int}(A) \cap \operatorname{int}(B)$*;*

(vii) $\operatorname{int}(\cup_{i \in I} A_i) \supseteq \cup_{i \in I} \operatorname{int}(A_i)$*;*

(viii) $\operatorname{int}(\cap_{i \in I} A_i) \subseteq \cap_{i \in I} \operatorname{int}(A_i)$*.*

Moreover, a set $A \subseteq S$ *is open if and only if* $\operatorname{int}(A) = A$*.*

Proof Parts (i) and (ii) are clear by definition of interior. Part (v) follows from part (vii), so we will only prove the latter.

(iii) This follows since the interior of an open set is the set itself.

(iv) Let $x \in \operatorname{int}(A)$. Then there exists a neighbourhood U of x such that $U \subseteq A$. Thus $U \subseteq B$, and the result follows from Proposition 1.2.33.

(vi) Let $x \in \operatorname{int}(A) \cap \operatorname{int}(B)$. Since $\operatorname{int}(A) \cap \operatorname{int}(B)$ is open by Exercise I-2.5.1, there exists a neighbourhood U of x such that $U \subseteq \operatorname{int}(A) \cap \operatorname{int}(B)$. Thus $U \subseteq A \cap B$. This shows that $x \in \operatorname{int}(A \cap B)$. This part of the result follows from part (viii).

(vii) Let $x \in \cup_{i \in I} \operatorname{int}(A_i)$. By Exercise I-2.5.1 the set $\cup_{i \in I} \operatorname{int}(A_i)$ is open. Thus there exists a neighbourhood U of x such that $U \subseteq \cup_{i \in I} \operatorname{int}(A_i)$. Thus $U \subseteq \cup_{i \in I} A_i$, from which we conclude that $x \in \operatorname{int}(\cup_{i \in I} A_i)$.

(viii) Let $x \in \operatorname{int}(\cap_{i\in I} A_i)$. Then there exists a neighbourhood U of x such that $U \subseteq \cap_{i\in I} A_i$. It therefore follows that $U \subseteq A_i$ for each $i \in I$, and so that $x \in \operatorname{int}(A_i)$ for each $i \in I$.

The final assertion follows directly from Proposition 1.2.33. ■

Now let us give the analogous result for the closure.

1.2.36 Proposition (Properties of closure) *Let* $(S, \mathscr{O})$ *be a topological space. For* $A, B \subseteq S$ *and for a family of subsets* $(A_i)_{i\in I}$ *of* S*, the following statements hold:*

(i) $\operatorname{cl}(\emptyset) = \emptyset$*;*

(ii) $\operatorname{cl}(S) = S$*;*

(iii) $\operatorname{cl}(\operatorname{cl}(A)) = \operatorname{cl}(A)$*;*

(iv) if $A \subseteq B$ *then* $\operatorname{cl}(A) \subseteq \operatorname{cl}(B)$*;*

(v) $\operatorname{cl}(A \cup B) = \operatorname{cl}(A) \cup \operatorname{cl}(B)$*;*

(vi) $\operatorname{cl}(A \cap B) \subseteq \operatorname{cl}(A) \cap \operatorname{cl}(B)$*;*

(vii) $\operatorname{cl}(\cup_{i\in I} A_i) \supseteq \cup_{i\in I} \operatorname{cl}(A_i)$*;*

(viii) $\operatorname{cl}(\cap_{i\in I} A_i) \subseteq \cap_{i\in I} \operatorname{cl}(A_i)$*.*

Moreover, a set $A \subseteq S$ *is closed if and only if* $\operatorname{cl}(A) = A$*.*

Proof Parts (i) and (ii) follow immediately from the definition of closure. Part (vi) follows from part (viii), so we will only prove the latter.

(iii) This follows since the closure of a closed set is the set itself.

(iv) Suppose that $x \in \operatorname{cl}(A)$. Then, for any neighbourhood U of x, the set $U \cap A$ is nonempty, by Proposition 1.2.33. Since $A \subseteq B$, it follows that $U \cap B$ is also nonempty, and so $x \in \operatorname{cl}(B)$.

(v) Let $x \in \operatorname{cl}(A \cup B)$. Then, for any neighbourhood U of x, the set $U \cap (A \cup B)$ is nonempty by Proposition 1.2.33. By Proposition I-1.1.4, $U \cap (A \cup B) = (U \cap A) \cup (U \cap B)$. Thus the sets $U \cap A$ and $U \cap B$ are not both nonempty, and so $x \in \operatorname{cl}(A) \cup \operatorname{cl}(B)$. That $\operatorname{cl}(A) \cup \operatorname{cl}(B) \subseteq \operatorname{cl}(A \cup B)$ follows from part (vii).

(vi) Let $x \in \operatorname{cl}(A \cap B)$. Then, for any neighbourhood U of x, the set $U \cap (A \cap B)$ is nonempty. Thus the sets $U \cap A$ and $U \cap B$ are nonempty, and so $x \in \operatorname{cl}(A) \cap \operatorname{cl}(B)$.

(vii) Let $x \in \cup_{i\in I} \operatorname{cl}(A_i)$ and let U be a neighbourhood of x. Then, for each $i \in I$, $U \cap A_i \neq \emptyset$. Therefore, $\cup_{i\in I}(U \cap A_i) \neq \emptyset$. By Proposition I-1.1.7, $\cup_{i\in I}(U \cap A_i) = U \cap (\cup_{i\in I} A_i)$, showing that $U \cap (\cup_{i\in I} A_i) \neq \emptyset$. Thus $x \in \operatorname{cl}(\cup_{i\in I} A_i)$.

(viii) Let $x \in \operatorname{cl}(\cap_{i\in I} A_i)$ and let U be a neighbourhood of x. Then the set $U \cap (\cap_{i\in I} A_i)$ is nonempty. This means that, for each $i \in I$, the set $U \cap A_i$ is nonempty. Thus $x \in \operatorname{cl}(A_i)$ for each $i \in I$, giving the result. ■

The interior, closure, and boundary enjoy some joint properties.

1.2.37 Proposition (Joint properties of interior, closure, boundary, and derived set) *Let* $(S, \mathscr{O})$ *be a topological space. For* $A \subseteq S$*, the following statements hold:*

(i) $S \setminus \operatorname{int}(A) = \operatorname{cl}(S \setminus A)$*;*

(ii) $S \setminus \operatorname{cl}(A) = \operatorname{int}(S \setminus A)$*.*

(iii) $\operatorname{cl}(A) = A \cup \operatorname{bd}(A)$*;*

(iv) $\mathrm{int}(A) = A - \mathrm{bd}(A)$;

(v) $\mathrm{cl}(A) = \mathrm{int}(A) \cup \mathrm{bd}(A)$;

(vi) $\mathrm{cl}(A) = A \cup \mathrm{der}(A)$;

(vii) $S = \mathrm{int}(A) \cup \mathrm{bd}(A) \cup \mathrm{int}(S \setminus A)$.

Proof (i) Let $x \in S \setminus \mathrm{int}(A)$. Since $x \notin \mathrm{int}(A)$, for every neighbourhood U of x it holds that $U \not\subseteq A$. Thus, for any neighbourhood U of x, we have $U \cap (S \setminus A) \neq \emptyset$, showing that $x \in \mathrm{cl}(S \setminus A)$.

Now let $x \in \mathrm{cl}(S \setminus A)$. Then for any neighbourhood U of x we have $U \cap (S \setminus A) \neq \emptyset$. Thus $x \notin \mathrm{int}(A)$, so $x \in S \setminus A$.

(ii) The proof here strongly resembles that for part (i), and we encourage the reader to provide the explicit arguments.

(iii) This follows from part (v).

(iv) Clearly $\int(A) \subseteq A$. Suppose that $x \in A \cap \mathrm{bd}(A)$. Then, for any neighbourhood U of x, the set $U \cap (S \setminus A)$ is nonempty. Therefore, no neighbourhood of x is a subset of A, and so $x \notin \mathrm{int}(A)$. Conversely, if $x \in \mathrm{int}(A)$ then there is a neighbourhood U of x such that $U \subseteq A$. The precludes the set $U \cap (S \setminus A)$ from being nonempty, and so we must have $x \notin \mathrm{bd}(A)$.

(v) Let $x \in \mathrm{cl}(A)$. For a neighbourhood U of x it then holds that $U \cap A \neq \emptyset$. If there exists a neighbourhood V of x such that $V \subseteq A$, then $x \in \mathrm{int}(A)$. If there exists *no* neighbourhood V of x such that $V \subseteq A$, then for every neighbourhood V of x we have $V \cap (S \setminus A) \neq \emptyset$, and so $x \in \mathrm{bd}(A)$.

Now let $x \in \mathrm{int}(A) \cup \mathrm{bd}(A)$. If $x \in \mathrm{int}(A)$ then $x \in A$ and so $x \in\subseteq \mathrm{cl}(A)$. If $x \in \mathrm{bd}(A)$ then it follows immediately from Proposition 1.2.33 that $x \in \mathrm{cl}(A)$.

(vi) Let $x \in \mathrm{cl}(A)$. If $x \notin A$ then, for every neighbourhood U of x, $U \cap A = U \cap (A \setminus \{x\}) \neq \emptyset$, and so $x \in \mathrm{der}(A)$.

If $x \in A \cup \mathrm{der}(A)$ then either $x \in A \subseteq \mathrm{cl}(A)$, or $x \notin A$. In this latter case, $x \in \mathrm{der}(A)$ and so the set $U \cap (A \setminus \{x\})$ is nonempty for each neighbourhood U of x, and we again conclude that $x \in \mathrm{cl}(A)$.

(vii) Clearly $\mathrm{int}(A) \cap \mathrm{int}(S \setminus A) = \emptyset$ since $A \cap (S \setminus A) = \emptyset$. Now let $x \in S \setminus (\mathrm{int}(A) \cup \mathrm{int}(S \setminus A))$. Then, for any neighbourhood U of x, we have $U \not\subseteq A$ and $U \not\subseteq (S \setminus A)$. Thus the sets $U \cap (S \setminus A)$ and $U \cap A$ must both be nonempty, from which we conclude that $x \in \mathrm{bd}(A)$. ■

Exercises

1.2.1 Let S be an uncountable set and define

$$\mathscr{O} = \{\emptyset\} \cup \{S\} \cup \{U \subseteq S \mid S \setminus U \text{ is countable}\}.$$

Show that $\mathscr{O}$ is a topology on S.

1.2.2 Answer the following questions.

(a) For a metric space (S, d) and $x \in S$ show that the set

$$\{\mathsf{B}_d(r, x) \mid r \in \mathbb{R}_{>0}\}$$

is totally ordered.

(b) For a topological space $(S, \mathscr{O})$ and $x \in S$ show that $\mathscr{N}_x$ is partially ordered.

(c) Let (S, d) be a metric space and let $x, y \in S$. Show that x is within distance r of y if and only if y is within distance r of x. That is, show that "closeness" is symmetric in a metric space.

(d)

Example where y is in every neighbourhood of x by no neighbourhood of y contains x

Section 1.3

Continuity

One of the most significant contributions made by general topology is that it allows one to define the notion of continuity in very general settings. We shall make great use of this in understanding important concepts in linear analysis in Chapters 3, 4, and 6.

1.3.1 Proposition (Continuity and closed sets) *For topological spaces* S *and* T *and for* $f\colon S \to T$*, the following statements are equivalent:*

(i) f *is continuous;*

(ii) $f^{-1}(A)$ *is closed for every closed subset* $A \subseteq T$*.*

1.3.2 Corollary (Homeomorphisms are open and closed)

Exercises

1.3.1

Section 1.4

Subspaces, products, and quotients

1.4.1 Subspaces of topological spaces

1.4.1 Definition (Relative topology)

Section 1.5

Convergence in topological spaces

1.5.1 Definition (Sequences in topological spaces)

1.5.2 Examples (Convergent sequences)

1. Let S be an uncountable set and let $\mathscr{O}$ be the collection of subsets of S whose complement is countable, along with $\varnothing$ and S. The reader is asked in Exercise 1.2.1 that $\mathscr{O}$ is a topology on S. We claim that the only convergent sequences in this topology are the eventually constant sequences. Indeed, suppose that a sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ converges to x. Note that $\{x\}$ is a neighbourhood of x in the topology $\mathscr{O}$; this immediate implies that there exists $N \in \mathbb{Z}_{>0}$ such that $x_j = x$ for $j \geq N$.

   As an aside, note that the convergent sequences in the topology $\mathscr{O}$ are the same as the convergent sequences in the discrete topology. However, $\mathscr{O}$ is not equal to the discrete topology, and this shows that the convergent sequences do not uniquely determine a topology.

1.5.3 Corollary (Closed subsets of metric spaces) *If* (S, d) *is a metric space then a subset* $\mathrm{A} \subseteq \mathrm{S}$ *is closed if and only if every convergent sequence* $(\mathrm{x_j})_{\mathrm{j}\in\mathbb{Z}}$ *in* A *has a limit in* A*.*

1.5.1 The most general setting for Landau symbols

Section 1.6

Compactness

1.6.1 Example The subset $\{(0,0)\} \cup \{(x,y) \mid x > 0\}$ is not locally compact since no compact subset contains $(0,0)$.

1.6.2 Example The long line is sequentially compact but not compact.

1.6.3 Theorem (Bolzano–Weierstrass Theorem)

1.6.4 Proposition (Closed subsets of compact sets are compact)

1.6.5 Proposition (Continuous images of compact sets are compact)

1.6.6 Proposition (Compact sets are closed)

1.6.7 Theorem (Continuous functions achieve their maximum and minimum on compact sets)

Exercises

1.6.1 For any topological space S and for $x \in S$, show that $\{x\}$ is compact.

1.6.2 Let S be a topological space.

(a) Show that a finite union of compact subsets of S is compact.

(b) Is it true that a countable union of compact subsets is compact?

Section 1.7

Connectedness

Section 1.8

Countability and separability conditions

1.8.1 Countability conditions

1.8.2 Separability definitions

1.8.1 Proposition (Metric spaces are Hausdorff)

1.8.2 Proposition (Accumulation point equals cluster point for T_1-spaces) *If* $(S, \mathscr{O})$ *is a* T_1*-topological space and if* $A \subseteq S$*, then* $x \in S$ *is an accumulation point if and only if it is a cluster point.*

Proof By Proposition 1.2.29 cluster points are always accumulation points. Suppose that x is not a cluster point and that N is a neighbourhood of x that contains only finitely many points $x_1, \dots, x_k$ from A. Since S is T_1, for each $j \in \{1, \dots, k\}$ there exists a neighbourhood N_j of x such that N_j does not contain x_j. Then the neighbourhood $\cap_{j=1}^{k} N_j$ does not contain any of the points $x_1, \dots, x_k$ and so x cannot be an accumulation point for A. ■

Section 1.9

The topologies of convergence of maps

1.9.1 The pointwise convergence topology

1.9.2 The uniform convergence topology

1.9.3 The compact-open topology

1.9.4 Maps between metric spaces

1.9.1 Theorem *Let* $(\mathrm{S}, \mathrm{d}_\mathrm{S})$ *and* $(\mathrm{T}, \mathrm{d}_\mathrm{T})$ *be metric spaces with* $(\mathrm{T}, \mathrm{d}_\mathrm{T})$ *complete, and let* $\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{S};\mathrm{T})$ *denote the set of continuous bounded functions from* S *to* T*:*

$$\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{S};\mathrm{T}) = \{\mathrm{f} \in \mathsf{C}^0(\mathrm{S};\mathrm{T}) \mid \textit{ there exists } \mathrm{y} \in \mathrm{T} \textit{ and } \mathrm{R} > 0 \textit{ such that } \mathrm{image(f)} \subseteq \mathsf{B}_{\mathrm{d_T}}(\mathrm{R}, \mathrm{y})\}.$$

If we define a metric $\mathrm{d}_{\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{S};\mathrm{T})}$ *on* $\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{S};\mathrm{T})$ *by*

$$\mathrm{d}_{\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{S};\mathrm{T})}(\mathrm{f}, \mathrm{g}) = \sup\{\mathrm{d_T}(\mathrm{f(x)}, \mathrm{g(x)}) \mid \mathrm{x} \in \mathrm{S}\},$$

then $(\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{S};\mathrm{T}), \mathrm{d}_{\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{S};\mathrm{T})})$ *is a complete metric space.*

Moreover, the result also holds if we replace $\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{S};\mathrm{T})$ *with the subset* $\mathsf{C}^0_{\mathrm{unif,bdd}}(\mathrm{S};\mathrm{T})$ *of* $\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{S};\mathrm{T})$ *consisting of uniformly continuous functions.*

Proof First let us show that $d_{\mathsf{C}^0_{\mathrm{bdd}}(S;T)}$ is a metric. It is clear that $d_{\mathsf{C}^0_{\mathrm{bdd}}(S;T)}$ is symmetric and definite, so we need only verify the triangle inequality. For $f, g, h \in \mathsf{C}^0_{\mathrm{bdd}}(S;T)$ we have

$$\begin{aligned} d_{\mathsf{C}^0_{\mathrm{bdd}}(S;T)}(f, g) &= \sup\{d_T(f(x), g(x)) \mid x \in S\} \\ &\le \sup\{d_T(f(x), h(x)) + d_T(h(x), g(x)) \mid x \in S\} \\ &\le \sup\{d_T(f(x_1), h(x_1)) + d_T(h(x_2), g(x_2)) \mid x_1, x_2 \in S\} \\ &\le \sup\{d_T(f(x_1), h(x_1)) \mid x_1 \in S\} + \sup\{d_T(h(x_2), g(x_2)) \mid x_2 \in S\} \\ &= d_{\mathsf{C}^0_{\mathrm{bdd}}(S;T)}(f, h) + d_{\mathsf{C}^0_{\mathrm{bdd}}(S;T)}(h, g), \end{aligned}$$

where we have used the triangle inequality for d_T and for $f\mathbb{R}$.

Let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathsf{C}^0_{\mathrm{bdd}}(S;T)$ and for $x \in S$ define $f(x) = \lim_{j\to\infty} f_j(x)$. This pointwise limit exists since $(f_j(x))_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in T (why?), and since (T, d_T) is complete. We first claim that f is bounded. To see this, for $\epsilon > 0$, let $N \in \mathbb{Z}_{>0}$ have the property that $\mathrm{d}_{\mathsf{C}^0_{\mathrm{bdd}}(S;T)}(f, f_N) < \epsilon$. Let $R > 0$ and $y \in T$ have the property that $d_T(f(x), y) < R$ for all $x \in S$. Now choose an arbitrary $x \in S$. Then, using the triangle inequality,

$$\mathrm{d}_T(f(x), y) \le d_T(f(x), f_N(t)) + d_T(f_N(x), y) \le \epsilon + R.$$

Thus $\mathrm{image}(f) \subseteq \mathsf{B}_{d_T}(R + \epsilon, y)$, so giving the boundedness of f.

Next we claim that, for any $\epsilon > 0$, there exists $N \in \mathbb{Z}_{>0}$ such that $d_T(f(x), f_j(x)) < \epsilon$ for all $x \in S$ whenever $j \geq N$. Since $(f_j)_{j\in\mathbb{Z}_{>0}}$ is Cauchy, for any $\epsilon > 0$ there exists $N \in \mathbb{Z}_{>0}$ such that $d_T(f_j(x), f_k(x)) < \frac{\epsilon}{2}$ for all $x \in S$. For each fixed $x \in S$, we may also find $N(x) \in \mathbb{Z}_{>0}$ such that $d_T(f(x), f_j(x)) < \frac{\epsilon}{2}$ for $j \geq N(x)$. Still for fixed x, let $k = \max\{N, N(x)\}$. For $j \geq N$ we then have

$$d_T(f_j(x), f(x)) \leq d_T(f_j(x), f_k(x)) + d_T(f_k(x), f(x)) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

where we have used the triangle inequality. Since this may be done for each $x \in S$, our claim follows.

Now we use the previous claim to prove that the limit function f is continuous. By the previous claim, for any $\epsilon > 0$, there exists $N \in \mathbb{Z}_{>0}$ such that $d_T(f_N(x), f(x)) < \frac{\epsilon}{3}$ for all $x \in S$. Now fix $x_0 \in S$ and consider the $N \in \mathbb{Z}_{>0}$ just defined. By continuity of f_N, there exists $\delta > 0$ such that, if $x \in S$ satisfies $d_S(x, x_0) < \delta$, then $d_T(f_N(x), f_N(x_0)) < \frac{\epsilon}{3}$. Then, for $x \in S$ satisfying $d_S(x, x_0) < \delta$, we have

$$\begin{aligned} d_T(f(x), f(x_0)) &\leq d_T(f(x), f_N(x)) + d_T(f_N(x), f_N(x_0)) + d_T(f_N(x_0), f(x_0)) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon, \end{aligned}$$

where we have again used the triangle inequality. Since this argument is valid for any $x_0 \in S$, it follows that f is continuous.

For the final assertion, suppose that $(f_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence of bounded uniformly continuous maps from S to T. For $\epsilon > 0$ choose $N \in \mathbb{Z}_{>0}$ such that $d_T(f_N(x), f(x)) < \frac{\epsilon}{3}$ for all $x \in S$. Now choose $\delta > 0$ such that $d_S(x_1, x_2) < \delta$ implies that $d_T(f_N(x_1), f_N(x_2)) < \frac{\epsilon}{3}$. Then, again provided that $d_S(x_1, x_2) < \delta$, we have

$$\begin{aligned} d_T(f(x_1), f(x_2)) &\leq d_T(f(x_1), f_N(x_1)) + d_T(f_N(x_1), f_N(x_2)) + d_T(f_N(x_2), f(x_2)) \\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon, \end{aligned}$$

so giving uniform continuity of the limit function f. ■

An application of Proposition 1.1.33 gives the following result.

1.9.2 Corollary *If* $(\mathrm{S}, \mathrm{d_S})$ *and* $(\mathrm{T}, \mathrm{d_T})$ *are metric spaces with* $(\mathrm{T}, \mathrm{d_T})$ *complete, then* $\mathrm{C}^0_{\mathrm{unif,bdd}}(\mathrm{S}; \mathrm{T})$ *is a closed subset of* $\mathrm{C}^0_{\mathrm{bdd}}(\mathrm{S}; \mathrm{T})$ *(it is implicit that we are using the metric* $\mathrm{d}_{\mathrm{C}^0_{\mathrm{bdd}}(\mathrm{S};\mathrm{T})}$ *on* $\mathrm{C}^0_{\mathrm{bdd}}(\mathrm{S}; \mathrm{T})$*).*

1.9.3 Remark Note that the preceding results do not require that the domain (S, d_S) be complete, only that the codomain (T, d_T) be complete. •

Exercises

1.9.1

Section 1.10

Metrisable spaces

Section 1.11

Locally compact topological spaces

This section we devote to the detailed study of a certain character of topological space, those for which every point possesses a precompact neighbourhood. Just why we would devote special attention to this particular class of topological spaces is not entirely evident at this point, except to say that many topological spaces of interest have this property. We shall see specific examples as we go along. The main utility of this class of topological spaces will be seen in Section 2.12 when we discuss measures on these spaces. There it will be seen that the extra topological structure of these spaces permits an elegant characterisation of certain measures on them. However, this is getting ahead of ourselves.

Do I need to read this section? The material in this section is quite specialised, and so can be omitted until one decides to read Section 2.12. •

1.11.1 Locally compact topological groups

Chapter 2

Measure theory and integration

The theory of measure and integration we present in this chapter represents one of the most important achievements of mathematics in the twentieth century. To a newcomer to the subject or to someone coming at the material from an "applied" perspective, it can be difficult to understand *why* abstract integration provides anything of value. This is the more so if one comes equipped with the knowledge of Riemann integration as we have developed in Sections I-3.4 and II-1.6. This theory of integration appears to be entirely satisfactory. There are certainly functions that are easily described, but not Riemann integrable (see Example I-3.4.10). However, these functions typically fall into the class of functions that one will not encounter in practice, so it is not clear that they represent a serious obstacle to the viability of Riemann integration. Indeed, if one's objective is only to compute integrals, then the Riemann integral is all that is needed. The multiple volumes of tables of integrals, many of them several hundred pages in length, are all compiled using good ol' Riemann integration. *But this is not the problem that is being addressed by modern integration theory!* The theory of measure and integration we present in this chapter is intended to provide a theory whereby *spaces* of integrable functions have satisfactory properties. This confusion concerning the objectives of modern integration theory is widespread. For example, an often encountered statement is that of Richard W. Hamming (1915–1998):

> Does anyone believe that the difference between the Lebesgue and Riemann integrals can have physical significance, and that whether say, an airplane would or would not fly could depend on this difference? If such were claimed, I should not care to fly in that plane.

We are uncertain what Hamming was actually saying when he made this statement. However, it is certainly the case that this statement gets pulled out by many folks as justification for the statement that the modern theory of integration is simply not worth learning. Our view on this is that it may well be the case that this is true. If all you want to be able to do is integrate functions, then there is no need to learn the modern theory of integration. However, if you find yourself talking about spaces of integrable functions (as we shall do constantly in Volume 4 in our discussion of signal theory), then you will find yourself needing a theory of integration that is better that Riemann integration. other places?

With the above as backdrop, in Section 2.1 we discuss in detail some of the limitations of the Riemann integral. After doing this we launch into a treatment of measure theory and integration. While there is no question that the special case of Lebesgue measure and integration is of paramount importance for us, we take the approach that measure theory and integration is actually easier to understand starting from a general point of view. Thus we start with general measure theory and the corresponding general integration theory. We then specialise to Lebesgue measure and integration.

Do I need to read this chapter? The reader ought to be able to decide based on the discussion above whether they want to read this chapter. If they elect to bypass it, then they will be directed back to it at appropriate points in the sequel.

That being said, it is worth attempting to disavow a common perception about the use of measure theory and integration. There appears to be a common feeling that the theory is difficult, weird, and overly abstract. Part of this may stem from the fact that many already have a comfort level with integration via the Riemann integral, and so do not feel compelled to relearn integration theory. But the fact is that measure theory is no more difficult to learn than anything else about real analysis. •

Contents

2.1 Some motivation for abstract measure theory and integration . . . 59
2.1.1 The Jordan measure and its limitations . . . 59
2.1.2 Some limitations of the Riemann integral . . . 69
2.1.3 An heuristic introduction to the Lebesgue integral . . . 72
2.1.4 Notes . . . 73
Exercises . . . 73
2.2 Measurable sets . . . 74
2.2.1 Algebras and σ-algebras . . . 74
2.2.2 Algebras and σ-algebras generated by families of subsets . . . 77
2.2.3 Products of measurable spaces . . . 82
Exercises . . . 84
2.3 Measures . . . 86
2.3.1 Functions on families of subsets . . . 86
2.3.2 Outer measures, measures, and their relationship . . . 92
2.3.3 Complete measures and completions of measures . . . 99
2.3.4 Outer and inner measures associated to a measure . . . 101
2.3.5 Probability measures . . . 109
2.3.6 Product measures . . . 111
2.3.7 Signed measures . . . 116
2.3.8 Complex measures . . . 125
2.3.9 Vector measures . . . 128
2.3.10 Spaces of positive, signed, complex, and vector measures . . . 129

2.3.11 Notes . . . 131
Exercises . . . 131
2.4 Lebesgue measure on $\mathbb{R}$ . . . 133
2.4.1 The Lebesgue outer measure and the Lebesgue measure on $\mathbb{R}$ . . . 133
2.4.2 Borel sets in $\mathbb{R}$ as examples of Lebesgue measurable sets . . . 138
2.4.3 Further properties of the Lebesgue measure on $\mathbb{R}$ . . . 143
2.4.4 Notes . . . 149
Exercises . . . 149
2.5 Lebesgue measure on $\mathbb{R}^n$ . . . 151
2.5.1 The Lebesgue outer measure and the Lebesgue measure on $\mathbb{R}^n$ . . . 151
2.5.2 Borel sets in $\mathbb{R}^n$ as examples of Lebesgue measurable sets . . . 153
2.5.3 Further properties of the Lebesgue measure on $\mathbb{R}^n$ . . . 158
2.5.4 Lebesgue measure on $\mathbb{R}^n$ as a product measure . . . 165
2.5.5 Coverings of subsets of $\mathbb{R}^n$ . . . 167
2.5.6 The Banach–Tarski Paradox . . . 171
2.5.7 Notes . . . 183
Exercises . . . 183
2.6 Measurable functions . . . 184
2.6.1 General measurable maps and functions . . . 184
2.6.2 Measurability and operations on functions . . . 190
2.6.3 Sequences of measurable functions . . . 196
2.6.4 $\mathbb{C}$- and vector-valued measurable functions . . . 203
2.6.5 Simple functions and approximations of measurable functions . . . 206
2.6.6 Topological characterisations of convergence for sequences of measurable functions . . . 210
Exercises . . . 216
2.7 Integration on measure spaces . . . 217
2.7.1 Definition of the integral . . . 217
2.7.2 The integral and operations on functions . . . 227
2.7.3 Limit theorems . . . 232
2.7.4 Integration with respect to probability measures . . . 237
2.7.5 Topological characterisations of limit theorems . . . 241
2.7.6 Image measure and integration by image measure . . . 243
2.7.7 The integral for $\mathbb{C}$- and vector-valued functions . . . 244
2.7.8 Integration with respect to signed, complex, and vector measures . . . 249
2.7.9 Notes . . . 260
Exercises . . . 260
2.8 Integration on products . . . 262
2.8.1 The product measure by integration . . . 262
2.8.2 The integral on product spaces . . . 265
2.8.3 Fubini's Theorem . . . 267
Exercises . . . 275
2.9 The single-variable Lebesgue integral . . . 276
2.9.1 Lebesgue measurable functions . . . 276
2.9.2 The (conditional) Lebesgue integral . . . 281
2.9.3 Properties of the Lebesgue integral . . . 284

2.9.4 Swapping operations with the Lebesgue integral 290
2.9.5 Locally Lebesgue integrable functions 293
2.9.6 Absolute continuity 296
2.9.7 The Fundamental Theorem of Calculus for the Lebesgue integral 303
2.9.8 Lebesgue points 315
2.9.9 Maximal functions 318
2.9.10 The change of variables formula 318
2.9.11 Topological characterisations of the deficiencies of the Riemann integral 320
Exercises 322
2.10 The multivariable Lebesgue integral 324
2.10.1 Lebesgue measurable functions 324
2.10.2 The (conditional) Lebesgue integral 327
2.10.3 Properties of the Lebesgue integral 327
2.10.4 Swapping operations with the Lebesgue integral 327
2.10.5 The change of variables formula 327
2.10.6 Locally integrable functions 327
2.11 Differentiation of measures 329
2.11.1 Absolutely continuous measures 329
2.11.2 Singular measures 340
2.11.3 Notes 340
2.12 Measures on locally compact topological spaces 341
2.12.1 Haar measure on a locally compact topological group 341
2.12.2 The case of $\mathbb{T}$ 341
2.12.3 The case of $\mathbb{R}$ 341
2.12.4 The case of $\mathbb{R}^n$ 348

Section 2.1

Some motivation for abstract measure theory and integration

In this section we illustrate the problems with the Riemann integral when it comes to dealing with spaces of integrable functions. We do this by first deriving a "measure theory," the Jordan measure, for Riemann integration, although this is not a theory of measure that satisfies the criterion we impose in our subsequent development of measure theory. What we shall see is that the difficulty arises from the fact that the Jordan measure only behaves well when one uses *finite* unions and intersections of sets. This leads to problems with sequential operations where there is an inherent need to be able to handle countable set theoretic unions and intersections. This is illustrated clearly in Example 2.1.10. We then illustrate why this phenomenon has repercussions for the Riemann integral. The problem, as we shall see, is that limits and Riemann integration do not commute; see Example 2.1.11.

Do I need to read this section? If you have already decided to read this chapter, and you do not already understand why it is necessary to move beyond the Riemann integral, then you should read this section. •

2.1.1 The Jordan measure and its limitations

We begin our discussion of the deficiencies of the Riemann integral by considering carefully the Jordan measure, which was touched lightly upon in Section II-1.6.4. Here we develop the Jordan measure in detail before finally tearing it down.

In Section II-1.6.4 we introduced the idea of a Jordan measurable set as a set A whose characteristic function χ_A is Riemann integrable. In Theorem II-1.6.13 we showed that a bounded set A is Jordan measurable if and only if $\mathrm{bd}(A)$ has zero volume if and only if $\mathrm{bd}(A)$ has zero measure. In this section we shall consider the Jordan measure in more detail and see that it has certain clear limitations.

First let us give a characterisation of Jordan measurable sets that will echo some of the constructions that will follow in our development of general measure theory. The basic building blocks for the Jordan measure are so-called elementary sets.

2.1.1 Definition (Elementary set) A subset $E \subseteq \mathbb{R}^n$ is ***elementary*** if $E = \cup_{j=1}^k C_j$ for bounded rectangles $C_1, \dots, C_k$. •

Note that, given a elementary set E, the expression of E as a union of bounded rectangles is not unique. Moreover, since there is no restriction that the rectangles do not overlap, the following result is of interest.

2.1.2 Proposition (Elementary sets are finite unions of disjoint rectangles) *If* E *is a elementary set then there exists disjoint rectangles* $\mathrm{C}_1, \dots, \mathrm{C}_\mathrm{k}$ *such that* $\mathrm{E} = \cup_{\mathrm{j}=1}^\mathrm{k} \mathrm{C}_\mathrm{j}$.

Proof By definition we can write an elementary set as $E = \cup_{j=1}^{\tilde{k}} \tilde{C}_j$ for rectangles $\tilde{C}_1, \dots, \tilde{C}_{\tilde{k}}$. We shall prove the proposition by induction on $\tilde{k}$. The result is clearly true for $\tilde{k} = 1$. Suppose that the result is true for $\tilde{k} \in \{1, \dots, \tilde{m}\}$ and suppose that $E = \cup_{j=1}^{\tilde{m}+1} \tilde{C}_j$ and write

$$E = \left(\cup_{j=1}^{\tilde{m}} (\tilde{C}_j \cap \tilde{C}_{m+1})\right) \cup \left(\tilde{C}_{m+1} \setminus (\cup_{j=1}^{\tilde{m}} \tilde{C}_j))\right).$$

By the induction hypothesis there exists disjoint rectangles $C_1, \dots, C_l$ such that

$$\cup_{j=1}^{\tilde{m}} \tilde{C}_j = \cup_{j=1}^{l} C_j.$$

Thus

$$E = \left(\cup_{j=1}^{l} (C_j \cap \tilde{C}_{m+1})\right) \cup \left(\cup_{j=1}^{l} (\tilde{C}_{m+1} - C_j)\right).$$

Thus the result boils down to the following lemma.

1 **Lemma** *If* C *and* C′ *are bounded rectangles then* C∩C′ *is a bounded rectangle if it is nonempty and* C − C′ *is a finite union of disjoint bounded rectangles if it is nonempty.*

Proof Suppose that

$$C = I_1 \times \cdots \times I_n, \qquad C' = I_1' \times \cdots \times I_n'$$

for bounded intervals $I_1, \dots, I_n$ and $I_1', \dots, I_n'$. Note that $\boldsymbol{x} \in C \cap C'$ if and only if $x_j \in I_j \cap I_j'$, $j \in \{1, \dots, n\}$. That is,

$$C \cap C' = (I_1 \cap I_1') \times \cdots \times (I_n \cap I_n').$$

Since $(I_j \cap I_j')$, $j \in \{1, \dots, n\}$, are bounded intervals if they are nonempty, it follows that $C \cap C'$ is a bounded rectangle if it is nonempty.

Note that $C - C' = C \setminus (C \cap C')$. We may as well suppose that each of the intersections $I_j \cap I_j'$, $j \in \{1, \dots, n\}$, is a nonempty bounded interval. Then write $I_j = J_j \cup (I_j \cap I_j')$ where $J_j \cap (I_j \cap I_j') = \emptyset$. This defines a partition of C where the interval I_j is partitioned as $(J_j, I_j \cap I_j')$, $j \in \{1, \dots, n\}$. Thus this gives C as a finite disjoint union of rectangles, the subrectangles of the partition. Moreover, $C \cap C'$ corresponds exactly to the subrectangle

$$(I_1 \cap I_1') \cap \cdots \cap (I_n \cap I_n')$$

of this partition. By removing this subrectangle, we have $C - C'$ as a finite union of disjoint bounded rectangles, as desired. ▼

This completes the proof. ■

The previous result makes plausible the following definition.

2.1.3 Definition (Jordan measure of an elementary set) If $E \subseteq \mathbb{R}^n$ is an elementary set and if $E = \cup_{j=1}^{k} C_j$ for disjoint bounded rectangles $C_1, \dots, C_k$, then the ***Jordan measure*** of E is

$$\rho(E) = \sum_{j=1}^{k} \mathrm{vol}(C_j).$$

•

This definition has the possible ambiguity that it depends on writing E as a finite union of disjoint bounded rectangles, and such a union is not uniquely defined. However, one can refer to Proposition II-1.6.32 to see that the definition is, in fact independent of how this union is made.

With the Jordan measure of elementary sets, we can introduce the following concepts which we shall see arise again when we are doing “serious” measure theory.

2.1.4 Definition (Inner and outer Jordan measure) If $A \subseteq \mathbb{R}^n$ is a bounded set then

(i) the ***Jordan outer measure*** of A is

$$\rho^*(A) = \inf\{\rho(E) \mid E \text{ an elementary set containing } A\}$$

and

(ii) the ***Jordan inner measure*** of A is

$$\rho_*(A) = \sup\{\rho(E) \mid E \text{ an elementary set contained in } A\}$$ •

Note that the Jordan outer and inner measures of a bounded set always exist, provided that, for the inner measure, we allow that the empty set be thought of as an elementary set, and that we adopt the (reasonable) convention that $\rho(\emptyset) = 0$.

The following result gives a characterisation of bounded Jordan measurable sets, including some of the characterisations we have already proved in Section II-1.6.4.

2.1.5 Theorem (Characterisations of bounded Jordan measurable sets) *For a bounded subset* $A \subseteq \mathbb{R}^n$ *the following statements are equivalent:*

(i) A *is Jordan measurable;*

(ii) $\mathrm{vol}(\mathrm{bd}(A)) = 0$*;*

(iii) χ_A *is Riemann integrable;*

(iv) $\rho^*(A) = \rho_*(A)$.

Proof The equivalent of the first three statements is the content of Theorems II-1.6.13 and II-1.6.16. Thus we only prove the equivalence of the last statement with the other three.

Let C be a fat compact rectangle containing A.

First suppose that A is Jordan measurable and let $\epsilon \in \mathbb{R}_{>0}$. Since χ_A is Riemann integrable there exists a partition $\boldsymbol{P}$ of C such that

$$A_+(\chi_A, \boldsymbol{P}) - A_-(\chi_A, \boldsymbol{P}) < \epsilon.$$

Let the subrectangles of $\boldsymbol{P}$ be divided into three sorts: (1) the first sort are those subrectangles that lie within A; (2) the second sort are those that intersect A; (3) the third sort are rectangles that do not intersect A. From the definition of χ_A, $A_+(\chi_A, \boldsymbol{P})$ is the total volume of the rectangles of the third sort and $A_-(\chi_A, \boldsymbol{P})$ is the total volume of the rectangles of the first sort. Moreover, by the definitions of these rectangles,

$$\rho^*(A) \leq A_+(\chi_A, \boldsymbol{P}), \quad \rho_*(A) \geq A_-(\chi_A, \boldsymbol{P}).$$

Thus $\rho^*(A) - \rho_*(A) < \epsilon$, giving $\rho^*(A) = \rho_*(A)$ since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary.

Now suppose that $\rho^*(A) = \rho_*(A)$, let $\epsilon \in \mathbb{R}_{>0}$, and let $\overline{E}_\epsilon$ and $\underline{E}_\epsilon$ be elementary subsets of $\mathbb{R}^n$ such that $\rho(\overline{E}_\epsilon) - \rho(\underline{E}_\epsilon) < \epsilon$. Since $\overline{E}_\epsilon$ is a disjoint union of finitely many bounded rectangles there exists a partition $\overline{P}_\epsilon$ of C such that $\overline{E}_\epsilon$ is a union of subrectangles from $\overline{\boldsymbol{P}}_\epsilon$. Similarly, there exists a partition $\underline{\boldsymbol{P}}_\epsilon$ such that $\underline{E}_\epsilon$ is a union of subrectangles of $\underline{\boldsymbol{P}}_\epsilon$. Now let $\boldsymbol{P}_\epsilon$ be a partition that refines both $\overline{\boldsymbol{P}}_\epsilon$ and $\underline{\boldsymbol{P}}_\epsilon$. Then we have

$$A_+(\chi_A, \boldsymbol{P}_\epsilon) \le \rho^*(\overline{E}_\epsilon), \quad A_-(\chi_A, \boldsymbol{P}_\epsilon \ge \rho_*(\underline{E}_\epsilon),$$

which gives

$$A_+(\chi_A, \boldsymbol{P}_\epsilon) - A_-(\chi_A, \boldsymbol{P}_\epsilon) < \epsilon,$$

as desired. ■

Note that it is only the basic definition of a Jordan measurable set, i.e., that its boundary have measure zero, that is applicable to unbounded sets. However, we can still use the characterisation of bounded Jordan measurable sets to give the measure of possibly unbounded sets. For the following definition we denote by

$$C_R = [-R, R] \times \cdots \times [-R, R]$$

the rectangle centred at $\mathbf{0}$ whose sides have length $2R$ for $R \in \mathbb{R}_{>0}$.

2.1.6 Definition (Jordan measure[1]) Let $\mathscr{J}(\mathbb{R}^n)$ denote the collection of Jordan measurable sets of $\mathbb{R}^n$ and define $\rho\colon \mathscr{J}(\mathbb{R}) \to \overline{\mathbb{R}}_{\ge 0}$ by

$$\rho(A) = \lim_{R\to\infty} \rho^*(A \cap C_R),$$

noting that $A \cap C_R$ is a bounded Jordan measurable set for each $R \in \mathbb{R}_{>0}$. For $A \in \mathscr{J}(\mathbb{R}^n)$, $\rho(A)$ is the ***Jordan measure*** of A. •

Of course, by Theorem 2.1.5 we could as well have defined

$$\rho(A) = \lim_{R\to\infty} \rho_*(A \cap C_R).$$

Let us look at some examples that flesh out the definition.

2.1.7 Examples (Jordan measurable sets)

1. $\mathbb{R}^n$ is itself Jordan measurable and $\rho(\mathbb{R}^n) = \infty$.
2. Let us consider the set

$$A = \{(x_1, x_2) \in \mathbb{R}^2 \mid |x_2| \le e^{-|x_1|}\}$$

[1]The Jordan measure is *not* a measure as we shall define the notion in Section 2.3. However, it is convenient to write as if it is to get prepared for the more general and abstract development to follow.

An application of Fubini's Theorem gives

$$\int_A \mathrm{d}\boldsymbol{x} = \int_{-\infty}^{\infty}\Big(\int_{-\mathrm{e}^{-|x_1|}}^{\mathrm{e}^{-|x_1|}} \mathrm{d}x_2\Big)\mathrm{d}x_1 = 4.$$

By the definition of the Riemann integral for unbounded domains (see Definition II-1.6.22) this means that $\rho(A) = 4$. Thus unbounded domains can have finite Jordan measure. •

The following property of Jordan measures—or more precisely the fact that *only* the following result applies—is crucial to why they are actually not so useful.

2.1.8 Proposition (Jordan measurable sets are closed under finite intersections and unions) *If* $A_1, \dots, A_k \in \mathscr{J}(\mathbb{R}^n)$ *are Jordan measurable then* $\cap_{j=1}^k A_j, \cup_{j=1}^k A_j \in \mathscr{J}(\mathbb{R}^n)$.

Proof This is straightforward and we leave the details to the reader as Exercise 2.1.1. ■

Having now built up the Jordan measure and given some of its useful properties, let us now proceed to show that it has some very undesirable properties. This destruction of the Jordan measure is tightly connected with our bringing down of the Riemann integral in the next section. Sometimes, in order to understand why something is useful (in this case, the Lebesgue measure), it helps to first understand why the alternatives are *not* useful. It is with this in mind that the reader should undertake to read the remainder of this section.

The most salient question about the Jordan measure is, "What are the Jordan measurable sets?" The first thing we shall note is that there are "nice" open sets that are not Jordan measurable. This is not good, since open sets form the building blocks of the topology of $\mathbb{R}^n$.

2.1.9 Example (A regularly open non-Jordan measurable set) We shall construct a subset $A \subseteq [0, 1]$ with the following properties:

1. A is open;
2. $A = \operatorname{int}(\operatorname{cl}(A))$ (an open set with this property is called ***regularly open***);
3. A is not Jordan measurable.

The construction is involved, and will be presented with the aid of a series of lemmata. If you are prepared to take the existence of a set A as stated on faith, you can skip the details. Let us denote $I = [0, 1]$.

Any $x \in I$ can be written in the form

$$\sum_{j=1}^{\infty} \frac{a_j}{3^j}$$

for $a_j \in \{0, 1, 2\}$. This is called a ***ternary decimal expansion*** of x, and we refer the reader to Exercise I-2.4.8 for details of this construction in base 10. There is a possible nonuniqueness in such decimal expansions that arises in the following manner. If $a_1,\dots,a_k \in \{0,1,2\}$ then the numbers

$$\sum_{j=1}^{k} \frac{a_j}{3^j} + \sum_{j=k+1}^{\infty} \frac{2}{3^j} \qquad \text{and} \qquad \sum_{j=1}^{k-1} \frac{a_j}{3^j} + \frac{(a_k+1) \mod 3}{3^k} + \sum_{j=k+1}^{\infty} \frac{0}{3^k}$$

are the same, where

$$(a_k+1) \mod 3 = \begin{cases} a_k+1, & a_k \in \{0,1\}, \\ 0, & a_k = 2. \end{cases}$$

Now, for $k \in \mathbb{Z}_{>0}$, define B_k to be the subset of I for which, if $x \in B_k$ is written as

$$x = \sum_{j=1}^{\infty} \frac{a_j}{3^j},$$

then $a_j = 1$ for $j \in \{2^{k-1}+1, 2^{k-1}+2, \dots, 2^k\}$. For numbers with nonunique ternary decimal expansions, we ask that *both* representations satisfy the condition.

1 Lemma *For* $k \in \mathbb{Z}_{>0}$*,* B_k *is a disjoint union of* $3^{2^{k-1}}$ *open intervals each of length* $\frac{1}{3^{2^k}}$*.*

Proof For $\boldsymbol{a} = (a_1,\dots,a_{2^{k-1}}) \in \{0,1,2\}$ define $I_{\boldsymbol{a}}$ to be the open interval whose left endpoint is

$$\sum_{j=1}^{2^{k-1}} \frac{a_j}{3^j} + \sum_{j=2^{k-1}+1}^{2^k} \frac{1}{3^j}$$

and whose right endpoint is

$$\sum_{j=1}^{2^{k-1}} \frac{a_j}{3^j} + \sum_{j=2^{k-1}+1}^{2^k} \frac{1}{3^j} + \sum_{j=2^k+1}^{\infty} \frac{2}{3^j}.$$

There are obviously $3^{2^{k-1}}$ such intervals and each such interval has length 3^{2^k}. One can directly verify that B_k is the union of all of these intervals. ▼

Now define $B = \cup_{k=1}^{\infty} B_k$ which is, therefore, open. The sets B_k, $k \in \mathbb{Z}_{>0}$, satisfy the following.

2 Lemma *If* $l, k \in \mathbb{Z}_{>0}$ *satisfy* $l < k$ *then* $\mathrm{bd}(B_l) \cap B_k = \varnothing$*.*

Proof Let $x \in \mathrm{bd}(B_l)$. Then

$$x = \sum_{j=1}^{\infty} \frac{a_j}{3^j}$$

where either $a_j = 0$ for all $j \geq 2^l$ or $a_j = 2$ for all $j \geq 2^l$. Thus $a_{2^k} \neq 1$ and so $x \notin B_k$. ▼

Now, for $k \in \mathbb{Z}_{>0}$, we define

$$A_k = B_k - \Big(\mathrm{cl}(B_{k+1}) \cup \Big(\cup_{j=1}^{k-1} B_j\Big)\Big).$$

These sets have the following property.

3 Lemma $A_k = B_k \cap (I \setminus \mathrm{cl}(B_{k+1})) \cap_{j=1}^{k-1} (I \setminus \mathrm{cl}(B_j))$. *In particular,* A_k *is open for each* $k \in \mathbb{Z}_{>0}$.

Proof By DeMorgan's Laws we have

$$A_k = B_k \cap (I \setminus \mathrm{cl}(B_{k+1})) \cap_{j=1}^{k-1} (I \setminus B_j).$$

By Lemma 2 we have

$$B_k \cap (I \setminus B_j) = B_k \cap (I \setminus \mathrm{cl}(B_j))$$

for each $j \in \{1, \dots, k-1\}$, and the stated formula for A_k follows from this. That A_k is open follows since finite intersections of open sets are open. ▼

Thus the set $A = \cup_{k=1}^{\infty} A_k$ is open, being a union of open sets, and is contained in B since $A_k \subseteq B_k$ for each $k \in \mathbb{Z}_{>0}$.

Now, for $k \in \mathbb{Z}_{>0}$, define

$$C_k = (B_k \cap B_{k+1}) \setminus \left(\cup_{j=1}^{k-1} B_j\right).$$

By the same argument as employed in the proof of Lemma 3, Lemma 2 implies that

$$C_k = B_k \cap B_{k+1} \cap_{j=1}^{k-1} (I \setminus \mathrm{cl}(B_j))$$

and so C_k, $k \in \mathbb{Z}_{>0}$, is open, being a finite intersection of open sets. Then define the open set $C = \cup_{k=1}^{\infty} C_k$. The relationship between the sets A_l, $l \in \mathbb{Z}_{>0}$, and C_k, $k \in \mathbb{Z}_{>0}$.

4 Lemma *For each* $l, k \in \mathbb{Z}_{>0}$, $A_l \cap C_k = \emptyset$.

Proof First suppose that $l = k$. By definition we have

$$A_k \subseteq I \cap \mathrm{cl}(B_{k+1}), \quad C_k \subseteq B_{k+1}$$

which immediately gives $A_k \cap C_k = \emptyset$. Now suppose that $l < k$. Again by definition we have

$$A_l \subseteq B_l, \quad C_k \subseteq I \setminus \mathrm{cl}(B_l),$$

giving $A_l \cap C_k = \emptyset$. Finally, for $l > k$ we have

$$A_l \subseteq I \setminus \mathrm{cl}(B_k), \quad C_k \subseteq B_k,$$

giving $A_l \cap C_k = \emptyset$. ▼

The following lemma then gives a relationship between A and C.

5 Lemma cl(A) = I \ C.

Proof By Lemma 4 we have $A \cap C = \emptyset$. That is, $A \subseteq I \setminus C$. Since $I \setminus C$ is closed it follows that $\mathrm{cl}(A) \subseteq I \setminus C$. The difficult bit if the converse inclusion. Let $x \in I \setminus C$. We consider three cases.

1. $x \in \cup_{k=1}^{\infty} \mathrm{bd}(A_k)$: In this case, since $\mathrm{bd}(A_k) \subseteq \mathrm{cl}(A_k) \subseteq \mathrm{cl}(A)$ for each $k \in \mathbb{Z}_{>0}$ it immediately follows that $x \in \mathrm{cl}(A)$.
2. $x \notin B$: In this case we can write

$$x = \sum_{j=1}^{\infty} \frac{a_j}{3^j}.$$

Since $x \notin B$, for every $k \in \mathbb{Z}_{>0}$ there exists $j \in \{2^{k-1}+1, \dots, 2^k\}$ such that $a_j \neq 1$. Now define a sequence $(y_k)_{k \in \mathbb{Z}_{>0}}$ by asking that $y_k = \sum_{j=1}^{\infty} \frac{b_j}{3^j}$ with

$$b_j = \begin{cases} 1, & j \in \{2^{k-1}+1, \dots, 2^k\}, \\ a_j, & \text{otherwise.} \end{cases}$$

We then have $|x - y_k| \le \frac{1}{3^{2^k}}$ (cf. the proof of Lemma 1) and so the sequence $(y_k)_{k \in \mathbb{Z}_{>0}}$ converges to x. Moreover, by construction,

$$y_k \in B_k, \; y_k \notin B_1 \cup \dots \cup B_{k-1}, \; y_k \notin \mathrm{cl}(B_{k+1}).$$

(Only the last of these statements is potentially not obvious. It, however, follows from the characterisation of B_{k+1}, and by implication the characterisation of $\mathrm{cl}(B_{k+1})$, obtained in Lemma 1.) That is, by definition of A_k, $y_k \in A_k \subseteq A$. Thus $x \in \mathrm{cl}(A)$ by Proposition I-2.5.18.
3. $x \notin \cup_{k=1}^{\infty} \mathrm{bd}(A_k)$ and $x \in B$: Let $k \in \mathbb{Z}_{>0}$ be the least index for which $x \in B_k$. Since $x \notin C$ it follows that $x \notin C_k$ and so $x \notin B_{k+1}$ and $x \notin B_j$ for $j \in \{1, \dots, k-1\}$. We also have $x \notin \mathrm{bd}(A_{k+1})$. We claim that $\mathrm{bd}(B_{k+1}) \subseteq \mathrm{bd}(A_{k+1})$. Indeed, for each $m \in \mathbb{Z}_{>0}$, by construction of the set A_m, $\mathrm{bd}(A_m)$ consists of those ternary decimal expansions $\sum_{j=1}^{\infty} \frac{a_j}{3^j}$ having the following three properties:
   (a) for $l < m$ there exists $j \in \{2^{l-1}+1, \dots, 2^l\}$ such that $a_j \neq 1$;
   (b) $a_j = 1$ for each $j \in \{2^{m-1}+1, \dots, 2^m\}$;
   (c) there exists $j \in \{2^m+1, \dots, 2^{m+1}\}$ such that $a_j \neq 1$.

   Using this characterisation, and by referring to the description of B_m in Lemma 1, we then see that, indeed, $\mathrm{bd}(B_{k+1}) \subseteq \mathrm{bd}(A_{k+1})$. Thus we conclude that $x \notin \mathrm{bd}(B_{k+1})$. Then, by definition of A_k, $x \in A_k \subseteq A \subseteq \mathrm{cl}(A)$, as desired. ▼

   We also then have

$$\mathrm{int}(\mathrm{cl}(A)) \subseteq \mathrm{int}(I \setminus C) = I \setminus \mathrm{cl}(C).$$

That is, $\mathrm{int}(\mathrm{cl}(A)) \cap \mathrm{cl}(C) = \emptyset$.

Now we can prove that A has one of the properties we set out for it to have.

6 Lemma A = int(cl(A)).

Proof Since $A \subseteq \mathrm{cl}(A)$ we have $A = \mathrm{int}(A) \subseteq \mathrm{int}(\mathrm{cl}(A))$. It is thus the converse inclusion we must prove.

We first claim that $\mathrm{int}(\mathrm{cl}(A)) \subseteq B$. Suppose that $x \notin B$. Let us write

$$x = \sum_{j=1}^{\infty} \frac{a_j}{3^j}.$$

Since $x \notin B$, for every $k \in \mathbb{Z}_{>0}$ there exists $j \in \{2^{k-1}+1, \dots, 2^k\}$ such that $a_j \neq 1$. Now, for $k \in \mathbb{Z}_{>0}$, define

$$y_k = \sum_{j=1}^{\infty} \frac{c_j}{3^j}$$

where

$$c_j = \begin{cases} a_j, & j \leq 2^{k-1}, \\ 1, & j > 2^{k-1}. \end{cases}$$

Then one can directly verify that

$$y_k \in B_k, \; y_k \in B_{k+1}, \; y_k \notin B_1 \cup \dots \cup B_{k-1}.$$

Thus, by definition of C_k, $y \in C_k$. Moreover, $|x - y_k| \leq \frac{1}{3^{2^{k-1}}}$ and so the sequence $(y_k)_{k\in\mathbb{Z}_{>0}}$ converges to x. Therefore, since $y_k \in C$ for each $k \in \mathbb{Z}_{>0}$, $x \in \mathrm{cl}(C)$ by Proposition I-2.5.18. Thus $x \notin \mathrm{int}(\mathrm{cl}(A))$ by our computation just preceding the statement of the lemma.

Now, if $x \in \mathrm{int}(\mathrm{cl}(A))$ then $x \in B$ and we let $k \in \mathbb{Z}_{>0}$ be the least integer for which $x \in B_k$. We claim that $x \notin \mathrm{cl}(B_{k+1})$. We suppose that $x \in \mathrm{cl}(B_{k+1})$ and arrive at a contradiction. There are two possibilities.

1. $x \in \mathrm{bd}(B_{k+1})$: First of all, using the characterisation of the sets B_l, $l \in \mathbb{Z}_{>0}$, from Lemma 1 and using the definition of the sets C_l, $l \in \mathbb{Z}_{>0}$, we deduce that $\mathrm{bd}(B_l) \subseteq \mathrm{bd}(C_l)$ for each $l \in \mathbb{Z}_{>0}$. Therefore, if $x \in \mathrm{bd}(B_{k+1})$ then $x \in \mathrm{bd}(C_{k+1}) \subseteq \mathrm{cl}(C_{k+1}) \subseteq \mathrm{cl}(C)$. This contradicts the fact that $x \in \mathrm{int}(\mathrm{cl}(A))$ and that $\mathrm{int}(\mathrm{cl}(A)) \cap \mathrm{cl}(C) = \emptyset$.
2. $x \in B_{k+1}$: In this case $x \in B_k \cap B_{k+1} \subseteq C_k \subseteq \mathrm{cl}(C)$, and we arrive at a contradiction, just as in the previous case.

Thus we have shown that $x \notin \mathrm{cl}(B_{k+1})$. But, by definition, this implies that $x \in A_k \subseteq A$, since $x \notin \cup_{j=1}^{k-1} B_j$ by definition of k. ▼

Finally, to complete the example, we need only show that A is not Jordan measurable. To do this, we shall show that $\mathrm{bd}(A)$ does not have measure zero. In fact, we shall show that $\mathrm{bd}(A)$ has positive measure, but this relies on actually knowing what "measure" means; it means Lebesgue measure. We shall subsequently carefully define Lebesgue measure, but all we need to know here is that (1) the

Lebesgue measure of a countable collection of intervals is less than or equal to the sum of the lengths of the intervals and (2) the Lebesgue measure of two disjoint sets is the sum of their measures. Let us denote by $\lambda(S)$ the Lebesgue measure of a set S. We note that, by Lemma 1,

$$\lambda(B_k) = 3^{2^{k-1}} \frac{1}{3^{2^k}} = \frac{1}{3^{2^{k-1}}}.$$

Thus

$$\lambda(B) \le \sum_{k=1}^{\infty} \frac{1}{3^{2^{k-1}}} < \sum_{j=1}^{\infty} \frac{1}{3^j} = \frac{1}{2}$$

(how would you compute this sum?). Since $A \subseteq B$ we also have $\lambda(A) < \frac{1}{2}$. Therefore, since $\mathrm{cl}(A) = I \setminus C$ and since $C \subseteq B$,

$$\lambda(A) + \lambda(\mathrm{bd}(A)) = \lambda(\mathrm{cl}(A)) \ge \lambda(I \setminus B) = 1 - \lambda(B) > \tfrac{1}{2} > \lambda(A),$$

which gives $\lambda(\mathrm{bd}(A)) \in \mathbb{R}_{>0}$, so A is not Jordan measurable. •

This is a rather complicated example. However, it says something important. It says that not all open sets are Jordan measurable, not even "nice" open sets (and regularly open sets are thought of as being pretty darn nice). Open subsets of $\mathbb{R}$ are pretty easy to describe. Indeed, by Proposition I-2.5.6 such sets are countable unions of open intervals. If one has an open subset of $[0,1]$, such as the one just constructed, this means that the total lengths of these intervals should sum to a finite number of value at most one. This should, if the world is right, be the "measure" of this open set. However, the example indicates that this is just not so if "measure" means "Jordan measure." We shall see that it *is* so for the Lebesgue measure.

In Proposition 2.1.8 we stated that finite unions and intersections of Jordan measurable sets are Jordan measurable. This no longer holds if one replaces "finite" with "countable."

2.1.10 Examples (Jordan measurable sets are not closed under countable intersections and unions)

1. Let $(q_j)_{j\in\mathbb{Z}_{>0}}$ be an enumeration of the rational numbers in the interval $[0,1]$. For each $j \in \mathbb{Z}_{>0}$ the set $\{q_j\}$ is Jordan measurable with Jordan measure 0. Thus, by Proposition 2.1.8 any finite union of these sets is also Jordan measurable with Jordan measure 0. However, the set $\cup_{j=1}^{\infty}\{q_j\}$ is not Jordan measurable by Example I-3.4.10.
2. Let $(q_j)_{j\in\mathbb{Z}_{>0}}$ be as above and define $A_j = [0,1] \setminus \{q_j\}$. Then A_j is Jordan measurable and has Jordan measure 1. Moreover, any finite intersection of these sets is Jordan measurable with Jordan measure 1. However, $\cap_{j=1}^{\infty} A_j$ is equal to the set of irrational numbers in the interval $[0,1]$ and is not Jordan measurable in exactly the same manner as the set $\cup_{j=1}^{\infty}\{q_j\}$ is not Jordan measurable, cf. Example I-3.4.10. •

A good question is, "Who cares if the Jordan measure is not closed under countable intersections and unions?" This is not obvious, but it certainly underlies, for example, the failure of the set in Example 2.1.9 to be Jordan measurable. Somewhat more precisely, this failure of the Jordan measure to not be closed under countable set theoretic operations is the reason why the Riemann integral does not have nice properties with respect to sequences, as we now explain explicitly.

2.1.2 Some limitations of the Riemann integral

In this section we simply give an example that illustrates a fundamental defect with the theory of Riemann integration. The problem we illustrate is the lack of commutativity of limits and Riemann integration. The reader may wish to refer to the discussion in Section II-1.7.3 concerning the Monotone and Dominated Convergence Theorems for the Riemann integral to get more insight into this.

2.1.11 Example (Limits do not commute with Riemann integration) First recall from Example I-3.4.10 that the function $f\colon [0,1] \to \mathbb{R}$ defined as taking value 1 on rational numbers, and value 0 on irrational numbers is not Riemann integrable. It is legitimate to inquire why one should care if such a degenerate function should be integrable. The reason is that the function f arises as the limit of a sequence of integrable functions. We explain this in the following example.

By Exercise I-2.1.3, the set of rational numbers in $[0,1]$ is countable. Thus it is possible to write the set of rational numbers as $(q_j)_{j\in\mathbb{Z}_{>0}}$. For each $j \in \mathbb{Z}_{>0}$ define $f_j\colon [0,1] \to \mathbb{R}$ by

$$f_j(x) = \begin{cases} 1, & x = q_j, \\ 0, & \text{otherwise}. \end{cases}$$

One may readily verify that f_j is Riemann integrable for each $j \in \mathbb{Z}_{>0}$, and that the value of the Riemann integral is zero. By Proposition I-3.4.22 it follows that for $k \in \mathbb{Z}_{>0}$, the function

$$g_k = \sum_{j=1}^{k} f_j$$

is Riemann integrable, and that the value of the Riemann integral is zero. Thus we have

$$\lim_{k\to\infty} g_k(x) = f(x), \quad \lim_{k\to\infty} \int_a^b g_k(x)\,\mathrm{d}x = 0,$$

the left limit holding for each $x \in [0,1]$ (i.e., the sequence $(g_k)_{k\in\mathbb{Z}_{>0}}$ converges pointwise to f). It now follows that

$$\lim_{k\to\infty} \int_a^b g_k(x)\,\mathrm{d}x \neq \int_a^b \lim_{k\to\infty} g_n(x)\,\mathrm{d}x.$$

Indeed, the expression on the right hand side is not even defined! •

It is perhaps not evident immediately why this lack of commutativity of limits and integrals is in any way debilitating, particularly given the inherent silliness of the functions in the preceding example. We shall not really understand the reasons for this in any depth until we consider in detail convergence theorems in Section 2.7.3.

Let us illustrate some additional "features" of the Riemann integral, the exact context for which we will only consider in detail in Chapter 3 (see, in particular, Sections 3.8.7 and 3.8.8). We shall freely use the language and notation from that chapter. Let us define

$$\mathsf{R}^{(1)}([0,1];\mathbb{R}) = \{f\colon [0,1] \to \mathbb{R} \mid f \text{ is Riemann integrable}\},$$

and recall from Propositions I-3.4.22 and I-3.4.25 that $\mathsf{R}^{(1)}([0,1];\mathbb{R})$ is a $\mathbb{R}$-vector space. Now let us define a seminorm $\|\cdot\|_1$ on $\mathsf{R}^{(1)}([0,1];\mathbb{R})$ by

$$\|f\|_1 = \int_0^1 |f(x)|\,\mathrm{d}x.$$

This fails to be a norm because there exist nonzero Riemann integrable functions f on $[0,1]$ for which $\|f\|_1 = 0$ (for example, take f to be a function that has a nonzero value at a single point in $[0,1]$). To produce a normed vector space we denote

$$Z([0,1];\mathbb{R}) = \{f \in \mathsf{R}^{(1)}([0,1];\mathbb{R}) \mid \|f\|_1 = 0\},$$

and by Theorem 3.1.8 note that

$$\mathsf{R}^1([0,1];\mathbb{R}) \triangleq \mathsf{R}^{(1)}([0,1];\mathbb{R})/Z([0,1];\mathbb{R})$$

is a normed vector space when equipped with the norm

$$\|f + Z([0,1];\mathbb{R})\|_1 \triangleq \|f\|_1,$$

where we use the abuse of notation of using the same symbol $\|\cdot\|_1$ for the norm. Note that $\mathsf{R}^1([0,1];\mathbb{R})$ is a vector space, not of functions, but of equivalence classes of functions under the equivalence relation that two Riemann integrable functions are equivalent when the absolute value of their difference has zero integral.

The crux of the matter is now the following result, the proof of which makes free use of concepts in this chapter that we have not yet introduced.

2.1.12 Proposition (The normed vector space of Riemann integrable functions is not complete) *The $\mathbb{R}$-normed vector space $(\mathsf{R}^1([0,1];\mathbb{R}), \|\cdot\|_1)$ is not complete.*

Proof Let $(q_j)_{j\in\mathbb{Z}_{>0}}$ be an enumeration of the rational numbers in $[0,1]$. Let $\ell \in (0,1)$ and for $j \in \mathbb{Z}_{>0}$ define

$$I_j = [0,1] \cap (q_j - \tfrac{\ell}{2^{j+1}}, q_j + \tfrac{\ell}{2^{j+1}})$$

to be the interval of length $\frac{\ell}{2^j}$ centred at q_j. Then define $A_k = \cup_{j=1}^k I_j$, $k \in \mathbb{Z}_{>0}$, and $A = \cup_{j\in\mathbb{Z}_{>0}} A_j$. Also define $f_k = \chi_{A_k}$, $k \in \mathbb{Z}_{>0}$, and $f = \chi_A$ be the characteristic functions of A_k and A, respectively. Note that A_k is a union of a finite number of intervals and so f_k is Riemann integrable for each $k \in \mathbb{Z}_{>0}$. However, we claim that f is not Riemann integrable. Indeed, the characteristic function of a set is Riemann integrable if and only the boundary of the set has measure zero; this is a direct consequence of Lebesgue's theorem stating that a function is Riemann integrable if and only if its set of discontinuities has measure zero (Theorem I-3.4.11). Note that since $\mathrm{cl}(\mathbb{Q} \cap [0,1]) = [0,1]$ we have

$$[0,1] = \mathrm{cl}(A) = A \cup \mathrm{bd}(A).$$

Thus

$$\lambda([0,1]) \leq \lambda(A) + \lambda(\mathrm{bd}(A)).$$

Since

$$\lambda(A) \leq \sum_{j=1}^{\infty} \lambda(I_j) \leq \ell,$$

it follows that $\lambda(\mathrm{bd}(A)) \geq 1 - \ell \in \mathbb{R}_{>0}$. Thus f is not Riemann integrable, as claimed.

Next we show that if $g\colon [0,1] \to \mathbb{R}$ satisfies $[g] = [f]$, then g is not Riemann integrable. To show this, it suffices to show that g is discontinuous on a set of positive measure. We shall show that g is discontinuous on the set $g^{-1}(0) \cap \mathrm{bd}(A)$. Indeed, let $x \in g^{-1}(0) \cap \mathrm{bd}(A)$. Then, for any $\epsilon \in \mathbb{R}_{>0}$ we have $(x-\epsilon, x+\epsilon) \cap A \neq \emptyset$ since $x \in \mathrm{bd}(A)$. Since $(x-\epsilon, x+\epsilon) \cap A$ is a nonempty open set, it has positive measure. Therefore, since f and g agree almost everywhere, there exists $y \in (x-\epsilon, x+\epsilon) \cap A$ such that $g(y) = 1$. Since this holds for every $\epsilon \in \mathbb{R}_{>0}$ and since $g(x) = 0$, it follows that g is discontinuous at x. Finally, it suffices to show that $g^{-1}(0) \cap \mathrm{bd}(A)$ has positive measure. But this follows since $\mathrm{bd}(A) = f^{-1}(0)$ has positive measure and since f and g agree almost everywhere.

We claim that the sequence $([f_k])_{k \in \mathbb{Z}_{>0}}$ is Cauchy in $\mathsf{R}^1([0,1];\mathbb{R})$. Let $\epsilon \in \mathbb{R}_{>0}$. Note that $\sum_{j=1}^{\infty} \lambda(I_j) \leq \ell$. This implies that there exists $N \in \mathbb{Z}_{>0}$ such that $\sum_{j=k+1}^{m} \lambda(I_j) < \epsilon$ for all $k, m \geq N$. Now note that for $k, m \in \mathbb{Z}_{>0}$ with $m > k$, the functions f_k and f_m agree except on a subset of $I_{k+1} \cup \cdots \cup I_m$. On this subset, f_m has value 1 and f_k has value 0. Thus

$$\int_0^1 |f_m(x) - f_k(x)|\,\mathrm{d}x \leq \lambda(I_{k+1} \cup \cdots \cup I_m) \leq \sum_{j=k+1}^{m} \lambda(I_j).$$

Thus we can choose $N \in \mathbb{Z}_{>0}$ sufficiently large that $\|f_m - f_k\|_1 < \epsilon$ for $k, m \geq N$. Thus the sequence $([f_k])_{k \in \mathbb{Z}_{>0}}$ is Cauchy, as claimed.

We next show that the sequence $([f_k])_{k \in \mathbb{Z}_{>0}}$ converges to $[f]$ in $\mathsf{L}^1([0,1];\mathbb{R})$ (see Section 3.8.7). Since the sequence $([f - f_k])_{k \in \mathbb{Z}_{>0}}$ is in the subset

$$\{[f] \in \mathsf{L}^1([0,1];\mathbb{R}) \mid |f(x)| \leq 1 \text{ for almost every } x \in [0,1]\},$$

by the Dominated Convergence Theorem, Theorem 2.7.28, it follows that

$$\lim_{k\to\infty} \|f - f_k\|_1 = \int_I \lim_{k\to\infty} |f - f_k|\,\mathrm{d}\lambda = 0.$$

This gives us the desired convergence of $([f_k])_{k \in \mathbb{Z}_{>0}}$ to $[f]$ in $\mathsf{L}^1([0,1];\mathbb{R})$. However, above we showed that $[f] \notin \mathsf{R}^1([0,1];\mathbb{R})$. Thus the Cauchy sequence $([f_k])_{k \in \mathbb{Z}_{>0}}$ in

$\mathsf{R}^1([0,1];\mathbb{R})$ is not convergent in $\mathsf{R}^1([0,1];\mathbb{R})$, giving the desired incompleteness of $(\mathsf{R}^1([0,1];\mathbb{R}), \|\cdot\|_1)$. ■

It should be emphasised that all of the above "problems" are not so much one with using the Riemann integral to compute the integral of a given function, as to use the notion of a Riemann integrable function in stating theorems, particularly those where limits are involved. This problem is taken care of by the Lebesgue integral, to which we turn our attention in Section 2.7.1 in a general setting for integration.

2.1.3 An heuristic introduction to the Lebesgue integral

Before we get to the powerful general theory, we provide in this section an alternate way of thinking about the integral of a function defined on a compact interval. The idea is an essentially simple one. One defines the Riemann integral by taking increasingly finer partitions of the independent variable axis, where on each subinterval of the partition the approximation is constant. For the Lebesgue integral, it turns out that what one should do instead is partition the *dependent* variable axis.

The reader should not treat the following discussion as the definition of the Lebesgue integral. This definition will be provided precisely in the general framework of Section 2.7.1. But let us be a little precise about the idea. We let $I = [a, b]$ and let $f: I \to \mathbb{R}$ be a positive bounded function. This means that $f(I) \subset [0, M]$ for some $M \in \mathbb{R}_{>0}$. We then let P be a partition of $[0, M]$ with endpoints $(y_0 = 0, y_1, \dots, y_{n-1}, y_n = M)$. Corresponding to this partition let us define sets

$$A_j = \{x \in I \mid f(x) \in [y_{j-1}, y_j)\},$$

and then define

$$f_P = \sum_{j=1}^{n} y_j \chi_{A_j}.$$

The function f_P is called a ***simple function***, as we shall see in Section 2.7, and approximates f from below as depicted in Figure 2.1. The integral of one of these approximations is then

$$\int_a^b f_P(x)\,\mathrm{d}x = \sum_{j=1}^{n} y_j \lambda(A_j),$$

where $\mu(A_j)$ is the "size" of the set A_j. If A_j is a union on intervals, then $\mu(A_j)$ is the sum of the lengths of these intervals. More generally, we shall define

$$\lambda(A) = \inf\Big\{\sum_{j=1}^{\infty} |b_j - a_j| \;\Big|\; A \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} (a_j, b_j)\Big\}$$

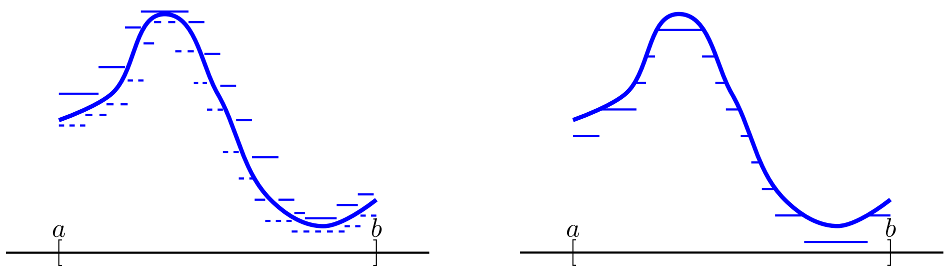


Figure 2.1 The idea behind the Riemann integral (left) and the Lebesgue integral (right)

for a very general class of subsets of $\mathbb{R}$. To define the integral of f we take

$$\text{“}\int_a^b f(x)\,\mathrm{d}x\text{”} = \sup\left\{ \int_a^b f_P(x)\,\mathrm{d}x \;\middle|\; P \text{ a partition of } [0, M] \right\}.$$

The idea is that by taking successively finer partitions of the image of f one can better approximate f.

For the elementary function we are depicting in Figure 2.1, the two approaches appear to be much the same. However, the power of the Lebesgue integral rests in its use of the “size” of the sets A_j on which the approximating function is constant. For step functions, these sets are always intervals, and it is there that the problems arise. By allowing the sets A_j to be quite general, the Lebesgue integral becomes a very powerful tool. However, it does need some buildup, and the first thing to do is remove the quotes from “size.”

2.1.4 Notes

Example 2.1.9 comes from [Börger 1999]. Frink, Jr. [1933] connects the Riemann integral and the Jordan measure.

[Cohn 2013, Halmos 1974]

Exercises

2.1.1 Prove Proposition 2.1.8.

Section 2.2

Measurable sets

The construction of the integral we provide in this chapter proceeds along different lines than does the usual construction of the Riemann integral. In Riemann integration one typically jumps right in with a function and starts constructing step function approximations, etc. However, one could also define the Riemann integral by first defining the Jordan measure as in Section 2.1.1, and then using this as the basis for defining the integral. But the idea is still that one uses step functions as approximations. In the theory for integration that we develop here, a crucial difference is the sort of functions we use to approximate the functions we wish to integrate. The construction of these approximating functions, in turn, rests on some purely set theoretic constructions that play the rôle of the Jordan measure (which, we remind the reader, is not a measure in the general sense we define in this chapter) in Riemann integration. In this section we provide the set theoretic constructions needed to begin this abstract form of integration theory.

Do I need to read this section? If you are reading this chapter, then this is where the technical material begins. If you are only interested in learning about Lebesgue measure, you can get away with knowing the definition of "measurable space" and then proceeding directly to Section 2.4. However, in Section 2.4 we will freely refer to things proved in this section, so as you read Section 2.4 you will eventually end up reading many things in this section anyway. •

2.2.1 Algebras and σ-algebras

The idea we develop in this section and the next is that of a means of measuring the size of a set in a general way. What one first must do is provide a suitable collection of sets whose size one wishes to measure. One's first reaction to this programme might be, "Why not measure the size of *all* subsets?" The answer to this question is not immediately obvious, and we shall say some things about this as we go along. For the moment, the reader should simply trust that the definitions we give have been thought over pretty carefully by lots of pretty smart people, and so are possibly "correct."[2]

2.2.1 Definition (Algebra, σ-algebra, measurable space) For a set X, a subset of subsets $\mathscr{A} \subseteq \mathbf{2}^X$ is an ***algebra***[3] if

(i) $X \in \mathscr{A}$,

(ii) $A \in \mathscr{A}$ implies $X \setminus A \in \mathscr{A}$, and

[2] That being said, *never* stop being a skeptic!

[3] Also sometimes called a ***field***.

(iii) $\cup_{j=1}^{k} A_j \in \mathscr{A}$ for any finite family $(A_1, \dots, A_k)$ of subsets,

and a ***σ-algebra***[4] on X if

(iv) $X \in \mathscr{A}$,

(v) $A \in \mathscr{A}$ implies $X \setminus A \in \mathscr{A}$, and

(vi) $\cup_{j \in \mathbb{Z}_{>0}} A_j \in \mathscr{A}$ for any countable family $(A_j)_{j \in \mathbb{Z}_{>0}}$ of subsets.

A pair $(X, \mathscr{A})$ is called a ***measurable space*** if $\mathscr{A}$ is a σ-algebra on X and elements of $\mathscr{A}$ are called ***$\mathscr{A}$-measurable***. •

We shall mainly be concerned with σ-algebras, although the notion of an algebra is occasionally useful even if one is working with σ-algebras.

2.2.2 Remark (Why are the axioms for a measurable space as they are?) In Remark 1.2.2 we attempted to justify why the axioms for a topological space are as they are. For topological spaces this justification is facilitated by the fact that most readers will already know about open subsets of Euclidean space. For readers new to measure theory, it is less easy to justify the axioms of a measurable space. In particular, why is it that we require *countable* unions of measurable subsets to be measurable? Why not finite unions (as with algebras) or arbitrary unions? Why not intersections instead of unions? The reason for this, at its core, is that we wish for the theory we develop to have useful properties with respect to sequential limit operations, and such limit operations have an intrinsic countability in them due to sequences being countable sets. It may be difficult to see just why this is important at this point, but this is the justification. •

Let us give some simple examples of σ-algebras.

2.2.3 Examples (Algebras, σ-algebras)

1. It is clear that the power set $\mathbf{2}^X$ of a set X is a σ-algebra.
2. For a set X, the collection of subsets $\{\emptyset, X\}$ is a σ-algebra.
3. For a set X the collection of subsets

$$\mathscr{A} = \{A \subseteq X \mid A \text{ or } X \setminus A \text{ is countable}\}$$

is a σ-algebra.

4. The collection $\mathscr{J}(\mathbb{R}^n)$ of Jordan measurable subsets of $\mathbb{R}^n$ (see Definition II-1.6.12) is an algebra by Proposition 2.1.8 and not a σ-algebra by virtue of Example 2.1.10. •

The following result records some useful properties of σ-algebras.

[4] Also sometimes called a ***σ-field***.

2.2.4 Proposition (Properties of σ-algebras) *Let $\mathscr{A}$ be a σ-algebra on X. The following statements hold:*

(i) $\varnothing \in \mathscr{A}$;

(ii) *if* $A_1, \dots, A_k \in \mathscr{A}$ *then* $\cup_{j=1}^k A_j \in \mathscr{A}$;

(iii) $\cap_{j \in \mathbb{Z}_{>0}} A_j \in \mathscr{A}$ *for any countable collection* $(A_j)_{j \in \mathbb{Z}_{>0}}$ *of subsets;*

(iv) *if* $A_1, \dots, A_k \in \mathscr{A}$ *then* $\cap_{j=1}^k A_j \in \mathscr{A}$.

Moreover, condition (vi) in Definition 2.2.1 can be equivalently replaced with condition (iii) above.

Proof (i) Since $X \in \mathscr{A}$ we must have $X \setminus X = \varnothing \in \mathscr{A}$.

(ii) We define a countable collection $(B_j)_{j \in \mathbb{Z}_{>0}}$ of subsets in $\mathscr{A}$ by

$$B_j = \begin{cases} A_j, & j \in \{1, \dots, k\}, \\ \varnothing, & j > k, \end{cases}$$

and the assertion now follows since

$$\cup_{j=1}^k A_j = \cup_{j \in \mathbb{Z}_{>0}} B_j \in \mathscr{A}.$$

(iii) This follows from De Morgan's Laws (Proposition I-1.1.5):

$$\bigcap_{j \in \mathbb{Z}_{>0}} A_j = X \setminus \Big(\bigcup_{j \in \mathbb{Z}_{>0}} (X \setminus A_j) \Big).$$

Since $X \setminus A_j \in \mathscr{A}$ it follows that $\cup_{j \in \mathbb{Z}_{>0}} (X \setminus A_j) \in \mathscr{A}$ since $\mathscr{A}$ is a σ-algebra. Therefore $X \setminus \big(\cup_{j \in \mathbb{Z}_{>0}} (X \setminus A_j)\big) \in \mathscr{A}$ and so this part of the result follows.

(iv) This follows again from De Morgans's Laws, along with part (ii).

The final assertion of the proposition follows from De Morgans's Laws, as can be gleaned from the arguments used in the proof of part (iii), along with a similar argument, swapping the rôles of union and intersection. ■

The following corollary is now obvious.

2.2.5 Corollary (σ-algebras are algebras) *A σ-algebra $\mathscr{A}$ on a set X is also an algebra on X.*

Another construction that is sometimes useful is the restriction of a measurable space $(X, \mathscr{A})$ to a subset $A \subseteq X$. If A is measurable, then there is a natural σ-algebra induced on A.

2.2.6 Proposition (Restriction of a σ-algebra to a measurable subset) *Let $(X, \mathscr{A})$ be a measurable space, let $A \in \mathscr{A}$, and define $\mathscr{A}_A \subseteq \mathbf{2}^A$ by*

$$\mathscr{A}_A = \{B \cap A \mid B \in \mathscr{A}\}.$$

Then $(A, \mathscr{A}_A)$ is a measurable space.

Proof We need to show that $\mathscr{A}_A$ is a σ-algebra on A. Clearly $A \in \mathscr{A}_A$ since $A = X \cap A$ and $X \in \mathscr{A}$. Also, since $A \setminus (B \cap A) = (X \setminus B) \cap A$ by Proposition I-1.1.5, it follows that $A \setminus (B \cap A) \in \mathscr{A}_A$ for $B \cap A \in \mathscr{A}_A$. Suppose that $(B_j \cap A)_{j \in \mathbb{Z}_{>0}}$ is a countable family of sets in $\mathscr{A}_A$. Since $\cup_{j \in \mathbb{Z}_{>0}} (B_j \cap A) = (\cup_{j \in \mathbb{Z}_{>0}} B_j) \cap A$ by Proposition I-1.1.7 it follows that $\cup_{j \in \mathbb{Z}_{>0}} (B_j \cap A) \in \mathscr{A}_A$. ■

2.2.2 Algebras and σ-algebras generated by families of subsets

It is often useful to be able to indirectly define algebras and σ-algebras by knowing that they contain a certain family of subsets. This is entirely analogous to the manner in which one defines a topology by a basis or subbasis; see Section 1.2.2.

Let us begin with the construction of a σ-algebra containing a family of subsets.

2.2.7 Proposition (σ-algebras generated by subsets) *If X is a set and if $\mathscr{S} \subseteq \mathbf{2}^X$ then there exists a unique σ-algebra $\sigma(\mathscr{S})$ with the following properties:*

(i) $\mathscr{S} \subseteq \sigma(\mathscr{S})$;

(ii) if $\mathscr{A}$ is any σ-algebra for which $\mathscr{S} \subseteq \mathscr{A}$ then $\sigma(\mathscr{S}) \subseteq \mathscr{A}$.

Proof We let $\mathscr{P}_{\mathscr{S}}$ be the collection of all σ-algebras with the property that if $\mathscr{A} \in \mathscr{P}_{\mathscr{S}}$ then $\mathscr{S} \subseteq \mathscr{A}$. Note that $\mathscr{P}_{\mathscr{S}}$ is nonempty since $\mathbf{2}^X \subseteq \mathscr{P}_{\mathscr{S}}$. We then define

$$\sigma(\mathscr{S}) = \bigcap\{\mathscr{A} \mid \mathscr{A} \in \mathscr{P}_{\mathscr{S}}\}.$$

If $\sigma(\mathscr{S})$ is a σ-algebra then clearly it satisfies the conditions of the statement of the result. Let us then show that $\sigma(\mathscr{S})$ is a σ-algebra. Since each element of $\mathscr{P}_{\mathscr{S}}$ is a σ-algebra we have $X \in \mathscr{A}$ whenever $\mathscr{A} \in \mathscr{P}_{\mathscr{S}}$. Therefore $X \in \sigma(\mathscr{S})$. If $A \in \sigma(\mathscr{S})$ it follows that $A \in \mathscr{A}$ whenever $\mathscr{A} \in \mathscr{P}_{\mathscr{S}}$. Therefore $X \setminus A \in \mathscr{A}$ whenever $\mathscr{A} \in \mathscr{P}_{\mathscr{S}}$, showing that $X \setminus A \in \sigma(\mathscr{S})$. Finally, if $(A_j)_{j\in\mathbb{Z}_{>0}} \subseteq \sigma(\mathscr{S})$ then $(A_j)_{j\in\mathbb{Z}_{>0}} \subseteq \mathscr{A}$ whenever $\mathscr{A} \in \mathscr{P}_{\mathscr{S}}$. Therefore, $\cup_{j\in\mathbb{Z}_{>0}} A_j \in \mathscr{A}$ whenever $\mathscr{A} \in \mathscr{P}_{\mathscr{S}}$. Therefore, $\cup_{j\in\mathbb{Z}_{>0}} A_j \in \sigma(\mathscr{S})$. ■

The previous proof applies equally well to algebras. Moreover, it is possible to give a more or less explicit characterisation of the smallest algebra containing a given collection of subsets. This is not possible for σ-algebras, cf. the proof of Theorem 2.2.14.

2.2.8 Proposition (Algebras generated by subsets) *If X is a set and if $\mathscr{S} \subseteq \mathbf{2}^X$ then there exists a unique algebra $\sigma_0(\mathscr{S})$ with the following properties:*

(i) $\mathscr{S} \subseteq \sigma_0(\mathscr{S})$;

(ii) if $\mathscr{A}$ is any algebra for which $\mathscr{S} \subseteq \mathscr{A}$ then $\sigma_0(\mathscr{S}) \subseteq \mathscr{A}$.

Moreover, $\sigma_0(\mathscr{S})$ is the set of finite unions of sets of the form $S_1 \cap \cdots \cap S_k$, where each of the sets $S_1, \dots, S_k$ is either in $\mathscr{S}$ or its complement is in $\mathscr{S}$.

Proof The existence of $\sigma_0(\mathscr{S})$ can be argued just as in the proof of Proposition 2.2.7. To see that $\sigma_0(\mathscr{S})$ admits the explicit stated form, let $\overline{\mathscr{S}}$ be the collection sets of the stated form. We first claim that $\overline{\mathscr{S}}$ is an algebra. To see that $X \in \overline{\mathscr{S}}$, let $S \in \mathscr{S}$ and note that $X \setminus S \in \overline{\mathscr{S}}$. Thus $X = S \cup (X \setminus S) \in \overline{\mathscr{S}}$. If $T \in \overline{\mathscr{S}}$ then we show that $X \setminus T \in \overline{\mathscr{S}}$ as follows. Note that $T = T_1 \cup \cdots \cup T_k$ where, for each $j \in \{1, \dots, k\}$,

$$T_j = \bigcap_{l_j=1}^{m_j} S_{jl_j}, \qquad S_{jl_j} \in \mathscr{S} \text{ or } X \setminus S_{jl_j} \in \mathscr{S},\ l_j \in \{1, \dots, m_j\}.$$

Let us for brevity denote $A = \{1, \dots, m_1\} \times \cdots \times \{1, \dots, m_k\}$. Then, using De Morgan's Laws and Proposition I-1.1.7,

$$X \setminus T = X \setminus \Big(\bigcup_{j=1}^{k}\Big(\bigcap_{l_j=1}^{m_j} S_{jl_j}\Big)\Big) = \bigcap_{j=1}^{k}\Big(X \setminus \Big(\bigcap_{l_j=1}^{m_j} S_{jl_j}\Big)\Big)$$
$$= \bigcap_{j=1}^{k}\Big(\bigcup_{l_j=1}^{m_j} X \setminus S_{jl_j}\Big) = \bigcup_{(l_1,\dots,l_k)\in A}\Big(\bigcup_{j=1}^{k} X \setminus S_{jl_j}\Big),$$

which then gives $X \setminus T \in \overline{\mathscr{S}}$. It is obvious that finite unions of sets from $\overline{\mathscr{S}}$ are in $\overline{\mathscr{S}}$, which shows that $\overline{\mathscr{S}}$ is an algebra, as desired. Moreover, it is clear that $\mathscr{S} \subseteq \overline{\mathscr{S}}$.

Now suppose that $\mathscr{A}$ is an algebra for which $\mathscr{S} \subseteq \mathscr{A}$. Since $\mathscr{A}$ is an algebra this implies that $X \setminus S \in \mathscr{A}$ for $S \in \mathscr{S}$ and, by Exercise 2.2.1, that $S_1 \cap \cdots \cap S_k \in \mathscr{A}$ for every collection $S_1, \dots, S_k$ for which either $S_j \in \mathscr{S}$ or $X \setminus S_j \in \mathscr{S}$ for each $j \in \{1, \dots, k\}$. Thus $\overline{\mathscr{S}} \subseteq \mathscr{A}$ and so $\overline{\mathscr{S}} = \sigma_0(\mathscr{S})$, as desired. ■

This gives the following result as a special case.

2.2.9 Corollary (The algebra generated by a finite collection of sets) *Let X be a set and let $S_1, \dots, S_k \subseteq X$ be a finite family of subsets. Then $\sigma_0(S_1, \dots, S_k)$ is the collection of finite unions of sets of the form $T_1 \cap \cdots \cap T_m$ where, for each $j \in \{1, \dots, m\}$, either $T_j \in \{S_1, \dots, S_k\}$ or $X \setminus T_j \in \{S_1, \dots, S_k\}$.*

The point is that you can specify any collection of subsets and define an algebra or σ-algebra associated with this collection in a natural way, i.e., by demanding that the conditions of an algebra or a σ-algebra hold. The preceding results makes sense of the next definition.

2.2.10 Definition (Algebras and σ-algebras generated by subsets) If X is a set and $\mathscr{S} \subseteq \mathbf{2}^X$, the algebra $\sigma_0(\mathscr{S})$ (resp. σ-algebra $\sigma(\mathscr{S})$) of Proposition 2.2.8 (resp. Proposition 2.2.7) is the ***algebra generated by $\mathscr{S}$*** (resp. ***σ-algebra generated by $\mathscr{S}$***). •

We now provide an alternative description of the σ-algebra generated by a collection of subsets. This description relies on the following concept.

2.2.11 Definition (Monotone class) For a set X, a ***monotone class*** on X is a collection $\mathscr{M} \subseteq \mathbf{2}^X$ of subsets of X with the following properties:

(i) $\cup_{j\in\mathbb{Z}} A_j \in \mathscr{M}$ for every family $(A_j)_{j\in\mathbb{Z}_{>0}}$ of subsets from $\mathscr{M}$ such that $A_j \subseteq A_{j+1}$ for every $j \in \mathbb{Z}_{>0}$;

(ii) $\cap_{j\in\mathbb{Z}_{>0}} A_j \in \mathscr{M}$ for every family $(A_j)_{j\in\mathbb{Z}_{>0}}$ of subsets from $\mathscr{M}$ such that $A_j \supseteq A_{j+1}$ for every $j \in \mathbb{Z}_{>0}$. •

Let us illustrate how the conditions of a monotone class can be used to relate algebras and σ-algebras.

2.2.12 Proposition (Algebras that are σ-algebras) *Let X be a set and let $\mathscr{A}$ be an algebra. If either*

(i) *$\cup_{j\in\mathbb{Z}_{>0}}A_j \in \mathscr{A}$ for every family $(A_j)_{j\in\mathbb{Z}_{>0}}$ of subsets from $\mathscr{A}$ for which $A_j \subseteq A_{j+1}$, $j\in\mathbb{Z}_{>0}$, or*

(ii) *$\cap_{j\in\mathbb{Z}_{>0}}A_j \in \mathscr{A}$ for every family $(A_j)_{j\in\mathbb{Z}_{>0}}$ of subsets from $\mathscr{A}$ for which $A_j \supseteq A_{j+1}$, $j\in\mathbb{Z}_{>0}$,*

then $\mathscr{A}$ is a σ-algebra.

Proof We clearly have $X\in\mathscr{A}$ and $X\setminus A\in\mathscr{A}$ for $A\in\mathscr{A}$.

Now suppose that the first of the two conditions in the proposition holds and let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a countable collection of subsets from $\mathscr{A}$. For $k\in\mathbb{Z}_{>0}$ define $B_k\in\cup_{j=1}^k A_j$. Since $\mathscr{A}$ is an algebra, $B_k\in\mathscr{A}$ for $k\in\mathbb{Z}_{>0}$. Moreover, we clearly have $B_k\subseteq B_{k+1}$ for each $k\in\mathbb{Z}_{>0}$ and $\cup_{j\in\mathbb{Z}_{>0}}A_j=\cup_{k\in\mathbb{Z}_{>0}}B_k$. Therefore, by assumption, $\cup_{j\in\mathbb{Z}_{>0}}A_j\in\mathscr{A}$, and so $\mathscr{A}$ is a σ-algebra.

Finally suppose that the second of the two conditions in the proposition holds and let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a countable collection of subsets from $\mathscr{A}$. Define $B_k=X\setminus\cup_{j=1}^k A_j$. Since $\mathscr{A}$ is an algebra we have $B_k\in\mathscr{A}$ for $k\in\mathbb{Z}_{>0}$. We also have $B_k\supseteq B_{k+1}$ for each $k\in\mathbb{Z}_{>0}$ and $\cap_{k=1}^\infty = X\setminus\cup_{j\in\mathbb{Z}_{>0}}A_j$. Thus $X\setminus\cup_{j\in\mathbb{Z}_{>0}}A_j\in\mathscr{A}$, and so $\cup_{j\in\mathbb{Z}_{>0}}A_j\in\mathscr{A}$ since $\mathscr{A}$ is an algebra. Thus $\mathscr{A}$ is a σ-algebra. ■

Next we state our alternative characterisation of the σ-algebra generated by an algebra of subsets. It is perhaps not immediately apparent why the result is useful, but we shall use it in our discussion of product measures in Section 2.8.1.

2.2.13 Theorem (Monotone Class Theorem) *Let X be a set and let $\mathscr{S}\subseteq\mathbf{2}^X$. Then there exists a unique monotone class $m(\mathscr{S})$ on X such that*

(i) *$\mathscr{S}\subseteq m(\mathscr{S})$ and*

(ii) *if $\mathscr{M}$ is any monotone class on X for which $\mathscr{S}\subseteq\mathscr{M}$ then $m(\mathscr{S})\subseteq\mathscr{M}$.*

Moreover, if $\mathscr{S}$ is an algebra then $m(\mathscr{S})=\sigma(\mathscr{S})$.

Proof We let $\mathscr{P}_{\mathscr{S}}$ be the collection of monotone classes with the property that if $\mathscr{M}\in\mathscr{P}_{\mathscr{S}}$ then $\mathscr{S}\subseteq\mathscr{M}$. Since $X\in\mathscr{P}_{\mathscr{S}}$ it follows that $\mathscr{P}_{\mathscr{S}}$ is not empty. We define

$$m(\mathscr{S})=\bigcap\{\mathscr{M}\mid \mathscr{M}\in\mathscr{P}_{\mathscr{S}}\}.$$

It is clear that $\mathscr{S}\subseteq m(\mathscr{S})$. Moreover, it is also clear that if $\mathscr{M}$ is a monotone class containing $\mathscr{S}$ then $m(\mathscr{S})\subseteq\mathscr{M}$. It remains to show that $m(\mathscr{S})$ is a monotone class. Let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a family of subsets from $m(\mathscr{S})$ such that $A_j\subseteq A_{j+1}$ for $j\in\mathbb{Z}_{>0}$. Since $A_j\in\mathscr{M}$ for each $j\in\mathbb{Z}_{>0}$ and $\mathscr{M}\in\mathscr{P}_{\mathscr{S}}$ it follows that $\cup_{j\in\mathbb{Z}_{>0}}A_j\in\mathscr{M}$ for every $\mathscr{M}\in\mathscr{P}_{\mathscr{S}}$. Thus $\cup_{j\in\mathbb{Z}_{>0}}A_j\in m(\mathscr{S})$. Similarly, let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a family of subsets from $m(\mathscr{S})$ for which $A_j\supseteq A_{j+1}$ for $j\in\mathbb{Z}_{>0}$. Since $A_j\in\mathscr{M}$ for every $j\in\mathbb{Z}_{>0}$ and $\mathscr{M}\in\mathscr{P}_{\mathscr{S}}$ it follows that $\cap_{j\in\mathbb{Z}_{>0}}A_j\in\mathscr{M}$ for every $\mathscr{M}\in\mathscr{P}_{\mathscr{S}}$. Thus $\cap_{j\in\mathbb{Z}_{>0}}A_j\in m(\mathscr{S})$, showing that $m(\mathscr{S})$ is indeed a monotone class.

Now let us prove the final assertion of the theorem, supposing that $\mathscr{S}$ is an algebra. We claim that $m(\mathscr{S})$ is an algebra. Indeed, let $S\in\mathscr{S}$ and define

$$\mathscr{M}_S=\{A\in m(\mathscr{S})\mid S\cap A, S\cap(X\setminus A),(X\setminus S)\cap A\in m(\mathscr{S})\}.$$

We claim that $\mathscr{M}_S$ is a monotone class. Indeed, let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a family of subsets from $\mathscr{M}_S$ such that $A_j \subseteq A_{j+1}$ for $j \in \mathbb{Z}_{>0}$. Thus

$$S \cap A_j, S \cap (X \setminus A_j), (X \setminus S) \cap A_j \in m(\mathscr{S}), \qquad j \in \mathbb{Z}_{>0}.$$

Then, using Propositions I-1.1.5 and I-1.1.7,

$$\begin{aligned}
&S \cap \left(\cup_{j\in\mathbb{Z}_{>0}} A_j\right) = \cup_{j\in\mathbb{Z}_{>0}} (S \cap A_j),\\
&S \cap \left(X \setminus \left(\cup_{j\in\mathbb{Z}_{>0}} A_j\right)\right) = S \cap \left(\cap_{j\in\mathbb{Z}_{>0}} X \setminus A_j\right) = \cap_{j\in\mathbb{Z}_{>0}} S \cap (X \setminus A_j),\\
&(X \setminus S) \cap \left(\cup_{j\in\mathbb{Z}_{>0}} A_j\right) = \cup_{j\in\mathbb{Z}_{>0}} (X \setminus S) \cap A_j.
\end{aligned}$$

Since

$$S \cap A_j \subseteq S \cap A_{j+1}, \quad S \cap (X \setminus A_j) \supseteq S \cap (X \setminus A_{j+1}), \quad (X \setminus S) \cap A_j \subseteq (X \setminus S) \cap A_{j+1},$$

for $j \in \mathbb{Z}_{>0}$, we conclude that

$$S \cap \left(\cup_{j\in\mathbb{Z}_{>0}} A_j\right), S \cap \left(X \setminus \left(\cup_{j\in\mathbb{Z}_{>0}} A_j\right)\right), (X \setminus S) \cap \left(\cup_{j\in\mathbb{Z}_{>0}} A_j\right) \in m(\mathscr{S}),$$

and so $\cup_{j\in\mathbb{Z}_{>0}} A_j \in \mathscr{M}_S$. A similarly styled argument gives $\cap_{j\in\mathbb{Z}_{>0}} \in \mathscr{M}_S$ for a countable family $(A_j)_{j\in\mathbb{Z}_{>0}}$ of subsets from $\mathscr{M}_S$ satisfying $A_j \supseteq A_{j+1}$, $j \in \mathbb{Z}_{>0}$. Thus $\mathscr{M}_S$ is indeed a monotone class.

We claim that $\mathscr{M}_S = m(\mathscr{S})$. To see this we first claim that $\mathscr{S} \subseteq \mathscr{M}_S$. Indeed, if $A \in \mathscr{S}$ then

$$S \cap A, S \cap (X \setminus A), (X \setminus S) \cap A \in \mathscr{S} \subseteq m(\mathscr{S})$$

since $\mathscr{S}$ is a field. Thus $\mathscr{M}_S$ is a monotone class containing $\mathscr{S}$ and so $m(\mathscr{S}) \subseteq \mathscr{M}_S$. Since $\mathscr{M}_S \subseteq m(\mathscr{S})$ by definition, we conclude that $\mathscr{M}_S = m(\mathscr{S})$. Note that $S \in \mathscr{S}$ is arbitrary in this construction.

Next we claim that $\mathscr{M}_S$, and so $m(\mathscr{S})$, is an algebra. First of all, since $X \in \mathscr{S}$ by virtue of $\mathscr{S}$ being an algebra, we have

$$X \in \mathscr{S} \subseteq m(\mathscr{S}) = \mathscr{M}_S.$$

Also, if $A \in \mathscr{M}_S$ we have

$$A \in \mathscr{M}_S \implies A \in \mathscr{M}_X \implies X \cap (X \setminus A) \in m(\mathscr{S}) \implies X \setminus A \in m(\mathscr{S}) = \mathscr{M}_S.$$

Also, let $A, B \in \mathscr{M}_S$. Then

$$A, B \in \mathscr{M}_S \implies A, B \in \mathscr{M}_A \implies B \cap A \in m(\mathscr{S}) = \mathscr{M}_S.$$

Thus the intersection of sets from $\mathscr{M}_S$ lies in $\mathscr{M}_S$. This means that if $A, B \in \mathscr{M}_S$ then

$$X \setminus A, X \setminus B \in \mathscr{M}_S \implies (X \setminus A) \cap (X \setminus B) = X \setminus (A \cup B) \in \mathscr{M}_S,$$

implying that $A \cup B \in \mathscr{M}_S$. Thus pairwise unions of sets from $\mathscr{M}_X$ are in $\mathscr{M}_S$. An elementary induction then gives $\cap_{j=1}^k A_j \in \mathscr{M}_S$ for every family of subsets $(A_1, \dots, A_j)$ from $\mathscr{M}_S$. This shows that $\mathscr{M}_S = m(\mathscr{S})$ is an algebra.

Since $\mathscr{M}_S$ is a monotone class it is a σ-algebra by Proposition 2.2.12. Thus $\sigma(\mathscr{S}) \subseteq \mathscr{M}_S = m(\mathscr{S})$. Moreover, $\sigma(\mathscr{S})$ is a monotone class by the properties of a σ-algebra and by Proposition 2.2.4. Since $\mathscr{S} \subseteq \sigma(\mathscr{S})$ we conclude from Proposition 2.2.12 that $m(\mathscr{S}) \subseteq \sigma(\mathscr{S})$, giving $m(\mathscr{S}) = \sigma(\mathscr{S})$, as desired. ■

The following "fun fact" about the σ-algebra generated by a collection of subsets is useful to understand how big this σ-algebra is. We will use this result in Proposition 2.4.13 to compare the cardinalities of Borel and Lebesgue measurable sets. Recall that $\aleph_0 = \operatorname{card}(\mathbb{Z}_{\geq 0})$.

2.2.14 Theorem (Cardinality of the σ-algebra generated by a collection of subsets) *Let* X *be a set and let* $\mathscr{S} \subseteq \mathbf{2}^X$ *be such that* $\varnothing \in \mathscr{S}$ *and that* $\operatorname{card}(\mathscr{S}) \geq 2$. *Then* $\operatorname{card}(\sigma(\mathscr{S})) \leq \operatorname{card}(\mathscr{S})^{\aleph_0}$.

Proof Let $\aleph_1$ be the smallest uncountable cardinal number (the cardinal number that the Continuum Hypothesis asserts is equal to $\operatorname{card}(\mathbb{R})$). Define $\mathscr{S}_0 = \mathscr{S}$. For a cardinal number $c < \aleph_1$ we shall use Transfinite Induction (Theorem I-1.5.14) to define $\mathscr{S}_c$ as follows. Suppose that $\mathscr{S}_{c'}$ has been defined for a cardinal number c' such that $0 < c' < c$. Then define $\mathscr{S}_c$ to be the collection of sets of the form $\cup_{j\in\mathbb{Z}_{>0}} A_j$ where either A_j or $X \setminus A_j$ is an element of the family $\cup_{0\leq c'<c}\mathscr{S}_{c'}$ of subsets of X. We claim that $\cup_{0\leq c<\aleph_1}\mathscr{S}_c = \sigma(\mathscr{S})$.

We first prove by Transfinite Induction that $\cup_{0\leq c<\aleph_1}\mathscr{S}_c \subseteq \sigma(\mathscr{S})$. Clearly $\mathscr{S}_0 \subseteq \sigma(\mathscr{S})$. Suppose that $\mathscr{S}_{c'} \subseteq \sigma(\mathscr{S})$ for $0 \leq c' < c < \aleph_1$. Then let $\cup_{j\in\mathbb{Z}_{>0}} A_j \in \mathscr{S}_c$ for set A_j such that either A_j or $X \setminus A_j$ are in the family $\cup_{0\leq c'<c}\mathscr{S}_{c'}$ of subsets of X. It follows from the induction hypothesis that $A_j, X \setminus A_j \in \sigma(\mathscr{S})$. Thus $\cup_{j\in\mathbb{Z}_{>0}} A_j \in \sigma(\mathscr{S})$ since a σ-algebra is closed under countable unions. Therefore, $\mathscr{S}_c \in \sigma(\mathscr{S})$ and so we conclude from Transfinite Induction that $\cup_{0\leq c<\aleph_1}\mathscr{S}_c \subseteq \sigma(\mathscr{S})$.

To prove that $\cup_{0\leq c<\aleph_1}\mathscr{S}_c = \sigma(\mathscr{S})$ it now suffices to show that $\cup_{0\leq c<\aleph_1}\mathscr{S}_c$ is a σ-algebra since it contains $\mathscr{S}$ and since $\sigma(\mathscr{S})$ is the smallest σ-algebra containing $\mathscr{S}$. Since $\varnothing \in \mathscr{S}$ we have

$$X = (X \setminus \varnothing) \cup \varnothing \cup \varnothing \cdots \in \mathscr{S}_1,$$

and so $X \in \cup_{0\leq c<\aleph_1}\mathscr{S}_c$. Now suppose that $A \in \cup_{0\leq c<\aleph_1}\mathscr{S}_c$ so that $A \in \mathscr{S}_{c_0}$ for some c_0 satisfying $0 \leq c_0 \leq \aleph_1$. For $c_1 > c_0$ it then holds that

$$X \setminus A = (X \setminus A) \cup (X \setminus A) \cup \cdots \in \mathscr{S}_{c_1},$$

and so $(X \setminus A) \in \cup_{0\leq c<\aleph_1}\mathscr{S}_c$. Finally, let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a countable family of subsets from $\cup_{0\leq c<\aleph_1}\mathscr{S}_c$. For $j \in \mathbb{Z}_{>0}$ let c_j be a cardinal number satisfying $0 \leq c_j < \aleph_1$ and $A_j \in \mathscr{S}_{c_j}$. Since $\aleph_1$ is uncountable it cannot be a countable union of countable sets (by Proposition I-1.7.16) and since each of the cardinal numbers c_j, $j \in \mathbb{Z}_{>0}$, are countable, it follows that there exists a cardinal number c_∞ such that $0 \leq c_\infty < \aleph_1$ and such that $c_j < c_\infty$. Then $\cup_{j\in\mathbb{Z}_{>0}} A_j \in \mathscr{S}_{c_\infty} \subseteq \cup_{0\leq c<\aleph_1}\mathscr{S}_c$, completing the proof that $\cup_{0\leq c<\aleph_1}\mathscr{S}_c = \sigma(\mathscr{S})$.

We now prove by Transfinite Induction that $\operatorname{card}(\mathscr{S}_c) \leq \operatorname{card}(\mathscr{S})^{\aleph_0}$ for every cardinal number c satisfying $0 \leq c \leq \aleph_1$. Certainly $\operatorname{card}(\mathscr{S}_0) \leq \operatorname{card}(\mathscr{S})^{\aleph_0}$. Now suppose that c is a cardinal number satisfying $0 \leq c < \aleph_1$ and suppose that $\operatorname{card}(\mathscr{S}_{c'}) \leq \operatorname{card}(\mathscr{S})^{\aleph_0}$ for cardinals c' satisfying $0 \leq c' < c$. Since c is countable it follows that

$$\operatorname{card}(\cup_{0\leq c'<c}\mathscr{S}_{c'}) \leq \aleph_0 \operatorname{card}(\mathscr{S})^{\aleph_0} = \operatorname{card}(\mathscr{S})^{\aleph_0}$$

by Theorem I-1.7.17, Exercises I-1.7.4 and I-1.7.5, and since $\operatorname{card}(\mathscr{S}) \geq 2$. Now, considering the definition of $\mathscr{S}_c$ we see that

$$\operatorname{card}(\mathscr{S}_c) = 2\operatorname{card}(\cup_{0\leq c'<c}\mathscr{S}_{c'}) \leq \operatorname{card}(\mathscr{S})^{\aleph_0},$$

as claimed.

From this we deduce that

$$\mathrm{card}(\sigma(\mathscr{S})) = \mathrm{card}(\cup_{0\le c<\aleph_1}\mathscr{S}_c) \le \mathrm{card}(\mathscr{S})^{\aleph_0}\aleph_1 = \mathrm{card}(\mathscr{S})^{\aleph_0},$$

using Theorem I-1.7.17 and the fact that $\mathrm{card}(\mathscr{S})^{\aleph_0} \ge \aleph_1$ since $\mathrm{card}(\mathscr{S}) \ge 2$ and using Exercises I-1.7.4 and I-1.7.5. ■

2.2.3 Products of measurable spaces

The development of measure theory on products is a little more challenging than, say, the development of topology on products. In this section we introduce the basic tool for studying measure theory for products by considering the products of sets equipped with algebras or σ-algebras of subsets.

We begin by considering products of sets equipped with algebras of subsets.

2.2.15 Definition (Measurable rectangles) For sets $X_1,\dots,X_k$ with algebras $\mathscr{A}_j \subseteq \mathbf{2}^{X_j}$, $j \in \{1,\dots,k\}$, a ***measurable rectangle*** is a subset

$$A_1 \times \cdots \times A_k \subseteq X_1 \times \cdots \times X_k$$

where $A_j \in \mathscr{A}_j$, $j \in \{1,\dots,k\}$. The set of measurable rectangles is denoted by $\mathscr{A}_1 \times \cdots \times \mathscr{A}_k$. •

By Corollary 2.2.5 the preceding definition can be applied to the case when each of the collections of subsets $\mathscr{A}_1,\dots,\mathscr{A}_k$ is a σ-algebra.

The following property of the set of measurable rectangles is then useful.

2.2.16 Proposition (Finite unions of measurable rectangles form an algebra) *For sets $\mathrm{X}_1,\dots,\mathrm{X}_\mathrm{k}$ with algebras $\mathscr{A}_\mathrm{j} \subseteq \mathbf{2}^{\mathrm{X}_\mathrm{j}}$, $\mathrm{j} \in \{1,\dots,\mathrm{k}\}$, the set of finite unions of sets from $\mathscr{A}_1\times\cdots\times\mathscr{A}_\mathrm{k}$ is an algebra on $\mathrm{X}_1\times\cdots\times\mathrm{X}_\mathrm{k}$, and is necessarily the algebra $\sigma_0(\mathscr{A}_1\times\cdots\times\mathscr{A}_\mathrm{k})$.*

Proof Clearly $X_1\times\cdots\times X_k$ is a measurable rectangle. Next, for measurable rectangles $A_1\times\cdots\times A_k$ and $B_1\times\cdots\times B_k$ we have

$$(A_1\times\cdots\times A_k)\cap(B_1\times\cdots\times B_k) = (A_1\cap B_1)\times\cdots\times(A_k\cap B_k).$$

This shows that the intersection of two measurable rectangles is a measurable rectangle. From Proposition I-1.1.4 we can then conclude that the intersection of two finite unions of measurable rectangles is a finite union of measurable rectangles. Next let $A_1\times\cdots\times A_k$ be a measurable rectangle and note that

$$(X_1\times\cdots\times X_k)\setminus(A_1\times\cdots\times A_k)$$

is the union of sets of the form $B_1\times\cdots\times B_k$ where $B_j \in \{A_j, X_j\setminus A_j\}$ and where at least one of the sets B_j is not equal to A_j. That is to say, the complement of a measurable rectangle is a finite union of measurable rectangles. By De Morgan's Laws we then conclude that the complement of a finite union of measurable rectangles is a finite union of measurable rectangles. By Exercise 2.2.1 this proves that the set of finite unions of measurable rectangles is an algebra. Moreover, if $\mathscr{A}$ is any σ-algebra containing $\mathscr{A}_1 \times \cdots \times \mathscr{A}_k$ then $\mathscr{A}$ must necessarily contain finite unions of measurable rectangles. Thus $\mathscr{A}$ is contained in the set of finite unions of measurable rectangles. By Proposition 2.2.8 this means that the algebra of finite unions of measurable rectangles is the algebra generated by $\mathscr{A}_1 \times \cdots \times \mathscr{A}_k$. ■

The principal object of interest to us will be the σ-algebra generated by the measurable rectangles. The following result gives a characterisation of this σ-algebra.

2.2.17 Proposition (The σ-algebra generated by the algebra of measurable rectangles) *For sets* $X_1, \dots, X_k$ *with algebras* $\mathscr{A}_j \subseteq \mathbf{2}^{X_j}$, $j \in \{1, \dots, k\}$, *we have*

$$\sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_k) = \sigma(\sigma(\mathscr{A}_1) \times \cdots \times \sigma(\mathscr{A}_k)).$$

Proof Clearly we have

$$\sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_k) \subseteq \sigma(\sigma(\mathscr{A}_1) \times \cdots \times \sigma(\mathscr{A}_k)).$$

To prove the opposite inclusion it suffices to show that

$$\sigma(\mathscr{A}_1) \times \cdots \times \sigma(\mathscr{A}_k) \subseteq \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_k)$$

since this will imply that the σ-algebra of the left-hand side is contained in the right-hand side. We prove the preceding inclusion by induction on k. For $k = 1$ the assertion is trivial. So suppose that for $k = m$ we have

$$\sigma(\mathscr{A}_1) \times \cdots \times \sigma(\mathscr{A}_m) \subseteq \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_m),$$

and suppose that we have a set X_{m+1} with an algebra $\mathscr{A}_{m+1}$. Fix $A_j \in \sigma(\mathscr{A}_j)$, $j \in \{1, \dots, m\}$, and define

$$\sigma'(\mathscr{A}_{m+1}) = \{A \in \sigma(\mathscr{A}_{m+1}) \mid A_1 \times \cdots \times A_m \times A \in \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_{m+1})\}.$$

We claim that $\sigma'(\mathscr{A}_{m+1})$ is a σ-algebra on X_{m+1}. Certainly $X_{m+1} \in \sigma'(\mathscr{A}_{m+1})$ since

$$A_1 \times \cdots \times A_m \times X_{m+1} \in \mathscr{A}_1 \times \cdots \times \mathscr{A}_m \times \mathscr{A}_{m+1} \subseteq \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_{m+1}).$$

Let $A \in \sigma'(\mathscr{A}_{m+1})$. Then we note that

$$A_1 \times \cdots \times A_m \times (X_{m+1} \setminus A) = (A_1 \times \cdots \times A_m \times X_{m+1}) \setminus (A_1 \times \cdots \times A_m \times A).$$

By assumption,

$$A_1 \times \cdots \times A_m \times A \in \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_m \times \mathscr{A}_{m+1})$$

from which we conclude that

$$(A_1 \times \cdots \times A_m \times X_{m+1}) \setminus (A_1 \times \cdots \times A_m \times A) \in \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_m \times \mathscr{A}_{m+1}).$$

Thus $X_{m+1} \setminus A \in \sigma'(sA_{m+1})$. Finally, if $(B_j)_{j\in\mathbb{Z}_{>0}}$ is a countable family of subsets from $\sigma'(\mathscr{A}_{m+1})$ we have

$$A_1 \times \cdots \times A_m \times \Big(\bigcup_{j\in\mathbb{Z}_{>0}} B_j \Big) = \bigcup_{j\in\mathbb{Z}_{>0}} A_1 \times \cdots \times A_m \times B_j \in \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_m \times \mathscr{A}_{m+1}).$$

Thus $\cup_{j\in\mathbb{Z}_{>0}} B_j \in \sigma'(\mathscr{A}_{m+1})$, showing that $\sigma'(\mathscr{A}_{m+1})$ is indeed a σ-algebra. Since $\mathscr{A}_{k+1} \subseteq \sigma'(\mathscr{A}_{m+1})$ and since $\sigma'(\mathscr{A}_{m+1}) \subseteq \sigma(\mathscr{A}_{m+1})$, we conclude that $\sigma(\mathscr{A}_{m+1}) = \sigma(\mathscr{A}_{m+1})$. This shows that

$$\sigma(\mathscr{A}_1) \times \cdots \times \sigma(\mathscr{A}_m) \times \sigma(\mathscr{A}_{m+1}) \subseteq \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_m \times \mathscr{A}_{m+1}),$$

as desired. ■

The following property of the product of σ-algebras is useful.

2.2.18 Proposition (Intersections of measurable sets with factors in products are measurable) *Let* $(\mathrm{X_j}, \mathscr{A}_\mathrm{j})$, $\mathrm{j} \in \{1, \dots, \mathrm{k}\}$, *be measurable spaces. For* $\mathrm{A} \in \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_\mathrm{k})$, *for* $\mathrm{j} \in \{1, \dots, \mathrm{k}\}$, *and for* $\mathrm{x_j} \in \mathrm{X_j}$ *define*

$$\mathrm{A}_{\mathrm{x_j}} = \{(\mathrm{x}_1, \dots, \mathrm{x}_{\mathrm{j}-1}, \mathrm{x}_{\mathrm{j}+1}, \dots, \mathrm{x_k}) \in \mathrm{X}_1 \times \cdots \times \mathrm{X}_{\mathrm{j}-1} \times \mathrm{X}_{\mathrm{j}+1} \times \cdots \times \mathrm{X_k} \mid (\mathrm{x}_1, \dots, \mathrm{x}_{\mathrm{j}-1}, \mathrm{x_j}, \mathrm{x}_{\mathrm{j}+1}, \dots, \mathrm{x_k}) \in \mathrm{A}\}.$$

Then $\mathrm{A}_{\mathrm{x_j}} \in \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_{\mathrm{j}-1} \times \mathscr{A}_{\mathrm{j}+1} \times \cdots \times \mathscr{A}_\mathrm{k})$.

Proof Let $\mathscr{F}_{x_j}$ be the subsets $A \subseteq X_1 \times \cdots \times X_k$ with the property that $A_{x_j} \in \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_{j-1} \times \mathscr{A}_{j+1} \times \cdots \times \mathscr{A}_k)$. We claim that if $B_j \in X_j$, $j \in \{1, \dots, k\}$, then $B_1 \times \cdots \times B_k \in \mathscr{F}_{x_j}$. Indeed, we have $A_{x_j} = B_1 \times \cdots \times B_{j-1} \times B_{j+1} \times \cdots \times B_k$ if $x_j \in B_j$ and $A_{x_j} = \varnothing$ otherwise. We also claim that $\mathscr{F}_{x_j}$ is a σ-algebra. We have just shown that $X_1 \times \cdots \times X_k \in \mathscr{F}_{x_j}$. If $A \in \mathscr{F}_{x_j}$ and $A_l \in \mathscr{F}_{x_j}$, $l \in \mathbb{Z}_{>0}$, then we have the easily verified identities

$$((X_1 \times \cdots \times X_k) \setminus A)_{x_j} = (X_1 \times \cdots \times X_k) \setminus A_{x_j}$$

and

$$\Big(\cup_{l\in\mathbb{Z}_{>0}} A_l\Big)_{x_j} = \cup_{j\in\mathbb{Z}_{>0}} (A_l)_{x_j},$$

which shows that $\mathscr{F}_{x_j}$ is indeed a σ-algebra. Since it contains the measurable rectangles we must have

$$\sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_k) \subseteq \mathscr{F}_{x_j}.$$

It, therefore, immediately follows that $A_{x_j} \in \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_{j-1} \times \mathscr{A}_{j+1} \times \cdots \times \mathscr{A}_k)$ whenever $A \in \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_k$, as desired. ■

Exercises

2.2.1 Let X be a set and let $\mathscr{A}$ be an algebra on X.

(a) Prove the following:

(i) $\varnothing \in \mathscr{A}$;

(ii) if $A_1, \dots, A_k \in \mathscr{A}$ then $\cap_{j=1}^k A_j \in \mathscr{A}$.

(b) Show that condition (iii) in Definition 2.2.1 can be equivalently replaced with condition (ii) above.

2.2.2 Let X be an infinite set. Indicate which of the following collections of subsets are algebras, σ-algebras, or neither:

(a) the collection of finite subsets X;

(b) the collection of subsets A for which $X \setminus A$ is finite;

(c) the collection of countable subsets X;

(d) the collection of subsets A for which $X \setminus A$ is countable.

2.2.3 Answer the following questions.

(a) Is the collection of open subsets of $\mathbb{R}$ an algebra or a σ-algebra?

(b) Is the collection of closed subsets of $\mathbb{R}$ an algebra or a σ-algebra?

2.2.4 Let X be a set and let $\mathscr{S} \subseteq \mathbf{2}^X$. Show that if

$$\mathscr{S}' = \Big\{ \bigcup_{j \in \mathbb{Z}_{>0}} A_j \ \Big| \ A_j \in \mathscr{S},\ j \in \mathbb{Z}_{>0} \Big\}$$

then the σ-algebras $\sigma(\mathscr{S})$ and $\sigma(\mathscr{S}')$ are generated by $\mathscr{S}$ and $\mathscr{S}'$ agree.

2.2.5 Let X and Y be disjoint sets and let $\mathscr{A}$ and $\mathscr{B}$ be σ-algebras on X and Y, respectively. Let

$$\mathscr{A} \cup \mathscr{B} = \{A \cup B \in \mathbf{2}^{X \cup Y} \mid A \in \mathscr{A},\ B \in \mathscr{B}\}.$$

Show that $\mathscr{A} \cup \mathscr{B}$ is a σ-algebra on $X \cup Y$.

Section 2.3

Measures

The nomenclature "measurable space" from the preceding section makes one think that one ought to be able to measure things in it. This is done with the concept of a measure that we now introduce, and which serves to provide a general framework for talking about the "size" of a subset. The notion of what we shall below call an "outer measure" is perhaps the most intuitive notion of size one can utilise. It has the great advantage of being able to be applied to measure the size of *all* subsets. However, and surprisingly, outer measure has an important defect, namely that it does not have the seemingly natural property of "countable-additivity." The way one gets around this is by restricting outer measure to a collection of subsets where this property of countable-additivity *does* hold. This leads to a natural σ-algebra. At the high level of abstraction in this section, it is not easy to see the justification for the definitions of outer measure and measure. This justification will only become clear in Section 2.4 where there is a fairly intuitive definition of outer measure on $\mathbb{R}$, but that natural outer measure is actually not a measure.

Do I need to read this section? In order to appreciate the framework in which the Lebesgue measure is developed in Sections 2.4 and 2.5, one should understand the notions of measure and outer measure. •

2.3.1 Functions on families of subsets

Before getting to the more specific definitions that we shall mainly use, it is useful to provide some terminology that helps to organise these definitions.

2.3.1 Definition (Properties of functions on subsets) For a set X and a collection $\mathscr{S} \subseteq \mathbf{2}^X$ of subsets of X, a map $\mu \colon \mathscr{S} \to \overline{\mathbb{R}}_{\geq 0}$ is:

(i) ***monotonic*** if $\mu(S) \leq \mu(T)$ for subsets $S, T \in \mathscr{S}$ such that $S \subseteq T$;

(ii) ***finitely-subadditive*** if $\mu\Big(\bigcup_{j=1}^{k} S_j\Big) \leq \sum_{j=1}^{k} \mu(S_j)$ for every finite family $(S_1, \dots, S_k)$ of sets from $\mathscr{S}$ whose union is also in $\mathscr{S}$;

(iii) ***countably-subadditive*** if $\mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} S_j\Big) \leq \sum_{j=1}^{\infty} \mu(S_j)$ for every countable family $(S_j)_{j\in\mathbb{Z}_{>0}}$ of sets from $\mathscr{S}$ whose union is also in $\mathscr{S}$;

(iv) ***monotonically increasing*** if, for every countable family of subsets $(S_j)_{j\in\mathbb{Z}_{>0}}$ from $\mathscr{S}$ for which $S_j \subseteq S_{j+1}$, $j \in \mathbb{Z}_{>0}$, and whose union is in $\mathscr{S}$, $\mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} S_j\Big) =$

$\lim_{j\to\infty} \mu(S_j)$;

(v) ***monotonically decreasing*** if, for every countable family of subsets $(S_j)_{j\in\mathbb{Z}_{>0}}$ from $\mathscr{S}$ for which $S_j \supseteq S_{j+1}$, $j \in \mathbb{Z}_{>0}$, for which $\mu(S_k) < \infty$ for some $k \in \mathbb{Z}_{>0}$, and whose intersection is in $\mathscr{S}$, $\mu\Big(\bigcap_{j\in\mathbb{Z}_{>0}} S_j\Big) = \lim_{j\to\infty} \mu(S_j)$.

If μ is $\overline{\mathbb{R}}$-valued then μ is:

(iii) ***finite*** if $X \in \mathscr{S}$ and if μ takes values in $\mathbb{R}$;

(iv) ***σ-finite*** if there exists subsets $(S_j)_{j\in\mathbb{Z}_{>0}}$ from $\mathscr{S}$ such that $|\mu(S_j)| < \infty$ for $j \in \mathbb{Z}_{>0}$ and such that $X = \cup_{j\in\mathbb{Z}_{>0}} S_j$.

(v) ***finitely-additive*** if $\mu\Big(\bigcup_{j=1}^{k} S_j\Big) = \sum_{j=1}^{k} \mu(S_j)$ for every finite family $(S_1, \dots, S_k)$ of pairwise disjoint sets from $\mathscr{S}$ whose union is also in $\mathscr{S}$;

(vi) ***countably-additive*** if $\mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} S_j\Big) = \sum_{j=1}^{\infty} \mu(S_j)$ for every countable family $(S_j)_{j\in\mathbb{Z}_{>0}}$ of pairwise disjoint sets from $\mathscr{S}$ whose union is also in $\mathscr{S}$.

(vii) ***consistent*** if at most one of ∞ and $-\infty$ is in image(μ). •

Initially, we shall only use the preceding definitions in the case where μ takes values in $\overline{\mathbb{R}}_{\geq 0}$. However, in Sections 2.3.7 and 2.3.8 we shall need to consider the case where μ takes values in $\overline{\mathbb{R}}$.

The following result records some obvious relationships between the preceding concepts.

2.3.2 Proposition (Relationships between properties of functions on subsets) *If* X *is a set, if* $\mathscr{S} \subseteq \mathbf{2}^X$*, and if* $\mu\colon \mathscr{S} \to \overline{\mathbb{R}}_{\geq 0}$*, then the following statements hold:*

(i) if $\mu(\emptyset) = 0$ *and if* μ *is countably-subadditive then it is finitely-subadditive;*

(ii) if $\mu(\emptyset) = 0$ *and if* μ *is finitely-additive then it is finitely-subadditive;*

(iii) if $\mu(\emptyset) = 0$ *and if* μ *is countably-additive then it is countably-subadditive;*

(iv) if μ *is countably-additive then it is monotonically increasing;*

(v) if μ *is countably-additive then it is monotonically decreasing;*

(vi) if μ *is finitely-additive then it is monotonic and, moreover,* $\mu(T \setminus S) = \mu(T) - \mu(S)$ *if* $\mu(S) < \infty$.

If μ *takes values in* $\overline{\mathbb{R}}$ *then the following statement holds:*

(vii) if $\mu(\emptyset) = 0$ *and if* μ *is countably-additive then it is finitely-additive.*

If μ *takes values in* $\overline{\mathbb{R}}$ *and if* $\mathscr{S}$ *has the property that* $S \in \mathscr{S}$ *implies that* $X \setminus S \in \mathscr{S}$*, then the following statement holds:*

(viii) if μ *is finitely additive then it is consistent.*

Proof (i) Let $(S_1, \dots, S_k)$ be a finite family of subsets from $\mathscr{S}$. Define $(T_j)_{j\in\mathbb{Z}_{>0}}$ by

$$T_j = \begin{cases} S_j, & j \in \{1, \dots, k\}, \\ \varnothing, & j > k. \end{cases}$$

Then

$$\mu\Big(\bigcup_{j=1}^{k} S_j\Big) = \mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} T_j\Big) \le \sum_{j=1}^{\infty} \mu(T_j) = \sum_{j=1}^{k} \mu(S_j),$$

since $\mu(\varnothing) = 0$.

The following lemma will be useful in the next two parts of the proof, as well as in various other arguments in this chapter.

1 Lemma *Let* X *be a set, let either* $\mathrm{J} = \{1, \dots, \mathrm{m}\}$ *for some* $\mathrm{m} \in \mathbb{Z}_{>0}$ *or* $\mathrm{J} = \mathbb{Z}_{>0}$*, and let* $(\mathrm{S_j})_{\mathrm{j}\in\mathrm{J}}$ *be a countable family of subsets of* X*. Then there exists a family* $(\mathrm{T_j})_{\mathrm{j}\in\mathrm{J}}$ *of subsets of* X *such that*

(i) $\mathrm{T_{j_1}} \cap \mathrm{T_{j_2}} = \varnothing$ *for* $\mathrm{j_1} \neq \mathrm{j_2}$*;*

(ii) $\mathrm{T_j} \subseteq \mathrm{S_j}$, $\mathrm{j} \in \mathrm{J}$*;*

(iii) $\cup_{\mathrm{j}\in\mathrm{J}}\mathrm{T_j} = \cup_{\mathrm{j}\in\mathrm{J}}\mathrm{S_j}$.

Moreover, if $\mathrm{S_j} \in \mathscr{A}$, $\mathrm{j} \in \mathbb{Z}_{>0}$*, for an algebra* $\mathscr{A}$ *on* X*, then the sets* $(\mathrm{T_j})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *can also be chosen to be in* $\mathscr{A}$*.*

Proof Let $j_0 \in J$ and define

$$T'_{j_0} = \bigcup_{j<j_0} (S_{j_0} \cap S_j), \quad T_{j_0} = S_{j_0} \setminus T'_{j_0}.$$

Thus T'_{j_0} is the set of points in S_{j_0} that are already contained in at least one of the "previous" subsets $\{S_j\}_{j<j_0}$, and T_{j_0} is the set of points in S_{j_0} not in one of the sets $\{S_j\}_{j<j_0}$. Thus we immediately have $T_{j_0} \subseteq S_{j_0}$ for each $j_0 \in J$. Let $j_1, j_2 \in J$ be distinct and suppose, without loss of generality, that $j_1 < j_2$. Then, by construction, T_{j_2} contains no points from S_{j_1} and since $T_{j_1} \subseteq S_{j_1}$ our claim follows. Finally, we show that $\cup_{j\in J} T_j = \cup_{j\in J} S_j$. This is clear since T_{j_0} is defined to contain those points from S_{j_0} not already in $S_1, \dots, S_{j_0-1}$.

The last assertion of the lemma follows since the sets T_j, $j \in \mathbb{Z}_{>0}$, are of the form $(X \setminus A) \cap B$ where $A \in \mathscr{A}$ and where B is a union of sets of the form $B_1 \cap B_2$ for $B_1, B_2 \in \mathscr{A}$. Thus $B \in \mathscr{A}$ by Exercise 2.2.1 and so $(X \setminus A) \cap B \in \mathscr{A}$, also by Exercise 2.2.1. ▼

(ii) By the lemma above let $(T_1, \dots, T_m)$ be pairwise disjoint, such that $T_j \subseteq S_j$ for $j \in \{1, \dots, m\}$, and such that $\cup_{j=1}^m T_j = \cup_{j=1}^m S_j$. Then, by finite-additivity,

$$\mu\Big(\bigcup_{k=1}^{m} S_k\Big) = \sum_{k=1}^{m} \mu(T_k).$$

But, for each $k \in \{1, \dots, m\}$, $S_k = S'_k \cup T_k$ and the union is disjoint. Monotonicity of μ gives $\mu(S_j) \ge \mu(T_j)$ for $j \in \{1, \dots, m\}$ which then gives

$$\mu\Big(\bigcup_{k=1}^{m} S_k\Big) = \sum_{k=1}^{m} \mu(T_k) \le \sum_{k=1}^{m} \mu(S_k),$$

as desired.

(iii) This follows from Lemma 1 just as does part (ii), with only trivial modifications to replace finite-additivity with countable-additivity.

(iv) Let $(S_j)_{j\in\mathbb{Z}_{>0}}$ be a countable family of subsets from $\mathscr{S}$ for which $S_j \subseteq S_{j+1}$, $j \in \mathbb{Z}_{>0}$. For $j \in \mathbb{Z}_{>0}$ define

$$T_j = \begin{cases} S_1, & j = 1, \\ S_j \setminus S_{j-1}, & j > 1. \end{cases}$$

Note that the sets $\{T_j\}_{j\in\mathbb{Z}_{>0}}$ are pairwise disjoint by construction and that

$$\bigcup_{j\in\mathbb{Z}_{>0}} S_j = \bigcup_{j\in\mathbb{Z}_{>0}} T_j.$$

Therefore, by countable-additivity,

$$\mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} S_j\Big) = \sum_{j=1}^{\infty} \mu(T_j).$$

But, since $\cup_{j=1}^{k} T_j = S_k$,

$$\sum_{j=1}^{\infty} \mu(T_j) = \lim_{k\to\infty} \sum_{j=1}^{k} \mu(T_j) = \lim_{k\to\infty} \mu\Big(\bigcup_{j=1}^{k} T_j\Big) = \lim_{k\to\infty} \mu(S_k),$$

which gives

$$\mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} S_j\Big) = \lim_{k\to\infty} \mu(S_k),$$

as desired.

(v) Let $(S_j)_{j\in\mathbb{Z}_{>0}}$ be a countable family of sets from $\mathscr{S}$ such that $S_j \supseteq S_{j+1}$, $j \in \mathbb{Z}_{>0}$, and such that $\mu(S_k)$ for some $k \in \mathbb{Z}_{>0}$. Define $(T_j)_{j\in\mathbb{Z}_{>0}}$ by $T_j = S_{j+k}$ so that

$$\bigcap_{j\in\mathbb{Z}_{>0}} S_j = \bigcap_{j\in\mathbb{Z}_{>0}} T_j.$$

Now define $(U_j)_{j\in\mathbb{Z}_{>0}}$ by $U_j = T_1 \setminus T_j$ so that $U_j \subseteq U_{j+1}$ for each $j \in \mathbb{Z}_{>0}$. We also have

$$\bigcup_{j\in\mathbb{Z}_{>0}} U_j = T_1 \setminus \Big(\bigcap_{j\in\mathbb{Z}_{>0}} T_j\Big).$$

By parts (vi) (since $\mu(T_1) < \infty$) and (iv) we then have

$$\begin{aligned} \mu(T_1) - \mu\Big(\bigcap_{j\in\mathbb{Z}_{>0}} S_j\Big) &= \mu\Big(T_1 \setminus \Big(\bigcap_{j\in\mathbb{Z}_{>0}} S_j\Big)\Big) = \mu\Big(T_1 \setminus \Big(\bigcap_{j\in\mathbb{Z}_{>0}} T_j\Big)\Big) \\ &= \mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} U_j\Big) = \mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} U_j\Big) = \lim_{j\to\infty} \mu(U_j) \\ &= \lim_{j\to\infty} \mu(T_1 \setminus T_j) = \mu(T_1) - \lim_{j\to\infty} \mu(T_j) \\ &= \mu(T_1) - \lim_{j\to\infty} \mu(S_j), \end{aligned}$$

from which we deduce
$$\mu\Big(\bigcap_{j\in\mathbb{Z}_{>0}} S_j\Big) = \lim_{j\to\infty} \mu(S_j)$$
since $\mu(T_1) < \infty$.

(vi) Let $S, T \in \mathscr{S}$ be such that $S \subseteq T$. Then, by finite-additivity,
$$\mu(S) \leq \mu(S) + \mu(T - S) = \mu(T),$$
as desired. The formula $\mu(T \setminus S) = \mu(T) - \mu(S)$ if $\mu(S) = \infty$ follows immediately from finite-additivity.

(vii) Let $(S_1, \dots, S_k)$ be a finite family of subsets from $\mathscr{S}$. Define $(T_j)_{j\in\mathbb{Z}_{>0}}$ by
$$T_j = \begin{cases} S_j, & j \in \{1, \dots, k\}, \\ \varnothing, & j > k, \end{cases}$$
noting that the family $(T_j)_{j\in\mathbb{Z}_{>0}}$ is pairwise disjoint. Then
$$\mu\Big(\bigcup_{j=1}^{k} S_j\Big) = \mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} T_j\Big) = \sum_{j=1}^{\infty} \mu(T_j) = \sum_{j=1}^{k} \mu(S_j),$$
since $\mu(\varnothing) = 0$.

(viii) Suppose that there exists sets $S_+, S_- \in \mathscr{S}$ such that $\mu(S_+) = \infty$ and $\mu(S_-) = -\infty$. Then, finite-additivity and the assumption that sets from $\mathscr{S}$ have complements in $\mathscr{S}$ implies that
$$\mu(X) = \mu(S_+) + \mu(X \setminus S_+) = \mu(A_-) + \mu(X \setminus S_-).$$
Since $\mu(S_+) = \infty$ and since $\mu(S_-) = -\infty$ and since the addition $\infty + (-\infty)$ is not defined, we must have
$$\mu(X \setminus S_+) \in \overline{\mathbb{R}} \setminus \{-\infty\}, \quad \mu(X \setminus S_-) \in \overline{\mathbb{R}} \setminus \{\infty\}.$$
Therefore,
$$\mu(X) = \infty, \quad \mu(X) = -\infty,$$
giving a contradiction. ■

The following relationships between finite-additivity, countable-additivity, and monotonicity are also useful.

2.3.3 Proposition (Additivity and monotonicity) *Let X be a set with $\mathscr{A} \subseteq \mathbf{2}^X$ an algebra on $\mathscr{A}$, and let $\mu_0\colon \mathscr{A} \to \overline{\mathbb{R}}$ be consistent, finitely-additive, and have the property that $\mu_0(\varnothing) = 0$. The following three statements are equivalent:*

(i) μ_0 is countably-additive;

(ii) for every sequence $(A_j)_{j\in\mathbb{Z}_{>0}}$ of subsets from $\mathscr{A}$ for which $A_j \subseteq A_{j+1}$, $j \in \mathbb{Z}_{>0}$, and for which $\cup_{j\in\mathbb{Z}_{>0}} A_j \in \mathscr{A}$, it holds that
$$\mu_0\Big(\bigcup_{j\in\mathbb{Z}_{>0}} A_j\Big) = \lim_{j\to\infty} \mu_0(A_j);$$

(iii) for every sequence $(A_j)_{j\in\mathbb{Z}_{>0}}$ *of subsets from* $\mathscr{A}$ *for which* $A_j \supseteq A_{j+1}$, $j \in \mathbb{Z}_{>0}$, *for which* $\cap_{j\in\mathbb{Z}_{>0}} A_j = \emptyset$, *it holds that* $\lim_{j\to\infty} \mu_0(A_j) = 0$.

Proof (i) $\Longrightarrow$ (ii) Let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of subsets from $\mathscr{A}$ for which $A_j \subseteq A_{j+1}$, $j \in \mathbb{Z}_{>0}$, and for which $\cup_{j\in\mathbb{Z}_{>0}} A_j \in \mathscr{A}$. Let us denote $A = \cup_{j\in\mathbb{Z}_{>0}} A_j$. Define $B_1 = A_1$ and for $j \geq 2$ define $B_j = A_j \setminus A_{j-1}$. Then the family $(B_j)_{j\in\mathbb{Z}_{>0}}$ is pairwise disjoint and satisfies $\cup_{j\in\mathbb{Z}_{>0}} B_j = A$. The assumed consistency and countable-additivity of μ_0 then gives

$$\mu_0(A) = \sum_{j=1}^{\infty} \mu_0(B_j).$$

Moreover, since $A_k = \cup_{j=1}^{k} B_j$ we have

$$\mu_0(A_k) = \sum_{j=1}^{k} \mu_0(B_j) \quad \Longrightarrow \quad \lim_{k\to\infty} \mu_0(A_k) = \sum_{j=1}^{\infty} \mu_0(B_j) = \mu_0(A),$$

as desired.

(ii) $\Longrightarrow$ (iii) Let us define $B_j = A_{k+j-1}$ for $j \in \mathbb{Z}_{>0}$. Then $\mu_0(B_1) < \infty$ and $\cap_{j\in\mathbb{Z}_{>0}} B_j = \emptyset$. Also define $C_j = B_1 \setminus B_{j+1}$ for $j \in \mathbb{Z}_{>0}$. Then the family of subsets $(C_j)_{j\in\mathbb{Z}_{>0}}$ is in $\mathscr{A}$ and satisfies $C_j \subseteq C_{j+1}$ for each $j \in \mathbb{Z}_{>0}$. Moreover,

$$\bigcup_{j\in\mathbb{Z}_{>0}} C_j = \bigcup_{j\in\mathbb{Z}_{>0}} B_1 \setminus B_{j+1} = B_1 \setminus \bigcap_{j\in\mathbb{Z}_{>0}} B_{j+1} = B_1,$$

using De Morgan's Laws. By assumption we then have

$$\lim_{j\to\infty} \mu_0(C_j) = \mu_0(B_1).$$

Therefore,

$$\lim_{j\to\infty} \mu_0(C_j) = \lim_{j\to\infty} \mu_0(B_1 \setminus B_{j+1}) = \lim_{j\to\infty} (\mu_0(B_1) - \mu_0(B_{j+1})) = \mu_0(B_1),$$

allowing us to conclude that

$$\lim_{j\to\infty} \mu_0(A_j) = \lim_{j\to\infty} \mu_0(B_j) = 0,$$

as desired.

(iii) $\Longrightarrow$ (i) Let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a family of pairwise disjoint sets and denote $A = \cup_{j\in\mathbb{Z}_{>0}} A_j$, supposing that $A \in \mathscr{A}$. For $k \in \mathbb{Z}_{>0}$ define $B_k = A \setminus \cup_{j=1}^{k} A_j$. Then $B_k \supseteq B_{k+1}$ and $\cap_{k\in\mathbb{Z}_{>0}} B_k = \emptyset$. By assumption we then have $\lim_{k\to\infty} \mu_0(B_k) = 0$. We have $A = B_k \cup (\cup_{j=1}^{k} A_j)$ with the union being disjoint. Finite-additivity of μ_0 gives

$$\mu_0(A) = \mu_0(B_k) + \sum_{j=1}^{k} \mu_0(A_j),$$

which gives

$$\mu_0(A) = \lim_{k\to\infty} \mu_0(B_k) + \sum_{j=1}^{\infty} \mu_0(A_j) = \sum_{j=1}^{\infty} \mu_0(A_j),$$

as desired. ■

2.3.2 Outer measures, measures, and their relationship

With the general properties of functions on subsets from the preceding section, we now introduce our first notion of "size" of a subset.

2.3.4 Definition (Outer measure) Let X be a set. An ***outer measure*** on X is a map $\mu^*\colon \mathbf{2}^X \to \overline{\mathbb{R}}_{\geq 0}$ with the following properties:

(i) $\mu^*(\varnothing) = 0$;

(ii) μ^* is monotonic;

(iii) μ^* is countably-subadditive. •

2.3.5 Remark (Why are the axioms for an outer measure as they are?) The notion of outer measure is intuitive, in the sense that its properties are included in those that we anticipate a reasonable notion of "size" to possess. What is not immediately clear is that these are the *only* properties that one might demand of our notion of size. This latter matter is difficult to address *a priori*, and indeed is only really addressed by knowing that these are indeed the properties that one uses in the development of the general theory. •

Let us consider some simple examples of outer measures. We shall postpone to Sections 2.4 and 2.5 the presentation of more interesting examples.

2.3.6 Examples (Outer measures)

1. For a set X, the map $\mu^*\colon \mathbf{2}^X \to \overline{\mathbb{R}}_{\geq 0}$ defined by $\mu^*(A) = 0$ is an outer measure. We call this the ***zero outer measure***.
2. Let us consider a set X with $\mu^*\colon \mathbf{2}^X \to \overline{\mathbb{R}}_{\geq 0}$ defined by
$$\mu^*(A) = \begin{cases} 0, & A = \varnothing, \\ \infty, & A \neq \varnothing. \end{cases}$$
It is then easy to see that μ^* is an outer measure.
3. For a set X define $\mu^*\colon \mathbf{2}^X \to \overline{\mathbb{R}}_{\geq 0}$ by
$$\mu^*(A) = \begin{cases} \operatorname{card}(A), & \operatorname{card}(A) < \infty, \\ \infty, & \text{otherwise.} \end{cases}$$
It is easy to verify that μ^* is an outer measure.
4. For a set X define $\delta_x^*\colon \mathbf{2}^X \to \overline{\mathbb{R}}_{\geq 0}$ by
$$\delta_x^*(A) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$
One can easily see that δ_x^* is indeed an outer measure. •

The notion of outer measure is a nice one in that it allows the measurement of size for any subset of the set X. However, it turns out that some outer measures lack an important property. Namely, there are outer measures μ^* (namely, the Lebesgue outer measure of Definition 2.4.1) that lack the property that, if $S, T \subseteq X$ are *disjoint*, then $\mu^*(S \cup T) = \mu^*(S) + \mu^*(T)$. Upon reflection, we hope the reader can see that this is indeed a property one would like any notion of size to possess. In order to ensure that this property is satisfied, it turns out that one needs to restrict oneself to measuring a subset of the collection of all sets. It is here where the notions of algebras and σ-algebras come into play.

2.3.7 Definition (Measure, measure space) Let X be a set and let $\mathscr{S} \subseteq \mathbf{2}^X$. A ***finitely-additive measure*** on $\mathscr{S}$ is a map $\mu\colon \mathscr{S} \to \overline{\mathbb{R}}_{\geq 0}$ with the following properties:

(i) $\mu(\emptyset) = 0$;

(ii) μ is finitely-additive.

A ***countably-additive measure***, or simply a ***measure***, on $\mathscr{S}$ is a map $\mu\colon \mathscr{S} \to \overline{\mathbb{R}}_{\geq 0}$ with the following properties:

(iii) $\mu(\emptyset) = 0$;

(iv) μ is countably-additive.

A triple $(X, \mathscr{A}, \mu)$ is called a ***measure space*** is $\mathscr{A}$ is a σ-algebra on X and if μ is a countably-additive measure on $\mathscr{A}$. •

Just as we are primarily interested in σ-algebras in preference to algebras, we are also primarily interested in countably-additive measures in preference to finitely-additive measures. However, finitely-additive measures will come up, usually in the course of a construction of a countably-additive measure.

2.3.8 Remark (Why are the axioms for a measure as they are?) Again, it is not perfectly evident why a measure has the stated properties. In particular, the conditions that (1) a measure space involves a σ-algebra and that (2) a measure be countably-additive seem like they ought to admit many viable alternatives. Why not allow a measure space to be defined using *any* collection of subsets? Why not finite-additivity? finite-subadditivity? countable-subadditivity? The reasons to restrict to a σ-algebra (possibly) smaller than the collection of all subsets will be made clear shortly. As concerns *countable*-additivity, the reasons for this are much like they are for the countability conditions for σ-algebras; countability is what we want here since we are after nice behaviour of our constructions with sequential operations. The requirement of disjointness in the definition is not so easy to understand. Indeed, in our definition of an outer measure in Definition 2.3.4 we relaxed this, and possibly the definition of an outer measure seems like the one that we should really be interested in. However, it is not, although the reasons for this will only be made clear as we go along. •

Let us give some simple examples of measures.

2.3.9 Examples (Measures)

1. For a measurable space $(X, \mathscr{A})$, the map $\mu\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ defined by $\mu^*(A) = 0$ is a measure. We call this the ***outer measure***.
2. For a measurable space $(X, \mathscr{A})$ define $\mu\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ by

$$\mu(A) = \begin{cases} 0, & A = \emptyset, \\ \infty, & A \neq \emptyset. \end{cases}$$

   This defines a measure on $(X, \mathscr{A})$.
3. If $(X, \mathscr{A})$ is a measurable space then define $\mu_\Sigma\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ by

$$\mu_\Sigma(A) = \begin{cases} \operatorname{card}(A), & \operatorname{card}(A) < \infty, \\ \infty, & \text{otherwise}. \end{cases}$$

   One may verify that this defines a measure for the measurable space $(X, \mathscr{A})$ called the ***counting measure***.
4. If $(X, \mathscr{A})$ is a measurable space and if $x \in X$ we define $\delta_x\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ by

$$\delta_x(A) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A. \end{cases}$$

   One may verify that this defines a measure and is called the ***point mass*** concentrated at x.
5. On the algebra $\mathscr{J}(\mathbb{R}^n)$ of Jordan measurable subsets of $\mathbb{R}^n$ the map $\rho\colon \mathscr{J}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ of Definition 2.1.6 is a finitely-additive measure. This follows from Proposition 2.1.8. •

Let us give some properties of measures that follow more or less directly from the definitions.

2.3.10 Proposition (Properties of measures) *For a set* X*, a collection of subsets* $\mathscr{S} \subseteq \mathbf{2}^{\mathsf{X}}$*, and a measure* μ *on* $\mathscr{S}$*, the following statements hold:*

(i) *μ is finitely-additive;*

(ii) *μ is monotonic and $\mu(\mathsf{T} \setminus \mathsf{S}) = \mu(\mathsf{T}) - \mu(\mathsf{S})$ if $\mu(\mathsf{S}) < \infty$;*

(iii) *μ is countably-subadditive;*

(iv) *μ is monotonically increasing;*

(v) *μ is monotonically decreasing.*

Proof (i) This follows immediately from Proposition 2.3.2(vii).

(ii) This follows from Proposition 2.3.2(vi) and part (i).

(iii) This follows from Proposition 2.3.2(iii).

(iv) This follows from Proposition 2.3.2(iv).

(v) This follows from Proposition 2.3.2(v). ■

Now let us examine the relationships between outer measure and measure. Let us begin with something elementary, given what we already know.

2.3.11 Proposition (When are measures outer measures?) *If* $(X, \mathscr{A}, \mu)$ *is a measure space then* μ *is an outer measure if and only if* $\mathscr{A} = \mathbf{2}^X$.

Proof This follows immediately from Proposition 2.3.10. ■

Since the outer measures in the examples are all actually measures, this leads one to the following line of questioning.

1. Are all outer measures measures?
2. Given a measure space $(X, \mathscr{A}, \mu)$ does there exist an outer measure μ^* on X for which $\mu = \mu^*|\mathscr{A}$?

We shall see in Corollary 2.3.29 that the answer to the second question is, "Yes." The answer to the first question is, "No," but we will have to wait until Section 2.4 (in particular, Example 2.4.3) to see an example of an outer measure that is not a measure. The key issue concerning whether an outer measure is a measure hinges on the following characterisation of a distinguished class of subsets of a set with an outer measure.

2.3.12 Definition (Measurable subsets for an outer measure) If μ^* is an outer measure on a set X, a subset $A \subseteq X$ is ***μ^*-measurable*** if

$$\mu^*(S) = \mu^*(S \cap A) + \mu^*(S \cap (X \setminus A))$$

for all $S \subseteq X$. The set of μ^*-measurable subsets is denoted by $\mathscr{M}(X, \mu^*)$. •

Note that an outer measure is finitely-subadditive by Proposition 2.3.2(i). Thus we always have

$$\mu^*(S) \leq \mu^*(S \cap A) + \mu^*(S \cap (X \setminus A)).$$

Therefore, a set A is *not* μ^*-measurable then we have

$$\mu^*(S) > \mu^*(S \cap A) + \mu^*(S \cap (X \setminus A)).$$

The definition of μ^*-measurability looks like it provides a "reasonable" property of a subset A: that the outer measure of a set S should be the outer measure of the points in S that are in A plus the outer measure of the points in S that are not in A. In Figure 2.2 we attempt to depict what is going on. What is not so obvious is that not all subsets need be μ^*-measurable. In the examples of outer measures in Example 2.3.6 above, they all turn out to be measures. It is only when we get to the more sophisticated construction of the Lebesgue measure in Section 2.4 that we see that nonmeasurable sets exist. Indeed, it is precisely in the constructions of Section 2.4 that the general ideas we are presently discussing were developed.

For the purposes of our present development, the following theorem is important in that it gives a natural passage from an outer measure to a measure space.

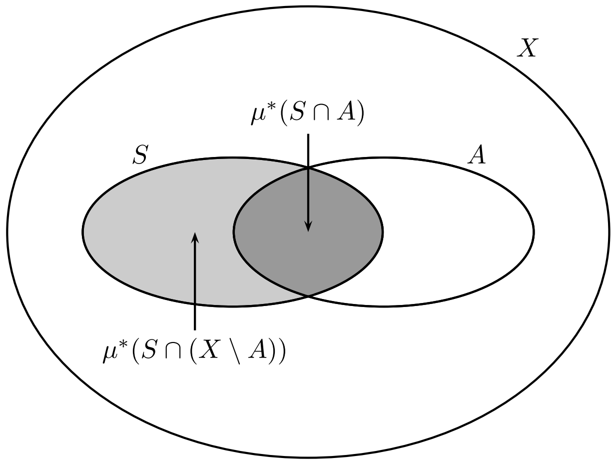


Figure 2.2 The notion of a μ^*-measurable set

2.3.13 Theorem (Outer measures give measure spaces) *If μ^* is an outer measure on a set* X *then* $(\mathsf{X}, \mathscr{M}(\mathsf{X}, \mu^*), \mu^*|\mathscr{M}(\mathsf{X}, \mu^*))$ *is a measure space.*

Proof Let us first show that $X \in \mathscr{M}(X, \mu^*)$. Let $S \in \mathbf{2}^X$ and note that

$$\mu^*(S \cap X) + \mu^*(S \cap (X \setminus X)) = \mu^*(S)$$

since $\mu^*(\emptyset) = 0$.

Now let us show that if $A \in \mathscr{M}(X, \mu^*)$ then $X \setminus A \in \mathscr{M}(X, \mu^*)$. This follows since

$$\mu^*(S \cap (X \setminus A)) + \mu^*(S \cap (X \setminus (X \setminus A))) = \mu^*(S \cap A) + \mu^*(S \cap (X \setminus A)) = \mu^*(S).$$

Next we show that if $A_1, \dots, A_n \in \mathscr{M}(X, \mu^*)$ then $\cup_{j=1}^n A_j \in \mathscr{M}(X, \mu^*)$. This will follow by a trivial induction if we can prove it for $n = 2$. Thus we let $A_1, A_2 \in \mathscr{M}(X, \mu^*)$, $S \subseteq X$, and compute

$$\begin{aligned}
&\mu^*(S \cap (A_1 \cup A_2)) + \mu^*(S \cap (X \setminus (A_1 \cup A_2))) \\
&= \mu^*((S \cap (A_1 \cup A_2)) \cap A_1) + \mu^*((S \cap (A_1 \cup A_2)) \cap (X \setminus A_1)) + \mu^*(S \cap (X \setminus (A_1 \cup A_2))) \\
&= \mu^*(S \cap A_1) + \mu^*(S \cap (X \setminus A_1) \cap A_2) + \mu^*(S \cap (X \setminus A_1) \cap (X \setminus A_2)) \\
&= \mu^*(S \cap A_1) + \mu^*(S \cap (X \setminus A_1)) = \mu^*(S).
\end{aligned}$$

In going from the first line to the second line we have used the fact that $A_1 \in \mathscr{M}(X, \mu^*)$. In going from the second line to the third line we have used some set theoretic identities for union and intersection that can be easily verified, e.g., by using Propositions I-1.1.4 and I-1.1.5. In going from the third line to the fourth line we have used the fact that $A_2 \in \mathscr{M}(X, \mu^*)$.

Next we show that property (vi) of Definition 2.2.1 holds. Thus we let $(A_j)_{j \in \mathbb{Z}_{>0}} \subseteq \mathscr{M}(X, \mu^*)$. To show that $\cup_{j \in \mathbb{Z}_{>0}} A_j \in \mathscr{M}(X, \mu^*)$ we may without loss of generality suppose that the sets $(A_j)_{j \in \mathbb{Z}_{>0}}$ are disjoint. Indeed, if they are not then we may replace their union with the union of the sets

$$\begin{aligned}\tilde{A}_1 &= A_1\\ \tilde{A}_2 &= A_2 \cap (X \setminus A_1)\\ &\vdots\\ \tilde{A}_j &= A_j \cap (X \setminus A_1) \cap \cdots \cap (X \setminus A_{j-1})\\ &\vdots\end{aligned}$$

where the collection $(\tilde{A}_j)_{j\in\mathbb{Z}_{>0}}$ is disjoint. First we claim that under this assumption that $(A_j)_{j\in\mathbb{Z}_{>0}} \subseteq \mathscr{M}(X, \mu^*)$ is disjoint we have

$$\mu^*(S) = \sum_{j=1}^{n} \mu^*(S \cap A_j) + \mu^*\Big(S \cap \Big(\bigcap_{j=1}^{n} (X \setminus A_j)\Big)\Big). \tag{2.1}$$

We prove this by induction. For $n = 1$ the claim follows since $A_1 \in \mathscr{M}(X, \mu^*)$. Now suppose the claim true for $n = k$ and compute

$$\begin{aligned}&\mu^*\Big(S \cap \Big(\bigcap_{j=1}^{k} (X \setminus A_j)\Big)\Big)\\ &= \mu^*\Big(S \cap \Big(\bigcap_{j=1}^{k} (X \setminus A_j)\Big) \cap A_{k+1}\Big) + \mu^*\Big(S \cap \Big(\bigcap_{j=1}^{k} (X \setminus A_j)\Big) \cap (X \setminus A_{k+1})\Big)\\ &= \mu^*(S \cap A_{k+1}) + \mu^*\Big(S \cap \Big(\bigcap_{j=1}^{k+1} (X \setminus A_j)\Big)\Big),\end{aligned}$$

so establishing (2.1) after an application of the induction hypothesis. In the first line we use the fact that $A_{k+1} \in \mathscr{M}(X, \mu^*)$ and in the second line we have used the fact that the set $(A_j)_{j\in\mathbb{Z}_{>0}}$ are disjoint.

By monotonicity of outer measures we have

$$\begin{aligned}&& \mu^*(S) &\geq \sum_{j=1}^{n} \mu^*(S \cap A_j) + \mu^*\Big(S \cap \Big(\bigcap_{j=1}^{\infty} (X \setminus A_j)\Big)\Big)\\ &\implies & \mu^*(S) &\geq \sum_{j=1}^{n} \mu^*(S \cap A_j) + \mu^*\Big(S \cap \Big(X \setminus \bigcup_{j=1}^{\infty} A_j\Big)\Big)\\ &\implies & \mu^*(S) &\geq \sum_{j=1}^{\infty} \mu^*(S \cap A_j) + \mu^*\Big(S \cap \Big(X \setminus \bigcup_{j=1}^{\infty} A_j\Big)\Big) \qquad (2.2)\\ &\implies & \mu^*(S) &\geq \mu^*\Big(S \cap \Big(\bigcup_{j=1}^{\infty} A_j\Big)\Big) + \mu^*\Big(S \cap \Big(X \setminus \bigcup_{j=1}^{\infty} A_j\Big)\Big).\end{aligned}$$

In the first line we have used (2.1) along with monotonicity of outer measures. In the second line we have used a simple set theoretic identity. In the third line we have simply taken the limit of a bounded monotonically increasing sequence of numbers. In the fourth line we have used countable-subadditivity of outer measures. This then gives

$$\mu^*(S) \geq \mu^*\Big(S \cap \Big(\bigcup_{j=1}^{\infty} A_j\Big)\Big) + \mu^*\Big(S \cap \Big(X \setminus \bigcup_{j=1}^{\infty} A_j\Big)\Big) \geq \mu^*(S),$$

by another application countable-subadditivity of outer measures. It therefore follows that $\cup_{j\in\mathbb{Z}_{>0}} A_j \in \mathscr{M}(X, \mu^*)$, as was to be shown.

The next thing we show is that $\mu \triangleq \mu^*|\mathscr{M}(X, \mu^*)$ is a measure on $(X, \mathscr{M}(X, \mu^*))$. Since

$$\mu^*(S) = \mu^*(S \cap \varnothing) + \mu^*(S \cap X) = \mu^*(\varnothing) + \mu^*(S),$$

for every $S \in \mathbf{2}^X$ it follows that $\mu(\varnothing) = \mu^*(\varnothing) = 0$. Now let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a collection of disjoint sets in $\mathscr{M}(X, \mu^*)$. We have

$$\mu\Big(\bigcup_{j=1}^{\infty} A_j\Big) = \mu^*\Big(\bigcup_{j=1}^{\infty} A_j\Big) \geq \sum_{j=1}^{\infty} \mu^*(A_j) + 0,$$

using (2.2) with $S = \cup_{j=1}^{\infty} A_j$. By monotonicity of outer measures we also have

$$\mu\Big(\bigcup_{j=1}^{\infty} A_j\Big) \leq \sum_{j=1}^{\infty} \mu^*(A_j),$$

and so μ is countably-additive. ■

The theorem immediately has the following corollary which helps to clarify the relationship between measures and outer measures.

2.3.14 Corollary (An outer measure is a measure if and only if all subsets are measurable) *If μ^* is an outer measure on X then $(X, \mathbf{2}^X, \mu^*)$ is a measure space if and only if every subset of X is μ^*-measurable.*

Proof From Theorem 2.3.13 it follows that $(X, \mathbf{2}^X, \mu^*)$ is a measure space if $\mathscr{M}(X, \mu^*) = \mathbf{2}^X$. For the converse, suppose that $A \subseteq X$ is not μ^*-measurable. Then there exists a set $S \subseteq X$ such that

$$\mu^*(S) \neq \mu^*(S \cap A) + \mu^*(S \cap (X \setminus A)).$$

However, since $S = (S \cap A) \cup (S \cap (X \setminus A))$ this prohibits μ^* from being a measure since, if it were a measure, we would have

$$\mu^*(S) = \mu^*(S \cap A) + \mu^*(S \cap (X \setminus A)).$$ ■

Thus the existence of nonmeasurable sets is exactly the obstruction to an outer measure being a measure. Said otherwise, if we wish for an outer measure to behave like a measure— i.e., have the property that

$$\mu^*\Big(\bigcup_{j\in\mathbb{Z}_{>0}} A_j\Big) = \sum_{j=1}^{\infty} \mu^*(A_j)$$

for a family $(A_j)_{j\in\mathbb{Z}_{>0}}$ of disjoint sets—then the sacrifice we have to make is that we possibly restrict the sets which we apply the outer measure to.

The following notions are also sometimes useful.

2.3.15 Definition (Continuous measure, discrete measure) Let $(X, \mathscr{A}, \mu)$ be a measure space for which $\{x\} \in \mathscr{A}$ for every $x \in X$. The measure μ is

(i) ***continuous*** if $\mu(\{x\}) = 0$ for every $x \in X$ and

(ii) ***discrete*** if there exists a countable subset $D \in \mathscr{A}$ such that $\mu(X \setminus D) = 0$. •

Let us consider how these various properties show up in our simple examples of measure spaces.

2.3.16 Examples (Properties of measures)

1. We consider the measure space $(X, \mathscr{A}, \mu)$ where $\mu(\emptyset) = 0$ and $\mu(A) = \infty$ for all nonempty measurable sets. This measure space is σ-finite if and only if $X = \emptyset$, is continuous if and only if $X = \emptyset$, and is discrete if and only if X is countable.
2. Let us consider a measurable space $(X, \mathscr{A})$ and for simplicity assume that $\{x\} \in \mathscr{A}$ for every $x \in X$. The counting measure is σ-finite if and only if X is countable, is not continuous, and is discrete if and only if X is countable.
3. For a measurable space $(X, \mathscr{A})$ the point mass measure δ_x is σ-finite if and only if X is a countable union of measurable sets, is not continuous, and is discrete if and only if there exists a countable set $D \in \mathscr{A}$ such that $x \notin D$. •

Let us close this section by introducing an important piece of lingo.

2.3.17 Notation (Almost everywhere, a.e.) Let $(X, \mathscr{A}, \mu)$ be a measure space. A property P of the set X holds ***µ-almost everywhere*** (***µ-a.e.***) if there exists a set $A \subseteq X$ for which $\mu(A) = 0$, and such that P holds for all $x \in X \setminus A$. If μ is understood, then we may simply write ***almost everywhere*** (***a.e.***). Some authors use "p.p." after the French "presque partout." Lebesgue, after all, was French. •

Let us finally show that the restriction of a measure to a subset makes sense if the subset is measurable.

2.3.18 Proposition (Restriction of measure to measurable subsets) *Let* $(\mathrm{X}, \mathscr{A}, \mu)$ *be a measure space, let* $\mathrm{A} \in \mathscr{A}$, *let* $(\mathrm{A}, \mathscr{A}_{\mathrm{A}})$ *be the measurable space of Proposition 2.2.6, and define* $\mu_{\mathrm{A}} \colon \mathscr{A}_{\mathrm{A}} \to \overline{\mathbb{R}}_{\geq 0}$ *by* $\mu_{\mathrm{A}}(\mathrm{A} \cap \mathrm{B}) = \mu(\mathrm{A} \cap \mathrm{B})$. *Then* $(\mathrm{A}, \mathscr{A}_{\mathrm{A}}, \mu_{\mathrm{A}})$ *is a measure space.*

Proof It is clear that $\mu_A(\emptyset) = 0$. Also let $(B_j \cap A)_{j \in \mathbb{Z}_{>0}}$ be a countable family of disjoint sets in $\mathscr{A}_A$. Since $B_j \cap A \in \mathscr{A}$ for $j \in \mathbb{Z}_{>0}$ this immediately implies that

$$\mu_A\Big(\bigcup_{j \in \mathbb{Z}_{>0}} B_j \cap A\Big) = \sum_{j=1}^{\infty} \mu_A(B_j \cap A),$$

thus showing that μ_A is a measure on $(A, \mathscr{A}_A)$. ■

2.3.3 Complete measures and completions of measures

In this section we consider a rather technical property of measure spaces, but one that does arise on occasion. It is a property that is at the same time (occasionally) essential and (occasionally) bothersome. This is especially true of the Lebesgue measure we consider in Sections 2.4 and 2.5. We shall point out instances of both of these attributes as we go along.

First we give the definition.

2.3.19 Definition (Complete measure) A measure space $(X, \mathscr{A}, \mu)$ is ***complete*** if for every pair of sets A and B with the properties that $A \subseteq B$, $B \in \mathscr{A}$, and $\mu(B) = 0$, we have $A \in \mathscr{A}$. •

Note that completeness has the interpretation that every subset of a set of measure zero should itself be in the set of measurable subsets, and have measure zero. This seems like a reasonable restriction, but it is one that is not met in certain common examples (see). In cases where we have a measure space that is not complete one can simply add some sets to the collection of measurable sets that make the resulting measure space complete. This is done as follows.

what?

2.3.20 Definition (Completion of a measure space) For a measure space $(X, \mathscr{A}, \overline{\mu})$ the ***completion $\mathscr{A}$ under μ*** is the collection $\mathscr{A}_\mu$ of subsets $A \subseteq X$ for which there exists $L, U \in \mathscr{A}$ such that $L \subseteq A \subseteq U$ and $\mu(U \setminus L) = 0$. Define $\overline{\mu}\colon \mathscr{A}_\mu \to \overline{\mathbb{R}}_{\geq 0}$ by $\overline{\mu}(A) = \mu(U) = \mu(L)$ where U and L are any sets satisfying $L \subseteq A \subseteq U$ and $\mu(U \setminus L) = 0$. The triple $(X, \mathscr{A}_\mu, \overline{\mu})$ is the ***completion*** of $(X, \mathscr{A}, \mu)$. •

The completion of a measure space is a complete measure space, as we now show.

2.3.21 Proposition (The completion of a measure space is complete) *If* $(\mathrm{X}, \mathscr{A}_\mu, \overline{\mu})$ *is the completion of* $(\mathrm{X}, \mathscr{A}, \mu)$ *then* $(\mathrm{X}, \mathscr{A}_\mu, \overline{\mu})$ *is a complete measure space for which* $\mathscr{A} \subseteq \mathscr{A}_\mu$.

Proof If $A \in \mathscr{A}$ then $A \subseteq A \subseteq A$ so that $A \in \mathscr{A}_\mu$. In particular, $X \in \mathscr{A}_\mu$. Note that $L \subseteq A \subseteq U$ and $\mu(U \setminus L) = 0$ implies that $(X \setminus U) \subseteq (X \setminus A) \subseteq (X \setminus L)$ and that $\mu((X \setminus L) \setminus (X \setminus U)) = 0$, thus showing that $X \setminus A \in \mathscr{A}_\mu$. Now let $(A_j)_{j\in\mathbb{Z}_{>0}} \subseteq \mathscr{A}_\mu$ and let $(L_j)_{j\in\mathbb{Z}_{>0}}$ and $(U_j)_{j\in\mathbb{Z}_{>0}}$ satisfy

$$L_j \subseteq A_j \subseteq U_j, \qquad \mu(U_j \setminus L_j) = 0, \qquad j \in \mathbb{Z}_{>0}. \tag{2.3}$$

A direct computation shows that

$$\bigcup_{j\in\mathbb{Z}_{>0}} L_j \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} A_j \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} U_j, \qquad \mu\Big(\Big(\bigcup_{j\in\mathbb{Z}_{>0}} U_j\Big) \setminus \Big(\bigcup_{j\in\mathbb{Z}_{>0}} L_j\Big)\Big) \leq \sum_{j=1}^\infty \mu(U_j \setminus L_j) = 0.$$

This shows that $\mathscr{A}_\mu$ is a σ-algebra.

Note that $\overline{\mu}(\varnothing) = 0$. Also let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a collection of disjoint subsets in $\mathscr{A}_\mu$ and take $(L_j)_{j\in\mathbb{Z}_{>0}}$ and $(U_j)_{j\in\mathbb{Z}_{>0}}$ to satisfy (2.3). Note that the sets $(L_j)_{j\in\mathbb{Z}_{>0}}$ are disjoint. From the definition of $\overline{\mu}$ it then follows that $\overline{\mu}$ is countably-additive. It remains to show that $(X, \mathscr{A}_\mu, \overline{\mu})$ is complete. If $A \in \mathscr{A}_\mu$ and $B \subseteq X$ satisfy $B \subseteq A$ and $\overline{\mu}(A) = 0$ then, since $A \in \mathscr{A}_\mu$, we have $U \in \mathscr{A}$ so that $A \subseteq U$ and $\mu(U) = 0$. Taking $L = \varnothing$ we have $L \subseteq B \subseteq U$ and $\mu(U \setminus L) = 0$, showing that $B \in \mathscr{A}_\mu$, as desired. ■

It turns out that the construction in Theorem 2.3.13 of a measure space from an outer measure yields a complete measure space.

2.3.22 Proposition (Completeness of measure space constructed from outer measures) *If μ^* is an outer measure on a set X then $(\mathsf{X}, \mathscr{M}(\mathsf{X}, \mu^*), \mu^*|\mathscr{M}(\mathsf{X}, \mu^*))$ is a complete measure space.*

Proof From Theorem 2.3.13 we need only prove completeness. We let $\mu = \mu^*|\mathscr{M}(X, \mu^*)$. Let $B \in \mathscr{M}(X, \mu^*)$ and let $A \subseteq B$. For $S \in \mathbf{2}^X$ we then have

$$\begin{aligned}\mu^*(S \cap A) + \mu^*(S \cap (X \setminus A)) &\leq \mu^*(S \cap B) + \mu^*(S \cap (X \setminus A)) \\ &= 0 + \mu^*(S \cap (X \setminus A)) \leq \mu^*(S),\end{aligned}$$

using the fact that $\mu^*(S \cap B) \leq \mu^*(B) = 0$ and monotonicity of outer measures. By countable-subadditivity of μ^* we have

$$\mu^*(S) \leq \mu^*(S \cap A) + \mu^*(S \cap (X \setminus A)),$$

and so it follows that $A \in \mathscr{M}(X, \mu^*)$. ■

Let us finally show that completeness is preserved by restriction.

2.3.23 Proposition (The restriction of a complete measure is complete) *If $(\mathsf{X}, \mathscr{A}, \mu)$ is a complete measure space then the measure space $(\mathsf{A}, \mathscr{A}_\mathsf{A}, \mu_\mathsf{A})$ of Proposition 2.3.18 is complete.*

Proof If $B \cap A \in \mathscr{A}_A$ satisfies $\mu_A(B \cap A) = 0$ then $\mu(B \cap A) = 0$. Therefore, by completeness of μ, if $C \subseteq (B \cap A)$ it follows that $\mu_A(C) = 0$. ■

2.3.4 Outer and inner measures associated to a measure

In this section we continue our exploration of the relationship between outer measure and measure, now going from a measure to an outer measure. We begin with a discussion of ways in which one may generate an outer measure from other data.

2.3.24 Proposition (Outer measure generated by a collection of subsets) *Let X be a set, let $\mathscr{S} \subseteq \mathbf{2}^\mathsf{X}$ have the property that $\varnothing \in \mathscr{S}$, and let $\mu_0 \colon \mathscr{S} \to \overline{\mathbb{R}}_{\geq 0}$ have the property that*

$$\inf\{\mu_0(\mathsf{S}) \mid \mathsf{S} \in \mathscr{S}\} = 0.$$

If we define $\mu^* \colon \mathbf{2}^\mathsf{X} \to \overline{\mathbb{R}}_{\geq 0}$ by

$$\mu^*(\mathsf{A}) = \inf\Big\{\sum_{\mathsf{j}=1}^{\infty} \mu_0(\mathsf{S}_\mathsf{j}) \Bigm| \mathsf{A} \subseteq \bigcup_{\mathsf{j} \in \mathbb{Z}_{>0}} \mathsf{S}_\mathsf{j},\ \mathsf{S}_\mathsf{j} \in \mathscr{S},\ \mathsf{j} \in \mathbb{Z}_{>0}\Big\},$$

then μ^* is an outer measure on X. Moreover, if $\mathscr{S}$ is an algebra on X and if μ_0 is a countably-additive measure, then $\mu^*(\mathsf{S}) = \mu_0(\mathsf{S})$ for every $\mathsf{S} \in \mathscr{S}$.

Proof First let us show that $\mu^*(\varnothing) = 0$. Let $\epsilon \in \mathbb{R}_{>0}$. By hypothesis there exists $S \in \mathscr{S}$ such that $\mu_0(S) \leq \epsilon$, and since $\varnothing \subseteq S$ we have

$$\mu^*(\varnothing) \leq \mu_0(S) \leq \epsilon.$$

As this holds for every $\epsilon \in \mathbb{R}_{>0}$ it follows that $\mu^*(\emptyset) = 0$. That $\mu^*(A) \leq \mu^*(B)$ if $A \subseteq B$ is clear. Now let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a countable family of subsets of X. If $\sum_{j=1}^{\infty} \mu^*(A_j) = \infty$ then the property of countable-subadditivity holds for the family $(A_j)_{j\in\mathbb{Z}_{>0}}$. Thus suppose that $\sum_{j=1}^{\infty} \mu^*(A_j) < \infty$ and let $\epsilon \in \mathbb{R}_{>0}$. For each $j \in \mathbb{Z}_{>0}$ let $(S_{jk})_{k\in\mathbb{Z}_{>0}}$ be a family of subsets from $\mathscr{S}$ with the properties that $A_j \subseteq \cup_{k\in\mathbb{Z}_{>0}} S_{jk}$ and

$$\sum_{k=1}^{\infty} \mu_0(S_{jk}) < \mu^*(A_j) + \frac{\epsilon}{2^j},$$

this being possible by definition of μ^*. Then

$$\bigcup_{j\in\mathbb{Z}_{>0}} A_j \subseteq \bigcup_{j,k\in\mathbb{Z}_{>0}} S_{jk} \quad\Longrightarrow\quad \mu^*\Big(\bigcup_{j\in\mathbb{Z}_{>0}} A_j\Big) \leq \mu^*\Big(\bigcup_{j,k\in\mathbb{Z}_{>0}} S_{jk}\Big).$$

Also

$$\bigcup_{j,k\in\mathbb{Z}_{>0}} S_{jk} \subseteq \bigcup_{j,k\in\mathbb{Z}_{>0}} S_{jk} \quad\Longrightarrow\quad \mu^*\Big(\bigcup_{j,k\in\mathbb{Z}_{>0}} S_{jk}\Big) \leq \sum_{j,k=1}^{\infty} \mu_0(S_{jk}) < \sum_{j=1}^{\infty} \mu^*(A_j) + \epsilon,$$

using the fact that $\sum_{j=1}^{\infty} \frac{1}{2^j} = 1$ (see Example I-2.4.2–1). From this we conclude that

$$\mu^*\Big(\bigcup_{j\in\mathbb{Z}_{>0}} A_j\Big) \leq \sum_{j=1}^{\infty} \mu^*(A_j)$$

since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary in the above development. Thus shows that μ^* is indeed an outer measure.

Now we prove the final assertion. Let $S \in \mathscr{S}$. Since $S \subseteq S$ we have $\mu^*(S) \leq \mu_0(S)$. Now let $(S_j)_{j\in\mathbb{Z}_{>0}}$ be a family of subsets such that $S \subseteq \cup_{j\in\mathbb{Z}_{>0}} S_j$. Then we define

$$\begin{aligned}
\tilde{S}_1 &= S_1 \\
\tilde{S}_2 &= S_2 \cap (X \setminus S_1) \\
&\vdots \\
\tilde{S}_j &= S_j \cap (X \setminus S_1) \cap \cdots \cap (X \setminus S_{j-1}) \\
&\vdots
\end{aligned}$$

noting that the family of sets $(\tilde{S}_j)_{j\in\mathbb{Z}_{>0}}$ is in $\mathscr{S}$ since $\mathscr{S}$ is an algebra. Moreover, by construction, the sets $(\tilde{S}_j)_{j\in\mathbb{Z}_{>0}}$ are pairwise disjoint and satisfy

$$\bigcup_{j\in\mathbb{Z}_{>0}} S_j = \bigcup_{j\in\mathbb{Z}_{>0}} \tilde{S}_j.$$

Since $\tilde{S}_j \subseteq S_j$ we have

$$\sum_{j=1}^{\infty} \mu_0(\tilde{S}_j) \leq \sum_{j=1}^{\infty} \mu_0(S_j).$$

Now, for each $j \in \mathbb{Z}_{>0}$, define $T_j = S \cap \tilde{S}_j$, noting that $T_j \in \mathscr{S}$ since $\mathscr{S}$ is an algebra. Note that $S = \cup_{j \in \mathbb{Z}_{>0}} T_j$. Moreover, by construction the family of sets $(T_j)_{j \in \mathbb{Z}_{>0}}$ is disjoint. Since μ_0 is a measure we have

$$\mu_0(S) = \mu_0\Big(\bigcup_{j \in \mathbb{Z}_{>0}} \tilde{T}_j\Big) = \sum_{j=1}^{\infty} \mu_0(\tilde{T}_j).$$

Since $T_j \subseteq \tilde{S}_j$ we have

$$\sum_{j=1}^{\infty} \mu_0(T_j) \le \sum_{j=1}^{\infty} \mu_0(\tilde{S}_j),$$

giving

$$\sum_{j=1}^{\infty} \mu_0(S_j) \ge \mu_0(S).$$

This allows us to conclude that $\mu^*(S) \ge \mu_0(S)$, and so $\mu^*(S) = \mu_0(S)$, as desired. ■

The outer measure of the preceding proposition has a name.

2.3.25 Definition (Outer measure generated by a collection of sets and a function on those sets) Let X be a set, let $\mathscr{S} \subseteq \mathbf{2}^X$ have the property that $\emptyset \in \mathscr{S}$, and let $\mu_0 \colon \mathscr{S} \to \overline{\mathbb{R}}_{\ge 0}$ have the property that

$$\inf\{\mu_0(S) \mid S \in \mathscr{S}\} = 0.$$

The outer measure $\mu^* \colon \mathbf{2}^X \to \overline{\mathbb{R}}_{\ge 0}$ defined in Proposition 2.3.24 is the outer measure ***generated*** by the pair $(\mathscr{S}, \mu_0)$. •

Let us give an application of the preceding constructions. A common construction with measures is the extension of a $\overline{\mathbb{R}}_{\ge 0}$-valued function on a collection of subsets to a measure on the σ-algebra generated by the subsets. There are a number of such statements, but the one that we will use is the following.

2.3.26 Theorem (Hahn–Kolmogorov[5] Extension Theorem) *Let* X *be a set, let* $\mathscr{A}$ *be an algebra on* X*, and let* $\mu_0 \colon \mathscr{A} \to \overline{\mathbb{R}}_{\ge 0}$ *be a* σ*-finite measure on* $\mathscr{A}$*. Then there exists a unique measure* μ *on* $\sigma(\mathscr{S})$ *such that* $\mu(\mathrm{A}) = \mu_0(\mathrm{A})$ *for every* $\mathrm{A} \in \mathscr{A}$*.*

Proof First let us assume that $\mu_0(X) < \infty$. Let $\mu^* \colon \mathbf{2}^X \to \overline{\mathbb{R}}_{\ge 0}$ be the outer measure generated by $\mathscr{A}$ and μ_0 as in Proposition 2.3.24. Then Proposition 2.3.24 ensures that $\mu^*(A) = \mu_0(A)$ for every $A \in \mathscr{A}$.

We wish to show that $\mu^*|\sigma(\mathscr{A})$ is a measure. To do this we define $d_{\mu^*} \colon \mathbf{2}^X \times \mathbf{2}^X \to \mathbb{R}_{\ge 0}$ by

$$d_{\mu^*}(S, T) = \mu^*(S \triangle T),$$

[5]Hans Hahn (1879–1934) was an Austrian mathematician whose contributions to mathematics were primarily in the areas of set theory and functional analysis. Andrey Nikolaevich Kolmogorov (1903–1987) is an important Russian mathematician. He made essential contributions to analysis, algebra, and dynamical systems. He also established the axiomatic foundations of probability theory.

recalling from Section I-1.1.2 the definition of the symmetric complement $\triangle$. We clearly have $d_{\mu^*}(S,T) = d_{\mu^*}(T,S)$ for every $S,T \subseteq X$. Since μ^* is an outer measure we have

$$\begin{aligned} d_{\mu^*}(S,U) &= \mu^*(S\triangle U) \leq \mu^*((S\triangle T)\cup(T\triangle U)) \\ &\leq \mu^*(S\triangle T) + \mu^*(T\triangle U) = d_{\mu^*}(S,T) + d_{\mu^*}(T,U) \end{aligned}$$

for every $S,T,U \subseteq X$, using Exercise I-1.1.2. Thus d_{μ^*} is a semimetric on $\mathbf{2}^X$. Moreover, $d_{\mu^*}(S,T) = 0$ if and only if $\mu^*(S-T) = 0$ and $\mu^*(T-S) = 0$. Thus the implication

$$d_{\mu^*}(S,T) = 0 \quad \Longrightarrow \quad S = T$$

holds only if $(\mu^*)^{-1}(0) = \varnothing$. That is, d_{μ^*} is a metric if and only if the only set of μ^*-measure zero is the empty set. We claim that $\mu^*\colon \mathbf{2}^X \to \mathbb{R}_{\geq 0}$ is continuous with respect to the semimetric topology defined by d_{μ^*}. To see this, let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta = \epsilon$. Then, if S,T satisfy $d_{\mu^*}(S,T) < \delta$, we have

ref for this?

$$\begin{aligned} |\mu^*(S) - \mu^*(T)| &= |\mu^*(S\triangle\varnothing) - \mu^*(T\triangle\varnothing)| \\ &= |d_{\mu^*}(S,\varnothing) - d_{\mu^*}(T,\varnothing)| \leq d_{\mu^*}(S,T) = \epsilon, \end{aligned}$$

using Exercise I-1.1.2 and Proposition 1.1.3 (noting that this holds for semimetrics, as well as for metrics).

ref?

Now define $\mathrm{cl}(\mathscr{A})$ to be the closure of $\mathscr{A} \subseteq \mathbf{2}^X$ using the semimetric d_{μ^*}. Thus $B \in \mathrm{cl}(\mathscr{A})$ if there exists a sequence $(A_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathscr{A}$ such that $\lim_{j\to\infty} d_{\mu^*}(B,A_j) = 0$. We claim that $\mathrm{cl}(\mathscr{A})$ is a σ-algebra. Certainly $\varnothing \in \mathrm{cl}(\mathscr{A})$. Let $B \in \mathrm{cl}(\mathscr{A})$. Then there exists a sequence $(A_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathscr{A}$ such that $\lim_{j\to\infty} d_{\mu^*}(B,A_j) = 0$. Using Exercise I-1.1.2 we have

$$d_{\mu^*}(X\setminus B, X\setminus A_j) = d_{\mu^*}(B,A_j), \qquad j \in \mathbb{Z}_{>0}.$$

Thus

$$\lim_{j\to\infty} d_{\mu^*}(X\setminus B, X\setminus A_j) = 0$$

and so $X\setminus B \in \mathrm{cl}(\mathscr{A})$. Now let $B,C \in \mathrm{cl}(\mathscr{A})$ and let $(S_j)_{j\in\mathbb{Z}_{>0}}$ and $(T_j)_{j\in\mathbb{Z}_{>0}}$ be sequences in $\mathscr{A}$ such that

$$\lim_{j\to\infty} d_{\mu^*}(B,S_j) = 0, \quad \lim_{j\to\infty} d_{\mu^*}(C,T_j) = 0.$$

Then

$$\begin{aligned} \lim_{j\to\infty} d_{\mu^*}(B\cup C, S_j\cup T_j) &= \lim_{j\to\infty} \mu^*((B\cup C)\triangle(S_j\cup T_j)) \\ &\leq \lim_{j\to\infty} \mu^*((B\triangle S_j)\cup(C\triangle T_j)) \\ &\leq \lim_{j\to\infty} \mu^*(B\triangle S_j) + \lim_{j\to\infty} \mu^*(C\triangle T_j) = 0, \end{aligned}$$

using Exercise I-1.1.2. Thus $B\cup C \in \mathrm{cl}(\mathscr{A})$. This shows that $\mathrm{cl}(\mathscr{A})$ is an algebra. Now let $(B_j)_{j\in\mathbb{Z}_{>0}}$ be a countable family of subsets from $\mathrm{cl}(\mathscr{A})$. Define $C_k = \cup_{j=1}^k B_j$ so that $C_k \in \mathrm{cl}(\mathscr{A})$, $k \in \mathbb{Z}_{>0}$. Then

$$\begin{aligned} \lim_{k\to\infty} d_{\mu^*}(\cup_{j\in\mathbb{Z}_{>0}} B_j, C_k) &= \lim_{k\to\infty} \mu^*((\cup_{j\in\mathbb{Z}_{>0}} B_j)\triangle(\cup_{j=1}^k B_j)) \\ &\leq \lim_{k\to\infty} \mu^*(\cup_{j=k+1}^\infty B_j). \end{aligned}$$

Since $\mu^*(X) < \infty$ by assumption, the sequence $(\mu^*(\cup_{j=1}^k B_j))_{k\in\mathbb{Z}_{>0}}$ is a bounded monotonically increasing sequence, and so converges. This implies that

$$\lim_{k\to\infty} d_{\mu^*}(\cup_{j\in\mathbb{Z}_{>0}} B_j, C_k) = \lim_{k\to\infty} \mu^*(\cup_{j=k+1}^{\infty} B_j) = 0.$$

Thus $\cup_{j\in\mathbb{Z}_{>0}} B_j \in \mathrm{cl}(\mathscr{A})$ since $\mathrm{cl}(\mathscr{A})$ is closed and since $C_k \in \mathrm{cl}(\mathscr{A})$ for each $k \in \mathbb{Z}_{>0}$. This shows that $\mathrm{cl}(\mathscr{A})$ is a σ-algebra, as desired. ref for semimetrics

We will now show that $\mu^*|\,\mathrm{cl}(\mathscr{A})$ is a measure. We certainly have $\mu^*(\emptyset) = 0$. We next claim that $\mu^*|\,\mathrm{cl}(\mathscr{A})$ is finitely-additive. To see this, let $B, C \in \mathrm{cl}(\mathscr{A})$ be disjoint and let $(S_j)_{j\in\mathbb{Z}_{>0}}$ and $(T_j)_{j\in\mathbb{Z}_{>0}}$ be sequences in $\mathscr{A}$ such that

$$\lim_{j\to\infty} d_{\mu^*}(B, S_j) = 0, \quad \lim_{j\to\infty} d_{\mu^*}(C, T_j) = 0.$$

We then have, using continuity of μ^* and additivity of $\mu^*|\mathscr{A} = \mu_0$,

$$\mu^*(B \cup C) = \lim_{j\to\infty} \mu^*(S_j \cup T_j) = \lim_{j\to\infty} \mu^*(S_j) + \lim_{j\to\infty} \mu^*(T_j - S_j) = \mu^*(B) + \mu^*(C).$$

A simple induction then gives finite-additivity. Finally, let $(B_j)_{j\in\mathbb{Z}_{>0}}$ be a countable collection of disjoint sets from $\mathrm{cl}(\mathscr{A})$. Because μ^* is an outer measure we have

$$\mu^*\Big(\bigcup_{j\in\mathbb{Z}_{>0}} B_j\Big) \le \sum_{j=1}^{\infty} \mu^*(B_j).$$

Since $\mu^*|\,\mathrm{cl}(\mathscr{A})$ is finitely-additive we have

$$\mu^*\Big(\bigcup_{j\in\mathbb{Z}_{>0}} B_j\Big) \ge \mu^*\Big(\bigcup_{j=1}^{k} B_j\Big) = \sum_{j=1}^{k} \mu^*(B_j)$$

for every $k \in \mathbb{Z}_{>0}$. Thus

$$\mu^*\Big(\bigcup_{j\in\mathbb{Z}_{>0}} B_j\Big) \ge \sum_{j=1}^{\infty} \mu^*(B_j),$$

which allows us to conclude countable-additivity of $\mu^*|\,\mathrm{cl}(\mathscr{A})$.

Since $\mathscr{A} \subseteq \mathrm{cl}(\mathscr{A})$ it follows from Proposition 2.2.7 that $\sigma(\mathscr{A}) \subseteq \mathrm{cl}(\mathscr{A})$. Since $\mu^*|\,\mathrm{cl}(\mathscr{A})$ is a measure, it is surely also true that $\mu \triangleq \mu^*|\sigma(\mathscr{A})$ is a measure. This proves the existence assertion of the theorem under the assumption that $\mu_0(X) < \infty$.

For uniqueness, let $\tilde{\mu}\colon \mathrm{cl}(\mathscr{A}) \to \mathbb{R}_{\ge 0}$ be a measure having the property that $\tilde{\mu}|\mathscr{A} = \mu_0$. Let $B \in \sigma(\mathscr{A})$ and let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a family of subsets such that $B \subseteq \cup_{j\in\mathbb{Z}_{>0}} A_j$. Since $\tilde{\mu}|\mathscr{A} = \mu_0$ we have

$$\tilde{\mu}(B) \le \sum_{j=1}^{\infty} \tilde{\mu}(A_j) = \sum_{j=1}^{\infty} \mu_0(A_j),$$

using Proposition 2.3.10. From this we infer that

$$\tilde{\mu}(B) \le \inf\Big\{\sum_{j=1}^{\infty} \mu_0(A_j) \;\Big|\; B \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} A_j,\ A_j \in \mathscr{A},\ j \in \mathbb{Z}_{>0}\Big\} = \mu(B).$$

In like manner we have that $\tilde{\mu}(X \setminus B) \leq \mu(X \setminus B)$. Thus

$$\tilde{\mu}(B) = \tilde{\mu}(X) - \tilde{\mu}(X \setminus B) \geq \mu(X) - \mu(X \setminus B) = \mu(B).$$

Thus $\tilde{\mu}(B) = \mu(B)$, as desired.

Finally, we prove the theorem, removing the assumption that $\mu_0(X) < \infty$. Since the hypotheses of the theorem include μ_0 being σ-finite, there exists a countable collection $(Y_j)_{j\in\mathbb{Z}_{>0}}$ of subsets from $\mathscr{A}$ such that $\mu_0(Y_j) < \infty$, $j \in \mathbb{Z}_{>0}$, and such that $X = \cup_{j\in\mathbb{Z}_{>0}} Y_j$. Then define

$$\begin{aligned}
X_1 &= Y_1 \\
X_2 &= Y_2 \cap (X \setminus Y_1) \\
&\vdots \\
X_j &= Y_j \cap (X \setminus Y_1) \cap \cdots \cap (X \setminus Y_{j-1}) \\
&\vdots
\end{aligned}$$

noting that the family of sets $(X_j)_{j\in\mathbb{Z}_{>0}}$ is in $\mathscr{A}$ since $\mathscr{A}$ is an algebra. Moreover, by construction the sets $(X_j)_{j\in\mathbb{Z}_{>0}}$ are pairwise disjoint, have the property that $\mu(X_j) < \infty$, $j \in \mathbb{Z}_{>0}$, and satisfy $X = \cup_{j\in\mathbb{Z}_{>0}} X_j$. Denote

$$\mathscr{A}_j = \{X_j \cap A \mid A \in \mathscr{A}\}, \quad \sigma(\mathscr{A})_j = \{X_j \cap B \mid B \in \sigma(\mathscr{A})\}, \quad \mu_{0,j} = \mu_0|\mathscr{A}_j.$$

We claim that $\sigma(\mathscr{A})_j = \sigma(\mathscr{A}_j)$. To show this one must show that $\sigma(\mathscr{A})_j$ is a σ-algebra on X_j containing $\mathscr{A}_j$ and that any σ-algebra containing on X_j containing $\mathscr{A}_j$ contains $\sigma(\mathscr{A})_j$. It is a straightforward exercise manipulating sets to show that $\sigma(\mathscr{A})_j$ is a σ-algebra containing $\mathscr{A}_j$, and we leave this to a sufficiently bored reader. So let $\mathscr{A}_j'$ be a σ-algebra on X_j containing $\mathscr{A}_j$. Let

$$\mathscr{A}' = \{A \cup B \mid A \in \mathscr{A}_j', B = (X \setminus X_j) \cap B', \ B' \in \sigma(\mathscr{A})\}.$$

By Exercise 2.2.5 we conclude that $\mathscr{A}'$ is a σ-algebra on $X = X_j \cup (X \setminus X_j)$. Moreover, $\mathscr{A} \subseteq \mathscr{A}'$ and so $\sigma(\mathscr{A}) \subseteq \mathscr{A}'$. But this means that if $X_j \cap B \in \sigma(\mathscr{A})_j$ then $X_j \cap B \in \mathscr{A}_j'$, giving our claim.

Now note that, for each $j \in \mathbb{Z}_{>0}$, the data X_j, $\mathscr{A}_j$, and $\mu_{0,j}$ satisfy the hypotheses used in the first part of the proof. Therefore, there exists a measure μ_j on $\sigma(\mathscr{A}_j) = \sigma(\mathscr{A})_j$ agreeing with $\mu_{0,j}$ on $\mathscr{A}_j$. Now define $\mu\colon \sigma(\mathscr{A}) \to \overline{\mathbb{R}}_{\geq 0}$ by

$$\mu(B) = \sum_{j=1}^{\infty} \mu_j(X_j \cap B).$$

That μ is a measure is easily verified using the fact that μ_j, $j \in \mathbb{Z}_{>0}$ is a measure and that the family of sets $(X_j)_{j\in\mathbb{Z}_{>0}}$ is pairwise disjoint. We leave the straightforward working out of this to the, again sufficiently bored, reader. It is also clear that $\mu|\mathscr{A} = \mu_0$. This gives the existence part of the proof. For uniqueness, suppose that $\tilde{\mu}\colon \sigma(\mathscr{A}) \to \overline{\mathbb{R}}_{\geq 0}$ is a measure such that $\tilde{\mu}|\mathscr{A} = \mu_0$ and let $B \in \sigma(\mathscr{A})$. By uniqueness from the first part of the proof we have $\tilde{\mu}(X_j \cap B) = \mu(X_j \cap B)$. Therefore, by countable-additivity of $\tilde{\mu}$,

$$\tilde{\mu}(B) = \sum_{j=1}^{\infty} \tilde{\mu}(X_j \cap B) = \sum_{j=1}^{\infty} \mu(X_j \cap B) = \mu(B),$$

as desired. ∎

The proof of the preceding theorem introduced an important construction. As we shall not make use of this in any subsequent part of the text, let us expound a little on this here.

2.3.27 Remark (Semimetrics and measures) A key ingredient in our proof of the Hahn–Kolmogorov Extension Theorem was a semimetric associated with a measure. This construction can be generalised somewhat. Let X be a set, let $\mathscr{S} \subseteq \mathbf{2}^X$, and let $\mu\colon \mathscr{S} \to \mathbb{R}_{\geq 0}$ be a finite-valued finitely-subadditive measure, i.e.,

$$\mu\Big(\cup_{j=1}^{k} A_j\Big) \leq \sum_{j=1}^{k} \mu(A_j), \qquad A_1, \dots, A_k \in \mathscr{A}.$$

Then we define $d_\mu\colon \mathscr{S} \times \mathscr{S} \to \mathbb{R}_{\geq 0}$ by $d_\mu(S, T) = \mu(S \triangle T)$, recalling from Section I-1.1.2 the definition of the symmetric complement $\triangle$. As in the above proof, we can verify that d_μ is a semimetric, and is a metric if and only if the only set of measure zero is the empty set. If μ is not finite-valued, then we can instead use

$$d'_\mu(S, T) = \max\{1, \mu(S \triangle T)\},$$

with the same conclusions.

In the proof we used this semimetric to define, in a topological sense, the closure $\mathrm{cl}(\mathscr{A})$ of the algebra $\mathscr{A}$, and we showed that $\sigma(\mathscr{A}) \subseteq \mathrm{cl}(\mathscr{A})$. In fact, although we did not need this in the proof above, $\mathrm{cl}(\mathscr{A})$ is the completion of $\sigma(\mathscr{A})$. This gives a neat loop-closing for the use of the word "completion" in this context, since it gives this a standard topological meaning. The Hahn–Kolmogorov Extension Theorem, then, becomes sort of a result about the extension of uniformly continuous functions to the completion, *a la* Theorem 1.1.37. When one digs more deeply into measure theory *per se*, these sorts of matters become more important. •

Now let us both specialise and extend our discussion of outer measures generated by a collection of subsets. We consider in detail the situation where we begin with a measure space.

2.3.28 Definition (Inner and outer measure of a measure) Let $(X, \mathscr{A}, \mu)$ be a measure space.

(i) The ***outer measure*** associated to μ is the map $\mu^*\colon \mathbf{2}^X \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$\mu^*(S) = \inf\{\mu(A) \mid A \in \mathscr{A},\ S \subseteq A\}.$$

(ii) The ***inner measure*** associated to μ is the map $\mu_*\colon \mathbf{2}^X \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$\mu_*(S) = \sup\{\mu(A) \mid A \in \mathscr{A},\ A \subseteq S\}. \qquad \bullet$$

The following corollary to Proposition 2.3.24 answers one of the basic questions we raised upon defining the concept of an outer measure.

2.3.29 Corollary (The outer measure of a measure is an outer measure) *If $(X, \mathscr{A}, \mu)$ be a measure space then the outer measure μ^* associated to μ is an outer measure as per Definition 2.3.4.*

Proof Since a σ-algebra is an algebra and since countable unions of measurable sets are measurable, this follows directly from Proposition 2.3.24. ■

One way to interpret the preceding result is that it provides a natural way of extending a measure, possibly only defined on a strict subset of the collection of all subsets, to a means of measuring "size" for all subsets, and that this extension is, in fact, an outer measure. This provides, then, a nice characterisation how a measure approximates sets "from above." What about the rôle of the inner measure that approximates sets "from below"? The following result clarifies this rôle, and illustrates one place where completeness is important.

2.3.30 Proposition (Sets for which inner and outer measure agree are in the completion) *Let $(X, \mathscr{A}, \mu)$ be a measure space and let $A \subseteq X$ be such that $\mu^*(A) < \infty$. Then $\mu_*(A) = \mu^*(A)$ if and only if $A \in \mathscr{A}_\mu$.*

Proof Suppose that $A \in \mathscr{A}_\mu$ and let $L, U \in \mathscr{A}$ satisfy $L \subseteq A \subseteq U$ and $\mu(U \setminus L) = 0$. Then

$$\mu(L) \leq \mu_*(A) \leq \mu^*(A) \leq \mu(U),$$

giving $\mu_*(A) = \mu^*(A)$ since $\mu(L) = \mu(U)$.

Conversely, suppose that $\mu_*(A) = \mu^*(A)$. Let $k \in \mathbb{Z}_{>0}$. Then there exists sets $M_k, V_k \in \mathscr{A}$ such that $M_k \subseteq A \subseteq V_k$ and such that

$$\mu_*(A) < \mu(M_k) + \tfrac{1}{k}, \quad \mu(V_k) < \mu^*(A) + \tfrac{1}{k}.$$

Then, for $k \in \mathbb{Z}_{>0}$ define

$$L_k = \cup_{j=1}^k M_j \in \mathscr{A}, \quad U_k = \cap_{j=1}^k V_j \in \mathscr{A},$$

noting that $M_k \subseteq L_k \subseteq A$, $A \subseteq U_k \subseteq V_k$, $L_k \subseteq L_{k+1}$, and $U_{k+1} \subseteq U_k$ for $k \in \mathbb{Z}_{>0}$. We then have

$$\mu_*(A) - \tfrac{1}{k} < \mu(M_k) \leq \mu(L_k) \leq \mu(U_k) \leq \mu(V_k) < \mu^*(A) + \tfrac{1}{k}.$$

Taking the limit as $k \to \infty$ gives

$$\lim_{k\to\infty} \mu(L_k) = \lim_{k\to\infty} \mu(L_k).$$

If we define $L = \cup_{k\in\mathbb{Z}_{>0}} L_k \in \mathscr{A}$ and $U = \cap_{k\in\mathbb{Z}_{>0}} U_k \in \mathscr{A}$ then we have $L \subseteq A \subseteq U$ and, by Proposition 2.3.10, $\mu(L) = \mu(U)$. Thus $A \subseteq \mathscr{A}_\mu$. ■

2.3.5 Probability measures

In this section we introduce the notion of a probability measure. As the name suggests, probability measures arise naturally in the study of probability theory, but this is something we will not take up here, postponing a general study of this for . what?

Let us first define what we mean by a probability measure.

2.3.31 Definition (Probability space, probability measure) A ***probability space*** is a measure space $(X, \mathscr{A}, \mu)$ for which $\mu(X) = 1$. The set X is called the ***sample space***, the σ-algebra $\mathscr{A}$ is called the set of ***events***, and the measure μ is called a ***probability*** measure. •

Let us give some examples.

2.3.32 Examples (Probability spaces)

1. Let us consider the classical example of a problem in so-called "discrete probability." We suppose that we have a coin which, when we flip it, has two outcomes, denoted "H" for "heads" and "T" for "tails." Let us suppose that we know that the coin is biased in a known way, so that the likelihood of seeing a head on any flip is $p \in [0, 1]$. Then the likelihood of seeing a tail on any flip is $1 - p$. We shall flip this coin once, and the record the outcome. Thus the sample space is $X = \{\mathrm{H}, \mathrm{T}\}$. The σ-algebra of events we take to be $\mathscr{A} = \mathbf{2}^X$. Thus there are four events: (a) $\varnothing$ (corresponding to an outcome of neither "heads" nor "tails"); (b) $\{\mathrm{H}\}$ (corresponding to an outcome of "heads"); (c) $\{\mathrm{T}\}$ (corresponding to an outcome of "tails"); (d) $\{\mathrm{H}, \mathrm{T}\}$ (corresponding to an outcome of either "heads" or "tails"). The probability measure is defined by

$$\mu(\{\mathrm{H}\}) = p, \quad \mu(\{\mathrm{T}\}) = (1 - p).$$

The probability measure for the events $\varnothing$ and $\{\mathrm{H}, \mathrm{T}\}$ must be 0 (because the measure of the empty set is always zero) and 1 (by countable additivity of the measure), respectively. Thus μ is a probability measure.

2. We have a biased coin as above. But now we perform an trial where we flip the coin n times and record the outcome each time. An element of the sample space X is an outcome of a single trial. Thus an element of the sample space is an element of $X = \{\mathrm{H}, \mathrm{T}\}^{\{1,\dots,n\}}$, the set of maps from $\{1, \dots, n\}$ to $\{\mathrm{H}, \mathrm{T}\}$. Note that $\mathrm{card}(X) = 2^n$. If $\phi \in X$ then the outcome of this trial is represented by the sequence

$$(\phi(1), \dots, \phi(n)) \in \{\mathrm{H}, \mathrm{T}\}^n.$$

The σ-algebra defining the set of events is the set of subsets of all trials: $\mathscr{A} = \mathbf{2}^X$. Now let us define a meaningful probability measure. For a trial $\phi \in X$ let $n_{\mathrm{H}}(\phi)$ be the number of heads appearing in the trial and let $n_{\mathrm{T}}(\phi)$ be the number of tails appearing in the trial. Obviously, $n_{\mathrm{H}}(\phi) + n_{\mathrm{T}}(\phi) = n$ for every $\phi \in X$. We then define

$$\mu(\phi) = p^{n_{\mathrm{H}}(\phi)}(1-p)^{n_{\mathrm{T}}(\phi)}.$$

This then defines μ on $\mathbf{2}^X$ by countable additivity. We should check that this is a probability measure, i.e., that $\mu(X) = 1$. For fixed $k \in \{1, \dots, k\}$, the number of trials in which k heads appears is

$$\binom{n}{k} \triangleq \frac{n!}{k!(n-k)!},$$

i.e., the binomial coefficient $B_{n,k}$ from Exercise I-2.2.1. Note that, according to Exercise I-2.2.1,

$$\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1.$$

Since the expression on the left is the sum over the trials with any possible number of heads, it is the sum over all possible trials.

3. Consider the problem of "randomly" choosing a number in the interval $[0,1]$. Thus $X = [0,1]$. We wish to use the Lebesgue measure as a probability measure. Note that, according to our constructions of Section 2.4, to do this pretty much necessitates taking $\mathscr{A} = \mathscr{L}([0,1])$ as the set of events.
4. Let $x_0 \in \mathbb{R}$ and let $\sigma \in \mathbb{R}_{>0}$. Let us consider the sample space $X = \mathbb{R}$, the set of events $\mathscr{A} = \mathscr{L}(\mathbb{R})$, and the measure $\gamma_{x_0,\sigma} \colon \mathscr{L}(\mathbb{R}) \to \mathbb{R}$ defined by

$$\gamma_{x_0,\sigma}(A) = \frac{1}{\sqrt{2\pi}\sigma} \int_{\mathbb{R}} \chi_A(x) \exp(-\tfrac{1}{2\sigma^2}(x-x_0)^2)\,\mathrm{d}x.$$

We claim that $\gamma_{x_0,\sigma}$ is a probability measure, i.e., that $\gamma_{x_0,\sigma}(\mathbb{R}) = 1$. The following lemma is useful in verifying this.

1 Lemma $\int_{\mathbb{R}} \mathrm{e}^{-\xi^2}\,\mathrm{d}\xi = \sqrt{\pi}.$

Proof By Fubini's Theorem we write

$$\Big(\int_{\mathbb{R}} \mathrm{e}^{-\xi^2}\,\mathrm{d}\xi\Big)^2 = \Big(\int_{\mathbb{R}} \mathrm{e}^{-x^2}\,\mathrm{d}x\Big)\Big(\int_{\mathbb{R}} \mathrm{e}^{-y^2}\,\mathrm{d}y\Big) = \int_{\mathbb{R}^2} \mathrm{e}^{-x^2-y^2}\,\mathrm{d}x\mathrm{d}y.$$

By Example II-1.6.40–3 we have

$$\Big(\int_{\mathbb{R}} \mathrm{e}^{-\xi^2}\,\mathrm{d}\xi\Big)^2 = \int_{\mathbb{R}_{>0}\times[-\pi,\pi]} r\mathrm{e}^{-r^2}\mathrm{d}r\mathrm{d}\theta = 2\pi \int_{\mathbb{R}_{>0}} r\mathrm{e}^{-r^2}\,\mathrm{d}r.$$

Now we make another change of variable $\rho = r^2$ to obtain

$$\Big(\int_{\mathbb{R}} \mathrm{e}^{-\xi^2}\,\mathrm{d}\xi\Big)^2 = \pi \int_{\mathbb{R}_{>0}} \mathrm{e}^{-\rho}\,\mathrm{d}\rho = \pi,$$

and so we get the result. ▼

By making the change of variable $\xi = \frac{1}{\sqrt{2}\sigma}(x - x_0)$, we can then directly verify that $\gamma_{x_0,\sigma}(\mathbb{R}) = 1$. This probability measure is called the ***Gaussian measure*** with ***mean*** x_0 and ***variance*** σ. •

2.3.6 Product measures

In Section 2.2.3 we showed how algebras on the factors of a product give algebras and σ-algebras on the product. In this section we investigate how to define measures on products given measures on each of the factors. The procedure for this is surprisingly technical; we use the Hahn–Kolmogorov Extension Theorem. It is also possible to define measures on products using the integral, after the integral has been defined. We refer to Section 2.8.1 for this construction.

For now, let us state and prove the basic result concerning the construction of measures on products of measure spaces.

2.3.33 Theorem (Measures on products of measure spaces) *If* $(\mathsf{X}_j, \mathscr{A}_j, \mu_j)$, $j \in \{1, \dots, k\}$, *are σ-finite measure spaces then there exists a unique measure*

$$\mu_1 \times \cdots \times \mu_k : \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_k) \to \overline{\mathbb{R}}_{\geq 0}$$

such that

$$\mu_1 \times \cdots \times \mu_k(\mathsf{A}_1 \times \cdots \times \mathsf{A}_k) = \mu_1(\mathsf{A}_1) \cdots \mu_k(\mathsf{A}_k)$$

for every $\mathsf{A}_1 \times \cdots \times \mathsf{A}_k \in \mathscr{A}_1 \times \cdots \times \mathscr{A}_k$.

Proof We use a couple of technical lemmata.

1 Lemma *Let* X *be a set and let* $\mathscr{S}_0 \subseteq \mathbf{2}^{\mathsf{X}}$ *be a family of subsets for which*

(i) $\mathsf{S}_1 \cap \mathsf{S}_2 \in \mathscr{S}_0$ *for every* $\mathsf{S}_1, \mathsf{S}_2 \in \mathscr{S}_0$ *and*

(ii) if $\mathsf{S} \in \mathscr{S}_0$ *then* $\mathsf{X} \setminus \mathsf{S} = \mathsf{S}_1 \cup \cdots \cup \mathsf{S}_k$ *for some pairwise disjoint* $\mathsf{S}_1, \dots, \mathsf{S}_k \in \mathscr{S}_0$.

Then $\sigma_0(\mathscr{S}_0)$ *is equal to the collection of finite unions of sets from* $\mathscr{S}_0$ *and, if* $\mu_0 : \mathscr{S}_0 \to \overline{\mathbb{R}}_{\geq 0}$ *is finitely-additive, then there exists a unique finitely-additive function* $\mu_0 : \sigma_0(\mathscr{S}) \to \overline{\mathbb{R}}_{\geq 0}$ *such that* $\mu|\mathscr{S}_0 = \mu_0$.

Proof First we claim that the set of finite unions of sets from $\mathscr{S}_0$, let us denote this collection of subsets by $\overline{\mathscr{S}}_0$, is an algebra. To see that $X \in \overline{\mathscr{S}}_0$, let $S \in \mathscr{S}_0$ and write, by hypothesis,

$$X = S \cup (X \setminus S) = S \cup (S_1 \cdots \cup S_k)$$

for some $S_1, \dots, S_k \in \mathscr{S}_0$. Thus $X \in \overline{\mathscr{S}}_0$. Now let $S \in \overline{\mathscr{S}}_0$ and write $S = S_1 \cup \cdots \cup S_k$ for $S_1, \dots, S_k \in \mathscr{S}_0$. Then, by De Morgan's Laws,

$$X \setminus S = (X \setminus S_1) \cap \cdots \cap (X \setminus S_k).$$

Thus $X \setminus S$ is, by assumption, a finite intersection of finite unions of sets from $\mathscr{S}_0$. Since intersections of finitely many sets from $\mathscr{S}_0$ are in $\mathscr{S}_0$, it then follows that $X \setminus S \in \overline{\mathscr{S}}_0$. Thus, by Exercise 2.2.1, $\overline{\mathscr{S}}_0$ is an algebra. Moreover, if $\mathscr{A}$ is any algebra containing $\mathscr{S}_0$

then $\mathscr{A}$ must necessarily contain the finite unions of sets from $\mathscr{S}_0$. Thus $\overline{\mathscr{S}}_0 \subseteq \mathscr{A}$. By Proposition 2.2.8 this shows that $\overline{\mathscr{S}}_0 = \sigma_0(\mathscr{S}_0)$, as desired.

Now let $A \in \sigma_0(\mathscr{S}_0)$ so that $A = A_1 \cup \cdots \cup A_k$ for some $A_1, \dots, A_k \in \mathscr{S}_0$. By Lemma 1 in the proof of Proposition 2.3.2, there are then *disjoint* sets $T_1, \dots, T_m \in \mathscr{S}_0$ such that $A = T_1 \cup \cdots \cup T_m$. We then define

$$\mu(A) = \mu_0(T_1) + \cdots + \mu_0(T_m).$$

We must show that this definition is independent of the particular way in which one writes A as a disjoint union of sets from $\mathscr{S}_0$. Suppose that $A = T'_1 \cup \cdots \cup T'_n$ for disjoint $T'_1, \dots, T'_n \in \mathscr{S}_0$. Then

$$A = \cup_{j=1}^m T_j = \cup_{l=1}^n T'_l = \cup_{j=1}^m \cup_{l=1}^n T_j \cap T'_l,$$

as may be easily verified. It then follows that

$$\mu\big(\cup_{j=1}^m T_j\big) = \sum_{j=1}^m \mu_0(T_j) = \sum_{j=1}^m \sum_{l=1}^n \mu_0(T_j \cap T'_l) = \sum_{l=1}^n \sum_{j=1}^m \mu_0(T'_l \cap T_j) = \sum_{l=1}^n \mu_0(T'_l),$$

giving the well-definedness of μ, and so the existence assertion of the lemma. Uniqueness follows immediately from finite-additivity of μ. ▼

2 Lemma *For sets* $\mathsf{X}_1, \dots, \mathsf{X}_k$ *with algebras* $\mathscr{A}_j \subseteq \mathbf{2}^{\mathsf{X}_j}$, $j \in \{1, \dots, k\}$, *let* $\mu_j \colon \mathscr{A}_j \to \overline{\mathbb{R}}_{\geq 0}$, $j \in \{1, \dots, k\}$, *be finitely-additive. Then there exists a unique finitely-additive*

$$\mu \colon \sigma_0(\mathscr{A}_1 \times \cdots \times \mathscr{A}_k) \to \overline{\mathbb{R}}_{\geq 0}$$

such that

$$\mu(\mathsf{A}_1 \times \cdots \times \mathsf{A}_k) = \mu_1(\mathsf{A}_1) \cdots \mu_k(\mathsf{A}_k) \tag{2.4}$$

for every $\mathsf{A}_j \in \mathscr{A}_j$, $j \in \{1, \dots, k\}$.

Proof Let us abbreviate $\mathscr{A} = \sigma_0(\mathscr{A}_1 \times \cdots \times \mathscr{A}_k)$. By Proposition 2.2.16, if $A \in \mathscr{A}$ then we can write

$$A = R_1 \cup \cdots \cup R_m$$

for disjoint measurable rectangles $R_1, \dots, R_m$. We then define

$$\mu(A) = \mu(R_1) + \cdots + \mu(R_m), \tag{2.5}$$

where $\mu(R_j)$, $j \in \{1, \dots, m\}$, is defined as in (2.4). We must show that this definition of μ is independent of the way in which one expresses A as a finite disjoint union of measurable rectangles. First let us suppose that

$$A = A_1 \times \cdots \times A_k \in \mathscr{A}_1 \times \cdots \times \mathscr{A}_k.$$

We shall prove by induction on k that if A is written as a finite disjoint union of measurable rectangles, $A = R_1 \cup \cdots \cup R_m$, that (2.5) holds. This assertion is vacuous for $k = 1$, so assume it holds for $k = n - 1$ and let

$$A_1 \times \cdots \times A_n = \cup_{j=1}^m B'_j \times B_j$$

where $B'_j \in \mathscr{A}_1 \times \cdots \times \mathscr{A}_{n-1}$ and $B_j \in \mathscr{A}_n$ for each $j \in \{1, \dots, m\}$. By the induction hypothesis and by our knowing the volumes of measurable rectangles, there exists a finitely-additive function $\mu' \colon \sigma_0(\mathscr{A}_1 \times \cdots \times \mathscr{A}_{n-1}) \to \overline{\mathbb{R}}_{\geq 0}$ such that

$$\mu'(A'_1 \times \cdots \times A'_{n-1}) = \mu_1(A'_1) \cdots \mu_{n-1}(A'_{n-1})$$

for every $A'_1 \times \cdots \times A'_{n-1} \in \mathscr{A}_1 \times \cdots \times \mathscr{A}_{n-1}$. We are charged with showing that

$$\begin{aligned}\mu(A_1 \times \cdots \times A_n) &= \mu_1(A_1) \cdots \mu_{n-1}(A_{n-1})\mu_n(A_n) \\ &= \mu'(A_1 \times \cdots \times A_{n-1})\mu_n(A_n) = \sum_{j=1}^{m} \mu'(B'_j)\mu_n(B_j),\end{aligned}$$

the last equality being the only that is not obvious.

From Lemma 1 in the proof of Proposition 2.3.2, there exists pairwise disjoint sets $C_1, \dots, C_r \subseteq A_n$ such that each of the sets $B_1, \dots, B_k$ is a finite union of the sets $C_1, \dots, C_r$. Thus, for each $j \in \{1, \dots, k\}$, there exists pairwise disjoint sets $S_{j1}, \dots, S_{jm_j} \subseteq A_n$, taken from the collection of sets $C_1, \dots, C_r$, for which $B_j = S_{j1} \cup \cdots \cup S_{jm_j}$. Thus

$$A_1 \times \cdots \times A_n = \cup_{j=1}^{m} B'_j \times \left(\cup_{l_j=1}^{m_j} S_{jl_j}\right) = \cup_{j=1}^{m} \cup_{l_j=1}^{m_j} B'_j \times S_{jl_j}.$$

Now, for each $s \in \{1, \dots, r\}$, let $J_s \subseteq \{1, \dots, k\}$ be defined so that $j \in J_s$ if and only if there exists $l_j \in \{1, \dots, m_j\}$ (necessarily unique) such that $S_{jl_j} = C_s$. Then define $B''_s = \cup_{j \in J_s} B'_j$. Since the measurable rectangles $B'_j \times B_j$, $j \in \{1, \dots, k\}$, are pairwise disjoint, it follows that the measurable rectangles B'_j, $j \in J_s$, are pairwise disjoint. Also note that we then have

$$A_1 \times \cdots \times A_n = \cup_{s=1}^{r} B''_s \times C_s,$$

noting that $C_1, \dots, C_s$ are pairwise disjoint. This implies that $\cup_{s=1}^{r} C_s = A_n$. This, in turn, forces us to conclude that $B''_s = A_1 \times \cdots \times A_{n-1}$ for each $s \in \{1, \dots, r\}$.

Now let us use the above facts, along with the induction hypothesis. Finite-additivity of μ_n gives

$$\mu_n(B_j) = \sum_{l_j=1}^{m_j} \mu_n(S_{jl_j}), \qquad j \in \{1, \dots, k\},$$

and

$$\sum_{s=1}^{r} \mu_n(C_s) = \mu_n(A_n).$$

Also, finite-additivity of μ' gives

$$\mu'(A_1 \times \cdots \times A_{n-1}) = \mu'(\cup_{j \in J_s} B'_j) = \sum_{j \in J_s} \mu'(B_j).$$

Putting this all together gives

$$\begin{aligned}\sum_{j=1}^{k} \mu'(B'_j)\mu_n(B_j) &= \sum_{j=1}^{k} \mu'(B_n) \sum_{l_j=1}^{m_j} \mu_n(S_{jl_j}) = \sum_{s=1}^{r} \sum_{j \in J_s} \mu'(B'_j)\mu_n(C_s) \\ &= \mu'(A_1 \times \cdots \times A_{n-1})\mu_n(A_n).\end{aligned}$$

This proves that the definition of volume of measurable rectangles is independent of how these rectangles are decomposed into finite disjoint unions of measurable rectangles.

The existence part of the lemma now follows from Lemma 1, along with Proposition 2.2.16. Uniqueness immediately follows from Proposition 2.2.16, along with the uniqueness assertion from Lemma 1. ▼

We complete the proof by induction on k, the assertion being clear when $k = 1$. So suppose that the conclusions of the theorem hold for $k = 1, \dots, m-1$ for some $m \geq 2$, and let $(X_j, \mathscr{A}_j, \mu_j)$, $j \in \{1, \dots, m\}$, be measure spaces satisfying the hypotheses of the theorem. Let us denote $Y = X_1 \times \cdots \times X_{m-1}$ and $Z = X_m$, $\mathscr{B} = \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_{m-1})$ and $\mathscr{C} = \mathscr{A}_m$, and $\nu = \mu_1 \times \cdots \times \mu_{m-1}$ and $\lambda = \mu_m$. We use the induction hypothesis to define $\nu\colon \mathscr{B} \to \overline{\mathbb{R}}_{\geq 0}$. We now wish to show that there exists a unique map $\nu \times \lambda\colon \mathscr{B} \times \mathscr{C} \to \overline{\mathbb{R}}_{\geq 0}$ such that

$$\nu \times \lambda(B \times C) = \nu(B)\lambda(C)$$

for every $B \times C \in \mathscr{B} \times \mathscr{C}$. Note that by Proposition 2.2.16 and Lemma 2, and since a countably-additive measure is also finitely-additive, there exists a unique finitely-additive measure $\nu_0\colon \sigma_0(\mathscr{B} \times \mathscr{C}) \to \overline{\mathbb{R}}_{\geq 0}$ such that

$$\nu_0(B \times C) = \nu(B)\lambda(C)$$

for every $B \times C \in \mathscr{B} \times \mathscr{C}$. By the Hahn–Kolmogorov Extension Theorem we need only show that ν_0 is countably-additive.

Let us first suppose that ν and λ are finite. Then, by Proposition 2.3.3, it suffices to show that if $(A_j)_{j \in \mathbb{Z}_{>0}}$ is a sequence of subsets from $\sigma_0(\mathscr{B} \times \mathscr{C})$ such that $A_j \supseteq A_{j+1}$ and such that $\cap_{j \in \mathbb{Z}_{>0}} A_j = \emptyset$, then $\lim_{j \to \infty} \nu_0(A_j) = 0$. By Proposition 2.2.16, for each $j \in \mathbb{Z}_{>0}$ we have

$$A_j = \cup_{k=1}^{m_j} B_{jk} \times C_{jk}$$

for nonempty sets $B_{j1}, \dots, B_{1m_j} \in \mathscr{B}$ and $C_{j1}, \dots, C_{jm_j} \in \mathscr{C}$. Moreover, as we argued in the proof of Lemma 2, we may suppose without loss of generality that the sets $B_{j1}, \dots, B_{jm_j}$ are pairwise disjoint. Now define $f_j\colon Y \to \mathbb{R}_{\geq 0}$ by

$$f_j(y) = \begin{cases} \lambda(C_{jk}), & y \in B_{jk}, \\ 0, & y \notin \cup_{k=1}^{m_j} B_{jk}. \end{cases}$$

For $y \in Y$ and $j \in \mathbb{Z}_{>0}$ there exists a unique $k(j,y) \in \{1, \dots, m_j\}$ such that $y \in B_{jk(j,y)}$. Moreover, if $j_1 < j_2$ we have

$$C_{j_1 k(j_1,y)} = \{z \in Z \mid (y,z) \in A_{j_1}\} \subseteq \{z \in Z \mid (y,z) \in A_{j_2}\} = C_{j_2 k(j_2,y)}$$

Therefore, the sequence $(f_j(y))_{j \in \mathbb{Z}_{>0}}$ is monotonically decreasing for each $y \in Y$. Moreover, $\lim_{j \to \infty} f_j(y) = 0$ since

$$\cap_{j \in \mathbb{Z}_{>0}} C_{jk(j,y)} \subseteq \cap_{j \in \mathbb{Z}_{>0}} \{z \in Z \mid (y,z) \in A_j\} = \emptyset.$$

Now let $\epsilon \in \mathbb{R}_{>0}$ and $j \in \mathbb{Z}_{>0}$ and define

$$B_{j,\epsilon} = \{y \in Y \mid f_j(y) > \epsilon\}.$$

We can easily see that $B_{j,\epsilon} \subseteq \cup_{k=1}^{m_j} B_{jk}$, that $B_{j,\epsilon} \supseteq B_{j+1,\epsilon}$ for $j \in \mathbb{Z}_{>0}$, and that $\cap_{j\in\mathbb{Z}_{>0}} B_{j,\epsilon} = \emptyset$. We therefore compute

$$\nu_0(A_j) = \sum_{k=1}^{m_j} \nu(B_{jk})\lambda(C_{jk}) \le \nu(B_{j,\epsilon})\lambda(Z) + \nu(Y)\epsilon.$$

Since $\lim_{j\to\infty} \nu(B_{j,\epsilon}) = 0$ by Proposition 2.3.3, it follows that

$$\lim_{j\to\infty} \nu_0(A_j) \le \epsilon\nu(Y),$$

giving $\lim_{j\to\infty} \nu_0(A_j) = 0$ since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary. This shows that ν_0 is a measure on $\sigma_0(\mathscr{B} \times \mathscr{C})$.

Next suppose that ν and λ are not finite, but are σ-finite. Then let $(S_k)_{k\in\mathbb{Z}_{>0}}$ and $(T_k)_{k\in\mathbb{Z}_{>0}}$ be subsets of Y and Z, respectively, such that $\nu(S_k) < \infty$ and $\lambda(T_k) < \infty$ for $k \in \mathbb{Z}_{>0}$, and such that $Y = \cup_{k\in\mathbb{Z}_{>0}} S_k$ and $Z = \cup_{k\in\mathbb{Z}_{>0}} T_k$. We may without loss of generality suppose that $S_k \subseteq S_{k+1}$ and $T_k \subseteq T_{k+1}$ for $k \in \mathbb{Z}_{>0}$. Let us denote

$$\mathscr{B}_k = \{B \cap S_k \mid B \in \mathscr{B}\}, \quad \mathscr{C}_k = \{C \cap T_k \mid C \in \mathscr{C}\}$$

and $\nu_k = \nu_0|sB_k \times \mathscr{C}_k$, noting from what we have already proved that ν_k is a measure. Then, for disjoint sets $(A_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathscr{B} \times \mathscr{C}$ we have

$$\begin{aligned}\sum_{j=1}^{\infty} \nu_0(A_j) &= \sum_{j=1}^{\infty} \lim_{k\to\infty} \nu_0(A_j \cap (S_k \times T_k)) = \lim_{k\to\infty} \sum_{j=1}^{\infty} \nu_k(A_j \cap (S_k \times T_k)) \\ &= \lim_{k\to\infty} \nu_k(\cup_{j\in\mathbb{Z}_{>0}} A_j \cap (S_k \times T_k)) = \nu_0(\cup_{j\in\mathbb{Z}_{>0}} A_j).\end{aligned}$$

This shows that ν_0 is a measure on $\mathscr{B} \times \mathscr{C}$.

Finally, to complete the proof by induction, one needs only to reinstate the definitions $Y = X_1 \times \cdots \times X_{m-1}$ and $Z = X_m$, $\mathscr{B} = \sigma(\mathscr{A}_1 \times \cdots \times \mathscr{A}_{m-1})$ and $\mathscr{C} = \mathscr{A}_m$, and $\nu = \mu_1 \times \cdots \times \mu_{m-1}$ and $\lambda = \mu_m$, and then apply the induction hypothesis. ■

Let us name the measure from the preceding theorem.

2.3.34 Definition (Product measure) If $(X_j, \mathscr{A}_j, \mu_j)$, $j \in \{1, \dots, k\}$, are σ-finite measure spaces then the measure $\mu_1 \times \cdots \times \mu_k$ is the ***product measure***. •

Let us give simple examples of product measures.

2.3.35 Examples (Product measures)

1. Let X and Y be sets with $\mathscr{A}$ and $\mathscr{B}$ σ-algebras on X and Y, respectively. Define $\mu\colon \mathscr{A} \to \overline{\mathbb{R}}_{\ge 0}$ and $\nu\colon \mathscr{B} \to \overline{\mathbb{R}}_{\ge 0}$ by

$$\mu(A) = \begin{cases} 0, & A = \emptyset, \\ \infty, & A \ne \emptyset, \end{cases} \qquad \nu(B) = \begin{cases} 0, & B = \emptyset, \\ \infty, & B \ne \emptyset. \end{cases}$$

Then the map $\mu \times \nu \colon \sigma(\mathscr{A} \times \mathscr{B}) \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$\mu \times \nu(S) = \begin{cases} 0, & S = \emptyset, \\ \infty, & S \neq \emptyset, \end{cases}$$

is a measure and satisfies $\mu \times \nu(A \times B) = \mu(A)\nu(B)$. Note, however, that since μ and ν are not σ-finite, we cannot use Theorem 2.3.33 to assert the existence of this measure except in the trivial case when $X = Y = \emptyset$.

2. Let X and Y be sets with $\mathscr{A}$ and $\mathscr{B}$ σ-algebras on X and Y, respectively. Define $\mu \colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ and $\nu \colon \mathscr{B} \to \overline{\mathbb{R}}_{\geq 0}$ by

$$\mu(A) = \begin{cases} \operatorname{card}(A), & \operatorname{card}(A) < \infty, \\ \infty, & \text{otherwise}, \end{cases} \qquad \nu(B) = \begin{cases} \operatorname{card}(B), & \operatorname{card}(B) < \infty, \\ \infty, & \text{otherwise}. \end{cases}$$

Then the map $\mu \times \nu \colon \sigma(\mathscr{A} \times \mathscr{B}) \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$\mu \times \nu(S) = \begin{cases} \operatorname{card}(S), & \operatorname{card}(S) < \infty, \\ \infty, & \text{otherwise}, \end{cases}$$

satisfies $\mu \times \nu(A \times B) = \mu(A)\nu(B)$. By Theorem 2.3.33 we can infer that $\mu \times \nu$ is a measure and is the unique measure with this property. •

2.3.36 Remark (Completeness of product measures) The product measure of complete measure spaces may be incomplete. We shall see a concrete instance of this in Section 2.5.4, but it is revealing to see how this can arise in a general way. Suppose that we have complete measure spaces $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$. Let $A \subseteq X$ be a nonempty set such that $\mu(A) = 0$ (thus A is measurable since $(X, \mathscr{A}, \mu)$ is complete) and let $B \subseteq Y$ be a nonmeasurable set. (Note that it might happen that there are no sets A and B with these properties.) Note that $A \times B \subseteq A \times Y$ and that $A \times Y$ is measurable, being a product of measurable rectangles. Moreover, $\mu \times \nu(A \times Y) = \mu(A)\nu(Y) = 0$ and so $A \times B$ is a subset of a set of measure zero. However, we claim that $A \times B$ is not $\sigma(\mathscr{A} \times \mathscr{B})$-measurable. Indeed, by Proposition 2.2.18, were $A \times B$ to be $\sigma(\mathscr{A} \times \mathscr{B})$-measurable, it would follow that B is $\mathscr{B}$-measurable, which we suppose not to be the case. •

2.3.7 Signed measures

In this section until now, a measure has been thought of as measuring the "size" of a measurable set, and so is an intrinsically nonnegative quantity. However, sometimes one wishes to use measures in ways more subtle than simply to measure "size," and in this case one wishes to allow for the measure of a set to be negative. In this section we carry out the steps needed to make such a definition, and we give a few basic properties of the sorts of measures we produce. The most interesting examples arise through integration; see Proposition 2.7.65. However, in Theorem 2.3.42 we will characterise signed measures to the degree that it is easy to see exactly what they "are."

We can begin with the definition.

2.3.37 Definition (Signed measure) For a measurable space $(X, \mathscr{A})$, a ***signed measure*** on $\mathscr{A}$ is a map $\mu\colon \mathscr{A} \to \overline{\mathbb{R}}$ such that

(i) $\mu(\varnothing) = 0$ and

(ii) μ is countably-additive.

A ***signed measure space*** is a triple $(X, \mathscr{A}, \mu)$ where $(X, \mathscr{A})$ is a measurable space and μ is a signed measure on $\mathscr{A}$. •

Note that, by Proposition 2.3.2(viii), a signed measure is consistent, and so a signed measure cannot take both values ∞ and $-\infty$. If, for emphasis, we wish to differentiate between a signed measure and a measure in the sense of Definition 2.3.7, we shall sometimes call the latter a ***positive measure***. However, whenever we say "measure," we always mean a measure in the sense of Definition 2.3.7.

Let us provide some simple examples of signed measures.

2.3.38 Examples (Signed measures)

1. Let X be a set and let $x_1, x_2 \in X$ be distinct points. Let us take $\mathscr{A} = \mathbf{2}^X$ and define $\mu\colon \mathbf{2}^X \to \overline{\mathbb{R}}$ by
$$\mu(A) = \begin{cases} m_1, & x_1 \in A,\ x_2 \notin A, \\ -m_2, & x_2 \in A,\ x_1 \notin A, \\ m_1 - m_2, & x_1, x_2 \in A, \\ 0, & x_1, x_2 \notin A, \end{cases}$$
for $m_1, m_2 \in \mathbb{R}$. Intuitively, μ has a positive mass m_1 at x_1 and a negative mass $-m_2$ and x_2.

2. Let $X = \mathbb{Z}$ be a set and take $\mathscr{A} = \mathbf{2}^X$. Suppose that the sequences $(p_j)_{j\in\mathbb{Z}_{\geq 0}}$ and $(n_j)_{j\in\mathbb{Z}_{>0}}$ of positive numbers are such that
$$\sum_{j=0}^{\infty} p_j < \infty, \qquad \sum_{j=1}^{\infty} n_j < \infty.$$
For $A \subseteq \mathbb{Z}$ define
$$\mu(A) = \sum_{j\in A\cap\mathbb{Z}_{\geq 0}} p_j - \sum_{j\in A\cap\mathbb{Z}_{<0}} n_{-j},$$
which can easily be verified to define a signed measure. •

Let us now indicate some of the essential features of signed measures.

2.3.39 Definition (Positive and negative sets, Hahn decomposition) For a signed measure space $(X, \mathscr{A}, \mu)$, a set $A \in \mathscr{A}$ is ***positive*** (resp. ***negative***) if, for every $B \subseteq A$ such that $B \in \mathscr{A}$, it holds that $\mu(B) \in \mathbb{R}_{\geq 0}$ (resp. $\mu(B) \in \mathbb{R}_{\leq 0}$). A ***Hahn decomposition*** for $(X, \mathscr{A}, \mu)$ is a pair (P, N) with the following properties:

(i) $P, N \in \mathscr{A}$;

(ii) $X = P \cup N$ and $P \cap N = \emptyset$;

(iii) P is a positive set and N is a negative set. •

It is clear that if A is a positive (resp. negative) set, every measurable subset of A is also positive (resp. negative).

We can prove that Hahn decompositions exist.

2.3.40 Theorem (Hahn Decomposition Theorem) *Every signed measure space possesses a Hahn decomposition. Moreover, if* (P_1, N_1) *and* (P_2, N_2) *are Hahn decompositions for a signed measure space* $(X, \mathscr{A}, \mu)$*, then* $P_1 \cap N_2$ *and* $P_2 \cap N_1$ *both have measure zero.*

Proof Since μ is consistent, we assume without loss of generality that μ cannot take the value $-\infty$. Let us define

$$L = \inf\{\mu(A) \mid A \text{ is a negative set}\}.$$

Note that there are negative sets since $\emptyset$ is negative. Also, $L > -\infty$. Indeed, if $L = -\infty$ this would imply that for each $j \in \mathbb{Z}_{>0}$ there exists a negative set A_j for which $\mu(A_j) < -j$. Let $B_k = \cup_{j=1}^{k} A_k$ so that $B_k \subseteq B_{k+1}$. Note that $\mu(B_k) < -k$. Countable-additivity of μ and Proposition 2.3.3 imply that

$$\mu\Big(\bigcup_{k \in \mathbb{Z}_{>0}} B_k\Big) = \lim_{k \to \infty} \mu(B_k) = -\infty,$$

and so indeed we must have $L > -\infty$ if μ cannot take the value $-\infty$. Now let $(A_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence of sets from $\mathscr{A}$ for which $\lim_{j \to \infty} \mu(A_j) = L$ and define $N = \cup_{j \in \mathbb{Z}_{>0}} A_j$. We claim that N is a negative set. Certainly $N \in \mathscr{A}$, N being a countable union of sets from $\mathscr{A}$. By Lemma 1 from the proof of Proposition 2.3.2 we can write $N = \cup_{j \in \infty} N_j$ for a pairwise disjoint family of negative sets $(N_j)_{j \in \mathbb{Z}_{>0}}$. Now, if $A \subseteq N$ is $\mathscr{A}$-measurable then $A = \cup_{j \in \mathbb{Z}_{>0}} A \cap N_j$. Since $A \cap N_j \subseteq N_j$ it follows that $\mu(A \cap N_j) \in \mathbb{R}_{\leq 0}$. Thus, by countable-additivity of μ,

$$\mu(A) = \sum_{j=1}^{\infty} \mu(A \cap N_j) \leq 0,$$

so showing that N is a negative set.

Now define $P = X \setminus N$. To prove that P is a positive set, we need a lemma.

1 Lemma *If* $(X, \mathscr{A}, \mu)$ *is a signed measure space and if* $A \in \mathscr{A}$ *satisfies* $\mu(A) \in \mathbb{R}_{<0}$*, then there exists a negative set* $B \subseteq A$ *such that* $\mu(B) \leq \mu(A)$.

Proof We define a sequence $(m_j)_{j \in \mathbb{Z}_{>0}}$ of nonnegative real numbers and a sequence $(A_j)_{j \in \mathbb{Z}_{>0}}$ of pairwise disjoint $\mathscr{A}$-measurable subsets of A with nonnegative measure as follows. Let

$$m_1 = \sup\{\mu(B) \mid B \in \mathscr{A},\ B \subseteq A\}.$$

Note that $m_1 \in \mathbb{R}_{\geq 0}$ since $\emptyset \in \mathscr{A}$ and $\emptyset \subseteq A$. Now let $A_1 \in \mathscr{A}$ be a subset of A that satisfies $\mu(A_1) \geq \min\{\frac{m_1}{2}, 1\}$, this being possible by the definition of m_1. Note that $\mu(A_1) \in \mathbb{R}_{\geq 0}$. Now suppose that we have defined $m_1, \dots, m_k \in \mathbb{R}_{\geq 0}$ and pairwise disjoint $\mathscr{A}$-measurable sets $A_1, \dots, A_k \subseteq A$ such that $\mu(A_j) \in \mathbb{R}_{\geq 0}$, $j \in \{1, \dots, k\}$. Then let

$$m_{k+1} = \sup\{\mu(B) \mid B \in \mathscr{A},\ B \subseteq A \setminus \cup_{j=1}^k A_j\}$$

and let $A_{k+1} \subseteq A \setminus \cup_{j=1}^k A_j$ have the property that $\mu(A_{k+1}) \geq \min\{\frac{m_{k+1}}{2}, 1\}$. As we argued above for m_1 and A_1, $m_{k+1}, \mu(A_{k+1}) \in \mathbb{R}_{\geq 0}$. It is clear that $A_{k+1} \cap A_j = \emptyset$, $j \in \{1, \dots, k\}$. Thus $(A_1, \dots, A_{k+1})$ are pairwise disjoint.

Let us take $B = A \setminus \cup_{j \in \mathbb{Z}_{>0}} A_j$. Note that

$$\mu\Big(\bigcup_{j \in \mathbb{Z}_{>0}} A_j\Big) = \sum_{j=1}^{\infty} \mu(A_j)$$

since the sets $(A_j)_{j \in \mathbb{Z}_{>0}}$ are pairwise disjoint. Therefore,

$$\mu(A) = \mu(B) + \mu\Big(\bigcup_{j \in \mathbb{Z}_{>0}} A_j\Big) \geq \mu(B).$$

Now we show that B is a negative set. Note that

$$\mu\Big(\bigcup_{j \in \mathbb{Z}_{>0}} A_j\Big) = \sum_{j=1}^{\infty} \mu(A_j) < \infty.$$

since $|\mu(A)| < \infty$. Thus the sum in the middle converges, and by Proposition I-2.4.7 it follows that $\lim_{j\to\infty} \mu(A_j) = 0$. Therefore, $\lim_{j\to\infty} m_j = 0$. Now let $E \subseteq B$ be $\mathscr{A}$-measurable. Thus $E \subseteq A \setminus \cup_{j=1}^k A_j$ for every $k \in \mathbb{Z}_{>0}$. Therefore, by definition of m_{k_1}, $\mu(E) \leq m_{k+1}$ for every $k \in \mathbb{Z}_{\geq 0}$. Therefore, it must be the case that $\mu(E) \in \mathbb{R}_{\leq 0}$ since $\lim_{j\to\infty} m_j = 0$. ▼

Now suppose that there exists a set $A \subseteq P$ such that $\mu(A) \in \mathbb{R}_{<0}$. Then, by the lemma, there exists a negative set $B \subseteq A$ such that $\mu(B) \leq \mu(A)$. Now $N \cup B$ is a negative set such that

$$\mu(N \cup B) = \mu(N) + \mu(B) \leq \mu(N) + \mu(A) < \mu(N) = L,$$

which contradicts the definition of L. Thus P is indeed positive.

To prove the final assertion of the theorem, note that both $P_1 \cap N_2$ and $P_2 \cap N_1$ are both positive and negative sets. It must, therefore, be the case that both have measure zero. ■

The Hahn decomposition can be illustrated for our examples above of signed measures.

2.3.41 Examples (Hahn decomposition)

1. We consider Example 2.3.38–1. A Hahn decomposition in this case consists of any subsets P and N such that
   (a) $P \cap N = \emptyset$,
   (b) $P \cup N = X$, and
   (c) $x_1 \in P$ and $x_2 \in N$.

   Note that there will generally be many possible Hahn decompositions in this case, since there are possible many sets of measure zero.
2. For Example 2.3.38–2, a Hahn decomposition is given by $P = \mathbb{Z}_{\geq 0}$ and $N = \mathbb{Z}_{<0}$. If none of the numbers p_j, $j \in \mathbb{Z}_{\geq 0}$, and n_j, $j \in \mathbb{Z}_{>0}$, are zero (as was assumed), then this is the *only* Hahn decomposition. •

As a direct consequence of the Hahn Decomposition Theorem we have the following decomposition of μ.

2.3.42 Theorem (Jordan Decomposition Theorem) *For a measurable space* $(\mathrm{X}, \mathscr{A})$ *the following statement hold:*

(i) if ν_+ and ν_- are two positive measures on $\mathscr{A}$, at least one of which is finite, then the map $\nu\colon \mathscr{A} \to \overline{\mathbb{R}}$ defined by $\nu(\mathrm{A}) = \nu_+(\mathrm{A}) - \nu_-(\mathrm{A})$ is a signed measure on $\mathscr{A}$;

(ii) if μ is a signed measure on $\mathscr{A}$ then there exist unique positive measures μ_+ and μ_- on $\mathscr{A}$ such that

(a) at least one of μ_+ and μ_- is finite,

(b) $\mu(\mathrm{A}) = \mu_+(\mathrm{A}) - \mu_-(\mathrm{A})$ for every $\mathrm{A} \in \mathscr{A}$, and

(c) $\mu_+(\mathrm{A}) = \mu(\mathrm{A})$ for every positive set A and $\mu_-(\mathrm{B}) = -\mu(\mathrm{B})$ for every negative set B;

(iii) if ν_+ and ν_- are positive measures on $\mathscr{A}$, at least one of which is finite, such that $\mu(\mathrm{A}) = \nu_+(\mathrm{A}) - \nu_-(\mathrm{A})$ for every $\mathrm{A} \in \mathscr{A}$ and if μ_+ and μ_- are as in part (ii), then $\nu_+(\mathrm{A}) \geq \mu_+(\mathrm{A})$ and $\nu_-(\mathrm{A}) \geq \mu_-(\mathrm{A})$ for every $\mathrm{A} \in \mathscr{A}$.

Proof (i) This is a straightforward verification that ν as defined in the statement of the theorem is countably-additive and satisfies $\mu(\emptyset) = 0$.

(ii) Let (P, N) be a Hahn decomposition for $(X, \mathscr{A}, \mu)$. Note that at most one of the relations $\mu(P) = \infty$ and $\mu(N) = -\infty$ can hold by consistency of μ. Define $\mu_+, \mu_-\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ by

$$\mu_+(A) = \mu(P \cap A), \quad \mu_-(A) = -\mu(N \cap A).$$

Clearly $\mu_+(\emptyset) = \mu_-(\emptyset) = 0$. Also, for a pairwise disjoint family $(A_j)_{j \in \mathbb{Z}_{>0}}$ of $\mathscr{A}$-measurable sets, we have

$$\mu_+\Big(\bigcup_{j \in \mathbb{Z}_{>0}} A_j\Big) = \mu\Big(P \cap \bigcup_{j \in \mathbb{Z}_{>0}} A_j\Big) = \mu\Big(\bigcup_{j \in \mathbb{Z}_{>0}} P \cap A_j\Big) = \sum_{j=1}^{\infty} \mu(P \cap A_j) = \sum_{j=1}^{\infty} \mu_+(A_j),$$

giving countable-additivity of μ_+. One similarly shows countable-additivity of μ_-. Also, if $A \in \mathscr{A}$, we have

$$\mu(A) = \mu(P \cap A) + \mu(N \cap A) = \mu_+(A) - \mu_-(A).$$

This gives the existence assertion of this part of the theorem.

We make an observation before we begin the proof of the uniqueness assertion of this part of the theorem. We continue with the notation from the proof of existence above, with μ_+ and μ_- as defined in that part of the proof, relative to the Hahn decomposition (P, N) for $(X, \mathscr{A}, \mu)$. Let (P', N') be another Hahn decomposition. We can then write

$$P = (P' \cap P) \cup (N' \cap P),$$

where, by Theorem 2.3.40, $\mu(N' \cap P) = 0$. Now note that, for every $A \in \mathscr{A}$,

$$\begin{aligned}\mu(P \cap A) &= \mu(((P' \cap P) \cup (N' \cap P)) \cap A)\\ &= \mu(((P' \cap P) \cap A) \cup ((N' \cap P) \cap A)) = \mu(((P' \cap P) \cap A)).\end{aligned}$$

Now we have

$$P' = (P' \cap P) \cup (P' \cap N),$$

where $\mu(P' \cap N) = 0$ by Theorem 2.3.40. Therefore,

$$\mu(P' \cap A) = \mu((P' \cap P) \cap A),$$

from which we deduce that $\mu(P \cap A) = \mu(P' \cap A)$. Similarly, we show that $\mu(N \cap A) = \mu(N' \cap A)$.

Let μ'_+ and μ'_- be positive measures satisfying

$$\mu(A) = \mu'_+(A) - \mu'_-(A), \qquad A \in \mathscr{A},$$

and suppose that $\mu'_+(A) = \mu(A)$ for every positive set A and that $\mu'_-(B) = \mu(B)$ for every negative set B. Let $A \in \mathscr{A}$ be a positive set and let $B \in \mathscr{A}$ be a negative set. Then, for the Hahn decomposition (P, N), we write

$$A = (P \cap A) \cup (N \cap A).$$

Since A is a positive set, we must have $\mu(N \cap A) = 0$. Define $P' = P \cup (N \cap A)$ and $N' = X \setminus P'$. Obviously P' is a positive set, being the union of a positive set with a set of measure zero. Since $N' = N \setminus (N \cap A)$, it follows that N' is a negative set. Thus (P', N') is a Hahn decomposition. Moreover, $P' \cap A = A$, and so

$$\mu'_+(A) = \mu(P' \cap A) = \mu(P \cap A) = \mu_+(A),$$

the second equality following from the remarks beginning this part of the proof. Similarly one shows that $\mu'_-(B) = \mu_-(B)$. Thus any positive measures μ'_+ and μ'_- having the three stated properties must agree with the measures μ_+ and μ_- explicitly constructed in part (ii).

(iii) For a positive set A we have

$$\mu(A) = \mu_+(A) = \nu_+(A) - \nu_-(A)$$

and so $\nu_+(A) \geq \mu_+(A)$ for every positive set A. For a negative set B we have $\mu_+(B) = 0$ and so we immediately have $\nu_+(B) \geq \mu_+(B)$. Therefore, for $A \in \mathscr{A}$ we have

$$\begin{aligned} & A = (P \cap A) \cup (N \cap A) \\ \implies \quad & \nu_+(A) = \nu_+(P \cap A) + \nu_+(N \cap A) \geq \mu_+(P \cap A) + \mu_+(N \cap A) = \mu_+(A). \end{aligned}$$

By the same arguments, *mutatis mutandis*, one shows that $\nu_-(A) \geq \mu_-(A)$ for every $A \in \mathscr{A}$. ■

Note that, without all of the assumptions from part (ii) of the theorem, uniqueness of μ_+ and μ_- cannot be guaranteed. Indeed, if μ is a positive measure then we can write

$$\mu(A) = \mu_+(A) - \mu_-(A) = \nu_+(A) - \nu_-(B)$$

where $\mu_+ = \mu$, μ_- is the zero measure, $\nu_+ = 2\mu$, and $\nu_- = \mu$. Note that $\nu_+(A) \geq \mu_+(A)$ and $\nu_-(A) \geq \mu_-(A)$, as asserted in part (iii).

Thus we make the following definition.

2.3.43 Definition (Jordan decomposition) If $(X, \mathscr{A}, \mu)$ is a signed measure space, the ***positive part*** and the ***negative part*** of μ are the positive measures μ_+ and μ_-, respectively, having the following properties:

(i) $\mu(A) = \mu_+(A) - \mu_-(A)$ for every $A \in \mathscr{A}$;

(ii) $\mu'_+(A) = \mu(A)$ for every positive set A;

(iii) $\mu'_-(B) = -\mu(B)$ for every negative set B.

The ***Jordan decomposition*** of μ is given by the representation $\mu = \mu_+ - \mu_-$ which signifies the first of the above properties of μ_+ and μ_-. •

2.3.44 Remark (Connections to functions with bounded variation) In Theorem I-3.3.3 we considered the Jordan decomposition for a function of bounded variation. This decomposition, like the one in Theorem 2.3.42, gives an additive decomposition with a (sort of) positive component and a (sort of) negative component. There is, as one might hope, a concrete relationship between the two Jordan decompositions.
what? However, this will not be realised until . •

For our ongoing examples we can illustrate the Jordan decomposition.

2.3.45 Examples (Jordan decomposition)

1. For the signed measure of Example 2.3.38–1, the positive and negative parts of the signed measure μ are defined by

$$\mu_+(A) = \begin{cases} m_1, & x_1 \in A, \\ 0, & x_1 \notin A, \end{cases} \quad \mu_-(A) = \begin{cases} m_2, & x_2 \in A, \\ 0, & x_2 \notin A. \end{cases}$$

2. For the signed measure of Example 2.3.38–2, the positive and negative parts of the signed measure μ are defined by

$$\mu_+(A) = \begin{cases} \sum_{j\in A\cap\mathbb{Z}_{\ge 0}} p_j, & A\cap\mathbb{Z}_{\ge 0} \neq \emptyset, \\ 0, & A\cap\mathbb{Z}_{\ge 0} = \emptyset, \end{cases} \quad \mu_-(A) = \begin{cases} \sum_{j\in A\cap\mathbb{Z}_{<0}} n_{-j}, & A\cap\mathbb{Z}_{<0} \neq \emptyset, \\ 0, & A\cap\mathbb{Z}_{<0} = \emptyset. \end{cases}$$

•

Now that we have at hand the decompositions which we use to characterise signed measures, we can use these to provide a new measure associated with a signed measure. The value of this construction may not be immediately apparent, but will be made clear in . what?

2.3.46 Definition (Variation and total variation of a signed measure) For a signed measure space $(X, \mathscr{A}, \mu)$, the ***variation*** of μ is the positive measure $|\mu|\colon \mathscr{A} \to \overline{\mathbb{R}}_{\ge 0}$ defined by

$$|\mu|(A) = \mu_+(A) + \mu_-(A),$$

where μ_+ and μ_- are the positive and negative parts, respectively, μ. The ***total variation*** of μ is $\|\mu\| = |\mu|(X)$ •

It is a simple verification to check that $|\mu|$ is indeed a positive measure. The following result characterises it among all positive measures which relate to μ in a prescribed manner.

2.3.47 Proposition (Property of the variation of a signed measure) *For $(X, \mathscr{A}, \mu)$ a signed measure space, $|\mu(A)| \le |\mu|(A)$ for all $A \in \mathscr{A}$. Moreover, if $\nu\colon \mathscr{A} \to \overline{\mathbb{R}}_{\ge 0}$ is a positive measure such that $|\mu(A)| \le \nu(A)$ for every $A \in \mathscr{A}$, then $|\mu|(A) \le \nu(A)$ for every $A \in \mathscr{A}$.*

Proof The first assertion of the result is clear, for if $A \in \mathscr{A}$ then

$$|\mu(A)| = |\mu_+(A) - \mu_-(A)| \le \mu_+(A) + \mu_-(A) = |\mu|(A).$$

For the second assertion, suppose that ν is a positive measure with the property that $|\mu(A)| \le \nu(A)$ for every $A \in \mathscr{A}$. If (P, N) is a Hahn decomposition for $(X, \mathscr{A}, \mu)$ then, for any $A \in \mathscr{A}$,

$$\mu_+(P \cap A) = |\mu(P \cap A)| \le \nu(P \cap A)$$

and

$$\mu_-(N \cap A) = |\mu(N \cap A)| \le \nu(N \cap A).$$

Therefore, using the definition of μ_+ and μ_-,

$$|\mu|(A) = \mu_+(A) + \mu_-(A) = \mu_+(P \cap A) + \mu_-(N \cap A) \le \nu(P \cap A) + \nu(N \cap A) = \nu(A),$$

as desired. ■

The following property of the variation of a signed measure is also useful.

2.3.48 Proposition (Characterisation of the variation of a signed measure) *For a signed measure space* $(\mathsf{X}, \mathscr{A}, \mu)$ *and for* $\mathsf{A} \in \mathscr{A}$*,*

$$|\mu|(\mathsf{A}) = \sup\Big\{\sum_{\mathsf{j}=1}^{\mathsf{k}} |\mu(\mathsf{A}_\mathsf{j})| \;\Big|\; (\mathsf{A}_1, \dots, \mathsf{A}_\mathsf{k}) \text{ is a partition of } \mathsf{A}\Big\}.$$

Proof Let $A \in \mathscr{A}$. For a partition $(A_1, \dots, A_k)$ of A we have

$$|\mu|(A) = \sum_{j=1}^{k} |\mu|(A_j) \geq \sum_{j=1}^{k} |\mu(A_j)|.$$

by Proposition 2.3.47 and using countable-additivity (and hence finite-additivity) of $|\mu|$. Taking the supremum of the expression on the right over all partitions gives

$$|\mu|(A) \geq \sup\Big\{\sum_{j=1}^{k} |\mu(B_j)| \;\Big|\; (B_1, \dots, B_k) \text{ is a partition of } A\Big\}.$$

We also have, for a Hahn decomposition (P, N) for $(X, \mathscr{A}, \mu)$ and a partition $(A_1, \dots, A_k)$ for A,

$$\mu_+(P \cap A) = |\mu(P \cap A)| = \Big|\sum_{j=1}^{k} \mu(P \cap A_j)\Big| \leq \sum_{j=1}^{k} |\mu(P \cap A_j)|$$

and similarly

$$\mu_-(N \cap A) \leq \sum_{j=1}^{k} |\mu(N \cap A_j)|.$$

Therefore, using the definition of μ_+ and μ_-,

$$|\mu|(A) = \mu_+(P \cap A) + \mu_-(N \cap A) \leq \sum_{j=1}^{k} |\mu(P \cap A_j)| + \sum_{j=1}^{k} |\mu(N \cap A_j)|.$$

Since $(P \cap A_1, \dots, P \cap A_k, N \cap A_1, \dots, N \cap A_k)$ is a partition of A we have

$$|\mu|(A) \leq \sup\Big\{\sum_{j=1}^{k} |\mu(B_j)| \;\Big|\; (B_1, \dots, B_k) \text{ is a partition of } A\Big\},$$

which gives the result. ■

The total variation is, in fact, an interesting quantity; it is a norm on the set of finite signed measures. This point of view will be taken up in Section 3.8.9.

As with measures, we can restrict signed measures to measurable subsets.

2.3.49 Proposition (Restriction of a signed measure) *If* $(\mathrm{X}, \mathscr{A}, \mu)$ *is a signed measure space and if* $\mathrm{A} \in \mathscr{A}$*, then* $(\mathrm{A}, \mathscr{A}_{\mathrm{A}}, \mu|\mathscr{A}_{\mathrm{A}})$ *is a signed measure space. (See Proposition 2.2.6 for the definition of* $\mathscr{A}_{\mathrm{A}}$*.)*

Proof This follows very much along the lines of Proposition 2.3.18. ■

2.3.8 Complex measures

Next we consider measures taking not just general real values, but complex values. As with signed measures, we shall not be able to see interesting examples of complex measures until we talk about integration; see Proposition 2.7.65.

We begin with the definition.

2.3.50 Definition (Complex measure) For a measurable space $(X, \mathscr{A})$, a ***signed measure*** on $\mathscr{A}$ is a map $\mu\colon \mathscr{A} \to \mathbb{C}$ such that

(i) $\mu(\varnothing) = 0$ and

(ii) $\mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} A_j\Big) = \sum_{j=1}^{\infty} \mu(A_j)$ for every family $(A_j)_{j\in\mathbb{Z}_{>0}}$ of pairwise disjoint sets from $\mathscr{A}$ (***countable-additivity***).

A ***complex measure space*** is a triple $(X, \mathscr{A}, \mu)$ where $(X, \mathscr{A})$ is a measurable space and μ is a complex measure on $\mathscr{A}$. •

Note that a complex measure is intrinsically finite since it must take values in $\mathbb{C}$. This makes complex measures a little different and more restrictive in scope than positive or signed measures.

For a complex measure space $(X, \mathscr{A}, \mu)$, we can define finite signed measures $\mathrm{Re}(\mu), \mathrm{Im}(\mu)\colon \mathscr{A} \to \mathbb{R}$ by

$$\mathrm{Re}(\mu)(A) = \mathrm{Re}(\mu(A)), \quad \mathrm{Im}(\mu)(A) = \mathrm{Im}(\mu(A)), \qquad A \in \mathscr{A}.$$

We obviously call $\mathrm{Re}(\mu)$ the ***real part*** of μ and $\mathrm{Im}(\mu)$ the ***imaginary part*** of μ. It is trivial to verify that $\mathrm{Re}(\mu)$ and $\mathrm{Im}(\mu)$ are indeed finite signed measures, and the reader can do this as Exercise 2.3.5. We can then write

$$\mu(A) = \mathrm{Re}(\mu)(A) + \mathrm{i}\,\mathrm{Im}(\mu)(A),$$

or $\mu = \mathrm{Re}(\mu) + \mathrm{i}\,\mathrm{Im}(\mu)$ for short. Since $\mathrm{Re}(\mu)$ and $\mathrm{Im}(\mu)$ are signed measures, they have Jordan decompositions

$$\mathrm{Re}(\mu) = \mathrm{Re}(\mu)_+ - \mathrm{Re}(\mu)_-, \quad \mathrm{Im}(\mu) = \mathrm{Im}(\mu)_+ - \mathrm{Im}(\mu)_-.$$

We can then write

$$\mu = \mathrm{Re}(\mu)_+ - \mathrm{Re}(\mu)_- + \mathrm{i}(\mathrm{Im}(\mu)_+ - \mathrm{Im}(\mu)_-),$$

to which we refer as the ***Jordan decomposition*** of the complex measure μ. It is clear that a finite signed measure can be thought of as a complex measure whose imaginary part is the zero measure.

Now let us turn to the variation of a complex measure.

2.3.51 Definition (Variation and total variation of a complex measure) Let $(X, \mathscr{A}, \mu)$ be a complex measure space. The ***variation*** of μ is the map $|\mu|\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$|\mu|(A) = \sup\Big\{ \sum_{j=1}^{k} |\mu(A_j)| \;\Big|\; (A_1, \dots, A_k) \text{ is a partition of } A \Big\}.$$

The ***total variation*** of μ is $\|\mu\| = |\mu|(X)$. •

Different from the case of a signed measure, it is not immediately clear that the variation is a measure. Thus we verify this.

2.3.52 Proposition (Variation is a positive finite measure) *If* $(X, \mathscr{A}, \mu)$ *is a complex measure space then* $|\mu|$ *is a finite positive measure that satisfies* $|\mu(A)| \leq |\mu|(A)$ *for every* $A \in \mathscr{A}$*. Moreover, if* $\nu\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ *is a positive measure satisfying* $|\mu(A)| \leq \nu(A)$ *for every* $A \in \mathscr{A}$*, then* $|\mu|(A) \leq \nu(A)$ *for every* $A \in \mathscr{A}$*.*

Proof It is evident that $|\mu|(\emptyset) = 0$. To verify countable-additivity of $|\mu|$, we first verify finite-additivity. Let $A_1, A_2 \in \mathscr{A}$ be disjoint and let $(B_1, \dots, B_k)$ be a partition of $A_1 \cup A_2$. We then have

$$\begin{aligned}\sum_{j=1}^{k} |\mu(B_j)| &= \sum_{j=1}^{k} |\mu(A_1 \cap B_j) + \mu(A_2 \cap B_j)| \\ &\leq \sum_{j=1}^{k} (|\mu(A_1 \cap B_j)| + |\mu(A_2 \cap B_j)|) \leq |\mu|(A_1) + |\mu|(A_2),\end{aligned}$$

the last inequality by definition of $|\mu|$. Since

$$|\mu|(A_1 \cup A_2) = \sup\Big\{ \sum_{j=1}^{k} |\mu(B_j)| \;\Big|\; (B_1, \dots, B_k) \text{ is a partition of } A_1 \cup A_2 \Big\},$$

we have

$$|\mu|(A_1 \cup A_2) \leq |\mu|(A_1) + |\mu|(A_2).$$

Now let $(B_{1,1}, \dots, B_{1,k_1})$ be a partition of A_1 and let $(B_{2,1}, \dots, B_{2,k_2})$ be a partition of A_2. Since

$$(B_{1,1}, \dots, B_{1,k_1}) \cup (B_{2,1}, \dots, B_{2,k_2})$$

is a partition of $A_1 \cup A_2$ we have

$$\sum_{j_1=1}^{k_1} |\mu(B_{1,j_1})| + \sum_{j_2=1}^{k_2} |\mu(B_{2,j_2})| \leq |\mu|(A_1 \cup A_2).$$

Since

$$|\mu|(A_1) = \sup\Big\{ \sum_{j_1=1}^{k_1} |\mu(B_{1,j_1})| \;\Big|\; (B_{1,1}, \dots, B_{1,k_1}) \text{ is a partition of } A_1 \Big\},$$

$$|\mu|(A_2) = \sup\Big\{ \sum_{j_2=1}^{k_2} |\mu(B_{2,j_2})| \;\Big|\; (B_{2,1}, \dots, B_{2,k_2}) \text{ is a partition of } A_2 \Big\}$$

we have

$$|\mu|(A_1) + |\mu|(A_2) \le |\mu|(A_1 \cup A_2).$$

Thus $|\mu|(A_1 \cup A_2) = |\mu|(A_1) + |\mu|(A_2)$, whence follows the finite additivity of $|\mu|$.

Now note that for $A \in \mathscr{A}$ we have

$$|\mu(A)| \le |\mathrm{Re}(\mu)(A)| + |\mathrm{Im}(\mu)(A)| \tag{2.6}$$

by . Therefore, for $A \in \mathscr{A}$ and for a finite partition $(A_1, \dots, A_k)$ for A, we have

something from the complex chapter

$$\begin{aligned}\sum_{j=1}^{k} |\mu(A_j)| &\le \sum_{j=1}^{k} (|\mathrm{Re}(\mu)(A_j)| + |\mathrm{Im}(\mu)(A_j)|) \\ &\le \sum_{j=1}^{k} (\mathrm{Re}(\mu)_+(A_j) + \mathrm{Re}(\mu)_-(A_j) + \mathrm{Im}(\mu)_+(A_j) + \mathrm{Im}(\mu)_-(A_j)) \\ &= \sum_{j=1}^{k} \mathrm{Re}(\mu)_+(A) + \mathrm{Re}(\mu)_-(A) + \mathrm{Im}(\mu)_+(A) + \mathrm{Im}(\mu)_-(A).\end{aligned}$$

Taking the supremum of the leftmost expression over all partitions we have

$$|\mu|(A) \le \mathrm{Re}(\mu)_+(A) + \mathrm{Re}(\mu)_-(A) + \mathrm{Im}(\mu)_+(A) + \mathrm{Im}(\mu)_-(A). \tag{2.7}$$

Therefore, if $(A_j)_{j \in \mathbb{Z}_{>0}}$ is a sequence of sets from $\mathscr{A}$ having the properties that $A_j \supseteq A_{j+1}$, $j \in \mathbb{Z}_{>0}$, and that $\cap_{j \in \mathbb{Z}_{>0}} A_j = \emptyset$, we have

$$\lim_{j \to \infty} |\mu|(A_j) \le \lim_{j \to \infty} (\mathrm{Re}(\mu)_+(A_j) + \mathrm{Re}(\mu)_-(A_j) + \mathrm{Im}(\mu)_+(A_j) + \mathrm{Im}(\mu)_-(A_j)) = 0.$$

Countable-additivity of $|\mu|$ now follows from Proposition 2.3.3.

The finiteness of $|\mu|$ follows immediately from (2.7), noting that the four positive measures on the right are finite.

For $A \in \mathscr{A}$ and for a partition $(A_1, \dots, A_k)$ of A we have

$$|\mu(A)| \le \sum_{j=1}^{k} |\mu(A_j)| \le |\mu|(A),$$

which gives the stated property of $|\mu|$.

Now suppose that ν is a positive measure on $\mathscr{A}$ for which $|\mu(A)| \le \nu(A)$ for every $A \in \mathscr{A}$. Therefore, for $A \in \mathscr{A}$ and for a partition $(A_1, \dots, A_k)$ of A, we have

$$\sum_{j=1}^{k} |\mu(A_j)| \le \sum_{j=1}^{k} \nu(A_j) = \nu(A).$$

Taking the supremum of the left-hand side over all partitions then gives $|\mu|(A) \le \nu(A)$, as desired. ■

Note that Proposition 2.3.48 ensures that if a finite signed measure μ is regarded as a complex measure with zero imaginary part, the definition of $|\mu|$ agrees when defined thinking of μ as a signed measure and when defined thinking of μ as a complex measure.

As with signed measures, the total variation for a complex measure is interesting, and will be studied in Section 3.8.9.

2.3.9 Vector measures

The development of vector measures follows rather like that for complex measures in the preceding section. While it is possible to consider measures taking values in general vector spaces, in this section we restrict ourselves to $\mathbb{R}^n$-valued measures.

2.3.53 Definition (Vector measure) For a measurable space $(X, \mathscr{A})$, a ***vector measure*** on $\mathscr{A}$ is a map $\boldsymbol{\mu}\colon \mathscr{A} \to \mathbb{R}^n$ such that

(i) $\boldsymbol{\mu}(\emptyset) = \mathbf{0}$ and

(ii) $\boldsymbol{\mu}\Big(\bigcup_{j\in\mathbb{Z}} A_j\Big) = \sum_{j=1}^{\infty} \boldsymbol{\mu}(A_j)$ for every family $(A_j)_{j\in\mathbb{Z}_{>0}}$ of pairwise disjoint sets from $\mathscr{A}$ (***countable-additivity***).

A ***vector measure space*** is a triple $(X, \mathscr{A}, \boldsymbol{\mu})$ where $(X, \mathscr{A})$ is a measurable space and $\boldsymbol{\mu}$ is a vector measure on $\mathscr{A}$. •

For a vector measure space $(X, \mathscr{A}, \boldsymbol{\mu})$ with $\boldsymbol{\mu}$ taking values in $\mathbb{R}^n$ and for $j \in \{1, \dots, n\}$ we can define a finite signed measure μ_j by $\mu_j(A) = \mathrm{pr}_j(\boldsymbol{\mu}(A))$, where $\mathrm{pr}_j\colon \mathbb{R}^n \to \mathbb{R}$ is the projection onto the jth component. We can write

$$\boldsymbol{\mu}(A) = \mu_j(A)\boldsymbol{e}_1 + \cdots + \mu_n(A)\boldsymbol{e}_n, \qquad A \in \mathscr{A},$$

where $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is the standard basis for $\mathbb{R}^n$. Of course, we can also decompose each of the signed measures $\mu_1, \dots, \mu_n$ into its positive and negative parts, and so arrive at the ***Jordan decomposition*** of $\boldsymbol{\mu}$:

$$\boldsymbol{\mu} = \mu_{1,+} - \mu_{1,-} + \cdots + \mu_{n,+} - \mu_{n,-}.$$

The definition of the variation for vector measures mirrors that for complex measures.

2.3.54 Definition (Variation and total variation of a vector measure) Let $(X, \mathscr{A}, \boldsymbol{\mu})$ be a vector measure space with $\boldsymbol{\mu}$ taking values in $\mathbb{R}^n$. The ***variation*** of $\boldsymbol{\mu}$ is the map $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$\|\boldsymbol{\mu}\|_{\mathbb{R}^n}(A) = \sup\Big\{\sum_{j=1}^{k} \|\boldsymbol{\mu}(A_j)\|_{\mathbb{R}^n} \;\Big|\; (A_1, \dots, A_k) \text{ is a partition of } A\Big\}.$$

The ***total variation*** of $\boldsymbol{\mu}$ is $|||\boldsymbol{\mu}|||_{\mathbb{R}^n} = \|\boldsymbol{\mu}\|_{\mathbb{R}^n}(X)$. •

As with complex measures, one can verify that the variation of a vector measure defines a positive measure.

2.3.55 Proposition (Variation is a positive finite measure) *If $(X, \mathscr{A}, \boldsymbol{\mu})$ is a vector measure space with $\boldsymbol{\mu}$ taking values in $\mathbb{R}^n$, then $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}$ is a finite positive measure that satisfies $\|\boldsymbol{\mu}(A)\|_{\mathbb{R}^n} \leq \|\boldsymbol{\mu}\|_{\mathbb{R}^n}(A)$ for every $A \in \mathscr{A}$. Moreover, if $\nu\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ is a positive measure satisfying $\|\boldsymbol{\mu}(A)\|_{\mathbb{R}^n} \leq \nu(A)$ for every $A \in \mathscr{A}$, then $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}(A) \leq \nu(A)$ for every $A \in \mathscr{A}$.*

Proof The proof is very similar to the corresponding Proposition 2.3.52 for complex measures, so we skip the details of the computations, only pointing out the important differences with the previous proof.

It is still clear that $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}(\emptyset) = 0$. The proof of finite-additivity of $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}$ follows in exactly the same manner as the complex case, but with the complex modulus $|\cdot|$ being replaced by the Euclidean norm $\|\cdot\|_{\mathbb{R}^n}$. In the proof of countable-additivity, the relation (2.6) in the complex case is replaced with the relation

$$\|\boldsymbol{\mu}(A)\|_{\mathbb{R}^n} \leq \sum_{j=1}^{n} |\mu_j(A)|,$$

following Proposition II-1.1.11. This results in the relation (2.7) in the complex case being replaced with the relation

$$\|\boldsymbol{\mu}(A)\|_{\mathbb{R}^n} \leq \sum_{j=1}^{n} (\mu_{j,+}(A) + \mu_{j,-}(A)) = \sum_{j=1}^{n} |\mu_j|(A). \tag{2.8}$$

Then the proof of countable-additivity, using Proposition 2.3.3, follows just as in the complex case, as does finiteness of $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}$.

The property for $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}$ in the proposition is proved just as in the complex case: for $A \in \mathscr{A}$ and for a partition $(A_1, \dots, A_k)$ for A, we have

$$\|\boldsymbol{\mu}(A)\|_{\mathbb{R}^n} \leq \sum_{j=1}^{k} \|\boldsymbol{\mu}(A_j)\|_{\mathbb{R}^n} \leq \|\boldsymbol{\mu}\|_{\mathbb{R}^n}(A).$$

If ν is a positive measure such that $\|\boldsymbol{\mu}(A)\|_{\mathbb{R}^n} \leq \nu(A)$ for every $A \in \mathscr{A}$, we have, just as in the complex case, for a partition $(A_1, \dots, A_k)$ of A:

$$\sum_{j=1}^{k} \|\boldsymbol{\mu}(A_j)\|_{\mathbb{R}^n} \leq \sum_{j=1}^{k} \nu(A_j) = \nu(A),$$

and taking the supremum of the left-hand side over all partitions gives $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}(A) \leq \nu(A)$. ■

2.3.10 Spaces of positive, signed, complex, and vector measures

In this section we briefly consider the various spaces of measures on a measurable space $(X, \mathscr{A})$. Further structural properties of these spaces will be explored in Section 3.8.9.

2.3.56 Definition (Spaces of positive, signed, complex, and vector measures) For a measurable space $(X, \mathscr{A})$, we use the following notation:

(i) $\mathsf{M}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ is the set of positive measures on $\mathscr{A}$;

(ii) $\mathsf{M}((X, \mathscr{A}); \overline{\mathbb{R}})$ is the set of signed measures on $\mathscr{A}$;

(iii) $\mathsf{M}((X, \mathscr{A}); \mathbb{R})$ is the set of finite signed measures on $\mathscr{A}$;

(iv) $\mathsf{M}((X, \mathscr{A}); \mathbb{C})$ is the set of complex measures on $\mathscr{A}$;

(v) $\mathsf{M}((X, \mathscr{A}); \mathbb{R}^n)$ is the set of vector measures on $\mathscr{A}$ taking values in $\mathbb{R}^n$.

For brevity, we may use $\mathsf{M}(X; \overline{\mathbb{R}}_{\geq 0}), \dots, \mathsf{M}(X; \mathbb{R}^n)$ if the σ-algebra $\mathscr{A}$ is understood. •

Let us first explore the algebraic structure of these spaces of measures.

2.3.57 Proposition (The vector space structure of spaces of measures) *For a measurable space* $(X, \mathscr{A})$*, the following statements hold:*

(i) the set $\mathsf{M}((X, \mathscr{A}); \mathbb{R})$ *has a* $\mathbb{R}$*-vector space structure with vector addition and scalar multiplication, respectively, defined by*

$$(\mu_1 + \mu_2)(A) = \mu_1(A) + \mu_2(A), \quad (a\mu)(A) = a(\mu(A))$$

for measures μ, μ_1, *and* μ_2 *in* $\mathsf{M}((X, \mathscr{A}); \mathbb{R})$, *and for* $a \in \mathbb{R}$*;*

(ii) the set $\mathsf{M}((X, \mathscr{A}); \mathbb{R}^n)$ *has a* $\mathbb{R}$*-vector space structure with vector addition and scalar multiplication, respectively, defined by*

$$(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2)(A) = \boldsymbol{\mu}_1(A) + \boldsymbol{\mu}_2(A), \quad (a\boldsymbol{\mu})(A) = a(\boldsymbol{\mu}(A))$$

for measures $\boldsymbol{\mu}, \boldsymbol{\mu}_1, \boldsymbol{\mu}_2 \in \mathsf{M}((X, \mathscr{A}); \mathbb{R}^n)$ *and for* $a \in \mathbb{R}$*;*

(iii) the set $\mathsf{M}((X, \mathscr{A}); \mathbb{C})$ *has a* $\mathbb{C}$*-vector space structure with vector addition and scalar multiplication, respectively, defined by*

$$(\mu_1 + \mu_2)(A) = \mu_1(A) + \mu_2(A), \quad (a\mu)(A) = a(\mu(A))$$

for measures $\mu, \mu_1, \mu_2 \in \mathsf{M}((X, \mathscr{A}); \mathbb{C})$ *and for* $a \in \mathbb{C}$.

Proof To check that $\mu_1 + \mu_2$ and $a\mu$ (or $\boldsymbol{\mu}_1 + \boldsymbol{\mu}_2$ and $a\boldsymbol{\mu}$) have the properties of a measure is straightforward. The remainder of the proof is just a matter of verifying the vector space axioms. The reader who believes this verification might be interesting is welcomed to perform it. ■

2.3.58 Remark (Vector space structures for infinite-valued measures) The reader will have noticed the absence from the above list the vector space structures for the set of positive measures and the set of signed measures. This absence is deserved since, using the natural vector space operations from the statement of the proposition, these sets of measures do not have vector space structures. Let us be sure we understand why in each case.

1. $\mathsf{M}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ is not a $\mathbb{R}$-vector space. The problem here is not just the fact that we allow infinite values for the measures. Even if we restrict to finite positive measures, we do not have a natural vector space structure for which vector addition is given by
$$(\mu_1 + \mu_2)(A) = \mu_1(A) + \mu_2(A).$$
To see this, let μ be a finite positive measure on $\mathscr{A}$. In order for the operation above to be vector space addition, there must exist a finite positive measure $-\mu$ on $\mathscr{A}$ such that $\mu + (-\mu)$ is the zero measure. Thus, for example, we would have to have $\mu(X) + (-\mu(X)) = 0$ and so $-\mu(X) \in \mathbb{R}_{<0}$ if $\mu(X) \in \mathbb{R}_{>0}$. In particular, $-\mu$ cannot be a positive measure.
2. $\mathsf{M}((X, \mathscr{A}); \overline{\mathbb{R}})$ is not a $\mathbb{R}$-vector space. Indeed, if $(X, \mathscr{A}, \mu_1)$ and $(X, \mathscr{A}, \mu_2)$ are signed measure spaces for which μ_1 takes the value ∞ and μ_2 takes the value $-\infty$, then (cf. the proof of Proposition 2.3.2(viii)) it follows that $\mu_1(X) = \infty$ and $\mu_2(X) = -\infty$. Therefore, $(\mu_1 + \mu_2)(X)$ cannot be defined in the natural way. •

Despite the fact that $\mathsf{M}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ is not a $\mathbb{R}$-vector space, we would like for it to have some structure since it comprises the set of positive measures on $\mathscr{A}$, and as such is an interesting object. The following result says that this set is, in fact, a convex cone.

2.3.59 Proposition (The set of positive measures is a convex cone) *Let* $(\mathrm{X}, \mathscr{A})$ *be a measurable space, let* $\mu, \mu_1, \mu_2 \in \mathsf{M}((\mathrm{X}, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$*, and let* $\mathrm{a} \in \mathbb{R}_{\geq 0}$*. Then the maps* $\mathrm{a}\mu, \mu_1 + \mu_2 \colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ *defined by*
$$(\mathrm{a}\mu)(\mathrm{A}) = \mathrm{a}(\mu(\mathrm{A})), \quad (\mu_1 + \mu_2)(\mathrm{A}) = \mu_1(\mathrm{A}) + \mu_2(\mathrm{A})$$
are positive measures on $\mathscr{A}$*. Moreover, for every* $\mu, \mu_1, \mu_2, \mu_3 \in \mathsf{M}((\mathrm{X}, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ *and for every* $\mathrm{a}, \mathrm{a}_1, \mathrm{a}_2 \in \mathbb{R}_{\geq 0}$*, the following statements hold:*

(i) $\mu_1 + \mu_2 = \mu_2 + \mu_1$;
(ii) $\mu_1 + (\mu_2 + \mu_3) = (\mu_1 + \mu_2) + \mu_3$;
(iii) $\mathrm{a}_1(\mathrm{a}_2\mu) = (\mathrm{a}_1\mathrm{a}_2)\mu$;
(iv) $\mathrm{a}(\mu_1 + \mu_2) = \mathrm{a}\mu_1 + \mathrm{a}\mu_2$;
(v) $(\mathrm{a}_1 + \mathrm{a}_2)\mu = \mathrm{a}_1\mu + \mathrm{a}_2\mu$.

Proof As with the proof of Proposition 2.3.57, the verification of the statements are simple matters of checking the properties. ■

2.3.11 Notes

Exercises

2.3.1 Let X be a set and let $\mathscr{A} \subseteq \mathbf{2}^X$ be an algebra. Let $\mu \colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ have the property that $\mu(\varnothing) = 0$. Show that μ is countably-additive if and only if it is finitely-additive and countably-subadditive.

2.3.2 Let X be a countable set, let $\mathscr{A}$ be the algebra $\mathscr{A} = \mathbf{2}^X$, and define $\mu\colon \mathbf{2}^X \to \overline{\mathbb{R}}_{\geq 0}$ by

$$\mu(A) = \begin{cases} 0, & \operatorname{card}(A) < \infty, \\ \infty, & \operatorname{card}(A) = \infty. \end{cases}$$

Answer the following questions.

(a) Show that μ is a σ-finite, finitely-additive measure.

(b) Show that if $(A_j)_{j\in\mathbb{Z}_{>0}}$ if a sequence of subsets from $\mathscr{A}$ for which $A_j \supseteq A_{j+1}$, $j \in \mathbb{Z}_{>0}$, for which $\cap_{j\in\mathbb{Z}_{>0}} A_j = \emptyset$, and for which $\mu(A_k) < \infty$ for some $k \in \mathbb{Z}_{>0}$, it holds that $\lim_{j\to\infty} \mu(A_j) = 0$.

(c) Show that μ is not countably-additive.

2.3.3 Let X be a set and consider the collection $\mathscr{S}$ of subsets of X defined by $\mathscr{S} = \{\emptyset\}$. Define $\mu_0\colon \mathscr{S} \to \overline{\mathbb{R}}_{\geq 0}$ by $\mu_0(\emptyset) = 0$. Compute the outer measure generated by $(\mathscr{S}, \mu_0)$.

2.3.4 For a measure space $(X, \mathscr{A}, \mu)$ do the following.

(a) Show that if $(A_j)_{j\in\mathbb{Z}_{>0}}$ is a countable collection of sets of measure zero then

$$\mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} A_j\Big) = 0.$$

(b) When will there exist an uncountable collection of sets of measure zero whose union has positive measure.

2.3.5 For a complex measure space $(X, \mathscr{A}, \mu)$, show that $\operatorname{Re}(\mu)$ and $\operatorname{Im}(\mu)$ are finite signed measures on $\mathscr{A}$.

2.3.6 Let $(X, \mathscr{A}, \boldsymbol{\mu})$ be a vector measure space with $\boldsymbol{\mu}$ taking values in $\mathbb{R}^n$. Show that for $A \in \mathscr{A}$ we have

$$\sum_{l=1}^{n} |\mu_l|(A) \leq \sqrt{n}\|\boldsymbol{\mu}\|_{\mathbb{R}^n}(A).$$

Hint: *Use Proposition II-1.1.11.*

Section 2.4

Lebesgue measure on ℝ

In this section we specialise the general constructions of the preceding section to a special measure on the set ℝ. Our construction proceeds by first defining an outer measure, then using Theorem 2.3.13 to infer from this a complete measure space. The idea of measure that we use in this section is to be thought of as a generalisation of "length," and we shall point out as we go along that it does indeed share the features of "length" where the latter makes sense. However, the measure we define can be applied to sets for which it is perhaps not clear that a naïve definition of length is possible.

We shall see as we progress through this section that the σ-algebra we define is (1) not the collection of all subsets of ℝ and (2) contains any reasonable set one could desire, and many more that one may not desire.

Do I need to read this section? If you are in the business of learning about the Lebesgue measure, this is where you go about it. •

2.4.1 The Lebesgue outer measure and the Lebesgue measure on ℝ

Our construction of the Lebesgue measure is carried out as per the idea in Section 2.3.2. That is to say, we construct an outer measure on ℝ and take the measurable sets for this outer measure as the σ-algebra for the Lebesgue measure.

We first define the outer measure we use.

2.4.1 Definition (Lebesgue outer measure on ℝ) The ***Lebesgue outer measure*** on ℝ is defined by

$$\lambda^*(S) = \inf\Big\{ \sum_{j=1}^{\infty} |b_j - a_j| \;\Big|\; S \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} (a_j, b_j) \Big\}.$$ •

Thus the Lebesgue outer measure of $S \subseteq \mathbb{R}$ is the smallest sum of the lengths of open intervals that are needed to cover S. Let us define the length of a general interval I by

$$\ell(I) = \begin{cases} b - a, & \mathrm{cl}(I) = [a, b], \\ \infty, & I \text{ is unbounded.} \end{cases}$$

We next verify that the Lebesgue outer measure is indeed an outer measure, and we give its value on intervals.

2.4.2 Theorem (Lebesgue outer measure is an outer measure) *The Lebesgue outer measure is an outer measure on ℝ. Furthermore, if* I *is an interval then* $\lambda^*(\mathrm{I})$ *is the length of* I.

Proof First we show that $\lambda^*(\emptyset)=0$. Indeed, let $(\epsilon_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence converging to zero in $\mathbb{R}_{>0}$ and note that $\emptyset\subseteq(-\epsilon_j,\epsilon_j)$, $j\in\mathbb{Z}_{>0}$. Since $\lim_{j\to\infty}|\epsilon_j+\epsilon_j|=0$, our assertion follows.

Next we show that λ^* is monotonic. This is clear since if $A\subseteq B\subseteq\mathbb{R}$ and if a collection of intervals $((a_j,b_j))_{j\in\mathbb{Z}_{>0}}$ covers B, then the same collection of intervals covers A.

For countable-subadditivity, let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a collection of subsets of $\mathbb{R}$. If $\sum_{j=1}^\infty\lambda^*(A_j)=\infty$ then countable-subadditivity follows trivially in this case, so we may as well suppose that $\sum_{j=1}^\infty\lambda^*(A_j)<\infty$. For $j\in\mathbb{Z}_{>0}$ and $\epsilon\in\mathbb{R}_{>0}$ let $((a_{j,k},b_{j,k}))_{k\in\mathbb{Z}_{>0}}$ be a collection of open sets covering A_j and for which

$$\sum_{k=1}^\infty|b_{j,k}-a_{j,k}|<\lambda^*(A_j)+\frac{\epsilon}{2^j}.$$

By Proposition I-1.7.16, $\mathbb{Z}_{>0}\times\mathbb{Z}_{>0}$ is countable. Therefore we may arrange the intervals $((a_{j,k},b_{j,k}))_{j,k\in\mathbb{Z}_{>0}}$ into a single sequence $((a_n,b_n))_{n\in\mathbb{Z}_{>0}}$ so that

1. $\cup_{j\in\mathbb{Z}_{>0}}A_j\subseteq\cup_{n\in\mathbb{Z}_{>0}}(a_n,b_n)$ and
2. $\displaystyle\sum_{n=1}^\infty|b_n-a_n|<\sum_{n=1}^\infty\Big(\lambda^*(A_n)+\frac{\epsilon}{2^n}\Big)=\sum_{n=1}^\infty\lambda^*(A_n)+\epsilon.$

This shows that

$$\lambda^*\Big(\bigcup_{j\in\mathbb{Z}_{>0}}A_j\Big)\le\sum_{n=1}^\infty\lambda^*(A_n),$$

giving countable-subadditivity.

We finally show that $\lambda^*(I)=\ell(I)$ for any interval I. We first take $I=[a,b]$. We may cover $[a,b]$ by $\{(a-\frac{\epsilon}{4},b+\frac{\epsilon}{4})\}\cup((0,\frac{\epsilon}{2^{j+1}}))_{j\in\mathbb{Z}_{>0}}$. Therefore,

$$\lambda^*([a,b])\le(b+\tfrac{\epsilon}{4}-a+\tfrac{\epsilon}{4})+\sum_{j=1}^\infty\frac{\epsilon}{2^{j+1}}=b-a+\epsilon,$$

where we use Example I-2.4.2–1. Since ϵ can be made arbitrarily small we have $\lambda^*([a,b])\le b-a$. Also, suppose that $((a_j,b_j))_{j\in\mathbb{Z}_{>0}}$ covers $[a,b]$. By Theorem I-2.5.27 there exists $n\in\mathbb{Z}_{>0}$ such that $[a,b]\subseteq\cup_{j=1}^n(a_j,b_j)$. Among the intervals $((a_j,b_j))_{j=1}^n$ we can pick a subset $((a_{j_k},b_{j_k}))_{k=1}^m$ with the properties that $a\in(a_{j_1},b_{j_1})$, $b\in(a_{j_m},b_{j_m})$, and $b_{j_k}\in(a_{j_{k+1}},b_{j_{k+1}})$. (Do this by choosing (a_{j_1},b_{j_1}) such that a is in this interval. Then choose (a_{j_2},b_{j_2}) such that b_{j_1} is in this interval. Since there are only finitely many intervals covering $[a,b]$, this can be continued and will stop by finding an interval containing b.) These intervals then clearly cover $[a,b]$ and also clearly satisfy $\sum_{k=1}^m|b_{j_k}-a_{j_k}|\ge b-a$ since they overlap. Thus we have

$$b-a\le\sum_{k=1}^m|b_{j_k}-a_{j_k}|\le\sum_{j=1}^\infty|b_j-a_j|.$$

Thus $b-a$ is a lower bound for the set

$$\Big\{\sum_{j=1}^\infty|b_j-a_j|\;\Big|\;[a,b]\subseteq\bigcup_{j\in\mathbb{Z}_{>0}}(a_j,b_j)\Big\}.$$

Since $\lambda^*([a,b])$ is the greatest lower bound we have $\lambda^*([a,b]) \geq b-a$. Thus $\lambda^*([a,b]) = b-a$.

Now let I be a bounded interval and denote $\mathrm{cl}(I) = [a,b]$. Since $I \subseteq [a,b]$ we have $\lambda^*(I) \leq b-a$ using monotonicity of λ^*. If $\epsilon \in \mathbb{R}_{>0}$ we may find a closed interval $J \subseteq I$ for which the length of I exceeds that of J by at most ϵ. Since $\lambda^*(J) \leq \lambda^*(I)$ by monotonicity of λ^*, it follows that $\lambda^*(I)$ differs from the length of I by at most ϵ. Thus

$$\lambda^*(I) \geq \lambda^*(J) = b-a-\epsilon.$$

Since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary $\lambda^*(I) \geq b-a$, showing that $\lambda^*(I) = b-a$, as desired.

Finally, if I is unbounded then for any $M \in \mathbb{R}_{>0}$ we may find a closed interval $J \subseteq I$ for which $\lambda^*(J) > M$. Since $\lambda^*(I) \geq \lambda^*(J)$ by monotonicity of λ^*, this means that $\lambda^*(I) = \infty$. ■

Now, having an outer measure on $\mathbb{R}$ one can ask, "Is λ^* a measure?" As we saw in Corollary 2.3.14 this amounts to asking, "Are all subsets of $\mathbb{R}$ λ^*-measurable?" Let us answer this question in the negative.

2.4.3 Example (A set that is not λ^*-measurable) Define an equivalence relation $\sim$ on $\mathbb{R}$ by

$$x \sim y \quad \iff \quad x-y \in \mathbb{Q}.$$

By Proposition I-1.2.9 it follows that $\mathbb{R}$ is the disjoint union of the equivalence classes for this equivalence relation. Moreover, each equivalence class has an element in the interval $(0,1)$ since, for any $x \in \mathbb{R}$, the set

$$\{x+q \mid q \in \mathbb{Q}\}$$

intersects $(0,1)$. By the Axiom of Choice, let $A \subseteq (0,1)$ be defined by asking that A contain exactly one element from each equivalence class. We claim that A is not λ^*-measurable.

Let $\{q_j\}_{j\in\mathbb{Z}_{>0}}$ be an enumeration of the set of rational numbers in $(-1,1)$ and for $j \in \mathbb{Z}_{>0}$ define

$$A_j = \{a+q_j \mid a \in A\}.$$

Note that $\cup_{j\in\mathbb{Z}_{>0}} A_j \subseteq (-1,2)$.

We claim that $A_j \cap A_k \neq \emptyset$ if and only if $j=k$. Indeed, suppose that $A_j \cap A_k = \{x\}$. Then

$$x = a_j + q_j = a_k + q_k, \qquad a_j, a_k \in A.$$

Therefore, $a_j \sim a_k$ and, by construction of A, this implies that $a_j = a_k$. Thus $q_j = q_k$ and so $j = k$.

We also claim that $(0,1) \subseteq \cup_{j\in\mathbb{Z}_{>0}} A_j$. Indeed, if $x \in (0,1)$ then there exists $a \in A$ such that $x \sim a$. Note that $x-a \in \mathbb{Q} \cap (-1,1)$ and so $x = a+q_j$ for some $j \in \mathbb{Z}_{>0}$. Thus $x \in A_j$.

Now suppose that A is λ^*-measurable. As we shall see in Theorem 2.4.23 below, this implies that A_j is λ^*-measurable for each $j \in \mathbb{Z}_{>0}$ and that $\lambda^*(A_j) = \lambda^*(A)$. We consider two cases.

1. $\lambda^*(A) = 0$: In this case, since the sets A_j, $j \in \mathbb{Z}_{>0}$, are disjoint, by properties of the measure we have
$$\mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} A_j\Big) = \sum_{j=1}^{\infty} \mu(A_j) = 0.$$
But this contradicts the fact that $(0,1) \subseteq \cup_{j\in\mathbb{Z}_{>0}} A_j$.
2. $\lambda^*(A) \in \mathbb{R}_{>0}$: In this case we have
$$\mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} A_j\Big) = \sum_{j=1}^{\infty} \mu(A_j) = \infty.$$
But this contradicts the fact that $\cup_{j\in\mathbb{Z}_{>0}} A_j \subseteq (-1,2)$.

The contradiction that arises for both possibilities forces us to conclude that A is not measurable. •

Thus, making the following definition is not a vacuous procedure, and gives a strict subset of $\mathbf{2}^{\mathbb{R}}$ of λ^*-measurable sets.

2.4.4 Definition (Lebesgue measurable subset of $\mathbb{R}$, Lebesgue measure on $\mathbb{R}$) Let λ^* be the Lebesgue outer measure on $\mathbb{R}$ and denote by $\mathscr{L}(\mathbb{R})$ the set of λ^*-measurable subsets of $\mathbb{R}$. The sets in $\mathscr{L}(\mathbb{R})$ are called ***Lebesgue measurable***, or merely ***measurable***, and the complete measure $\lambda\colon \mathscr{L}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ induced by λ^* is called the ***Lebesgue measure*** on $\mathbb{R}$. •

The fairly concrete Example 2.4.3 can actually be sharpened considerably.

2.4.5 Theorem (The wealth of nonmeasurable subsets) *If* $\mathrm{A} \in \mathscr{L}(\mathbb{R})$ *satisfies* $\lambda(\mathrm{A}) \in \mathbb{R}_{>0}$ *then there exists* $\mathrm{S} \subseteq \mathrm{A}$ *that is not in* $\mathscr{L}(\mathbb{R})$*.*

Proof We have $A = \cup_{k\in\mathbb{Z}_{>0}}[-k,k] \cap A$ giving
$$0 < \lambda(A) \leq \sum_{k=1}^{\infty} \lambda([-k,k] \cap A).$$
Thus there exists $N \in \mathbb{Z}_{>0}$ such that $\lambda([-N,N] \cap A > 0$. Therefore, without loss of generality we may suppose that $A \subseteq [-N,N]$ for some $N \in \mathbb{Z}_{>0}$. Let $C \subseteq A$ be a countable subset of A and denote by H_C the subgroup of $(\mathbb{R},+)$ generated by C (see Definition I-4.1.14). Therefore, by Proposition I-4.1.13 it follows that
$$\mathsf{H}_C = \Big\{\sum_{j=1}^{k} n_j x_j \;\Big|\; k \in \mathbb{Z}_{>0},\ n_1,\dots,n_k \in \mathbb{Z}_{>0},\ x_1,\dots,x_k \in C\Big\}.$$
Note that H_C is then a countable union of countable sets and so is countable by Proposition I-1.7.16. Now note that the cosets of H_C form a partition of $\mathbb{R}$. Let $S' \subseteq \mathbb{R}$ be chosen (using the Axiom of Choice) such that S' contains exactly one representative from each coset of H_C. Then define
$$S = \{x \in A \mid x \in (x' + \mathsf{H}_C) \cap A,\ x' \in S'\}.$$

We will show that $S \notin \mathscr{L}(\mathbb{R})$.

For subsets $X, Y \subseteq \mathbb{R}$ let us denote

$$X + Y = \{x + y \mid x \in X,\ y \in Y\}, \quad X - Y = \{x - y \mid x \in X,\ y \in Y\}.$$

Let $B = \mathsf{H}_C \cap (A - A)$. Since $C - C \subseteq B$ we conclude that B is countable. We claim that if $(x_1 + S) \cap (x_2 + S) \neq \emptyset$ for $x_1, x_2 \in B$ then $x_1 = x_2$. Indeed, let $x \in (x_1 + S) \cap (x_2 + S)$ so that

$$x = x_1 + y_1 = x_2 + y_2, \qquad y_1, y_2 \in S.$$

Since $x_1, x_2 \in \mathsf{H}_C$ this implies that $y_2 - y_1 \in \mathsf{H}_C \cap S$ and so $y_1 = y_2$ by construction of S. Thus $(x + S)_{x \in B}$ is a family of pairwise disjoint sets. Moreover, $x + S \subseteq [-3N, 3N]$ for every $x \in B$ since $B, S \subseteq [-N, N]$. We further claim that $A \subseteq B + S$. Indeed, if $x \in A$ then x is in some coset of H_C: $x = y' + \mathsf{H}_C$ for $y' \in S'$. Then, since $x \in A$, there exists $y \in S$ such that $y + \mathsf{H}_C = y' + \mathsf{H}_C$. Thus $x = y + h$ for $y \in S$ and $h \in \mathsf{H}_C$. Therefore, $h = x - z \in A - A$ and so $h \in B$. Thus $x \in B + S$ as desired.

Now suppose that $S \in \mathscr{L}(\mathbb{R})$. There are two possibilities.

1. $\lambda(S) = 0$: In this case we have

$$\lambda(B + S) = \sum_{x \in B} \lambda(x + S) = \sum_{x \in B} \lambda(S) = 0,$$

where we have used the translation-invariance of the Lebesgue measure which we shall prove as Theorem 2.4.23 below. Since $A \subseteq B + S$ and $\lambda(A) \in \mathbb{R}_{>0}$ this is impossible.

2. $\lambda(S) \in \mathbb{R}_{>0}$: In this case we have

$$\lambda(B + S) = \sum_{x \in B} \lambda(x + S) = \sum_{x \in B} \lambda(S) = \infty.$$

Again, this is impossible, this time because $B + S \subseteq [-3N, 3N]$.

The impossibility of the two possible choices if S is Lebesgue measurable forces us to conclude that S is not Lebesgue measurable. ■

The reader might benefit by comparing the proof of the preceding theorem with the more concrete construction of Example 2.4.3.

We will very often wish to consider the Lebesgue measure not on all of $\mathbb{R}$, but on subsets of $\mathbb{R}$. Generally the subsets we consider will be intervals, but let us indicate how to restrict the Lebesgue measure to quite general subsets.

2.4.6 Proposition (Restriction of Lebesgue measure to measurable subsets) *Let $\mathrm{A} \in \mathscr{L}(\mathbb{R})$ and denote*

(i) $\mathscr{L}(\mathrm{A}) = \{\mathrm{B} \cap \mathrm{A} \mid \mathrm{B} \in \mathscr{L}(\mathbb{R})\}$ *and*

(ii) $\lambda_{\mathrm{A}} \colon \mathscr{L}(\mathrm{A}) \to \overline{\mathbb{R}}_{\geq 0}$ *given by* $\lambda_{\mathrm{A}}(\mathrm{B} \cap \mathrm{A}) = \lambda(\mathrm{B} \cap \mathrm{A})$.

Then $(\mathrm{A}, \mathscr{L}(\mathrm{A}), \lambda_{\mathrm{A}})$ *is a complete measure space.*

Proof This follows from Propositions 2.2.6, 2.3.18, and 2.3.23. ■

2.4.2 Borel sets in $\mathbb{R}$ as examples of Lebesgue measurable sets

As we saw in Example 2.4.3, there are subsets of $\mathbb{R}$ that are not Lebesgue measurable. This then forces us to ask, "Which subsets of $\mathbb{R}$ *are* Lebesgue measurable?" To completely answer this question is rather difficult. What we shall do instead is provide a large collection of subsets that (1) are Lebesgue measurable, (2) are somewhat easy to understand (or at least convince ourselves that we understand), and (3) in an appropriate sense approximately characterise the Lebesgue measurable sets.

The sets we describe are given in the following definition. Denote by $\mathscr{O}(\mathbb{R}) \subseteq \mathbf{2}^{\mathbb{R}}$ be the collection of open subsets of $\mathbb{R}$.

2.4.7 Definition (Borel subsets of $\mathbb{R}$) The collection of ***Borel sets*** in $\mathbb{R}$ is the σ-algebra generated by $\mathscr{O}(\mathbb{R})$ (see Proposition 2.2.7). We denote by $\mathscr{B}(\mathbb{R})$ the Borel sets in $\mathbb{R}$. If $A \in \mathscr{B}(\mathbb{R})$ then we denote

$$\mathscr{B}(A) = \{A \cap B \mid B \in \mathscr{B}(\mathbb{R})\}$$ •

It is not so easy to provide a characterisation of the general Borel set, but certainly Borel sets can account for many sorts of sets. Borel sets are a large class of sets, and we shall pretty much only encounter Borel sets except when we are in the process of trying to be pathological. Furthermore, as we shall shortly see, Borel sets are Lebesgue measurable, and so serve to generate a large class of fairly easily described Lebesgue measurable sets.

Let us give some simple classes of Borel sets.

2.4.8 Examples (Borel sets)

1. All open sets are Borel sets, obviously.
2. All closed sets are Borel sets since closed sets are complements of open sets, and since σ-algebras are closed under complementation.
3. All intervals are Borel sets; Exercise 2.4.3.
4. The set $\mathbb{Q}$ of rational numbers is a Borel set; Exercise 2.4.4.
5. A subset $A \subseteq \mathbb{R}$ is a $\mathbf{G}_{\boldsymbol{\delta}}$ if $A = \cap_{j\in\mathbb{Z}_{>0}} O_j$ for a family $(O_j)_{j\in\mathbb{Z}_{>0}}$ of open sets. A G_δ is a Borel set; Exercise 2.4.5.
6. A subset $A \subseteq \mathbb{R}$ is an $\mathbf{F}_{\boldsymbol{\sigma}}$ if $A = \cup_{j\in\mathbb{Z}_{>0}} C_j$ for a family $(C_j)_{j\in\mathbb{Z}_{>0}}$ of closed sets. An F_σ is a Borel set; Exercise 2.4.5.

The practice of calling a set "a G_δ" or "an F_σ" is one of the unfortunate traditions involving poor notation in mathematics, notwithstanding that "G" stands for "Gebiet" ("open" in German), "F" stands for "fermé" ("closed" in French), "δ" stands for "Durchschnitt" ("intersection" in German), and "σ" stands for "Summe" ("sum" in German).

Let us first prove a result which gives interesting and sometimes useful alternative characterisations of Borel sets.

2.4.9 Proposition (Alternative characterisations of Borel sets) *$\mathscr{B}(\mathbb{R})$ is equal to the following collections of sets:*

(i) the σ-algebra $\mathscr{B}_1$ generated by the closed subsets;

(ii) the σ-algebra $\mathscr{B}_2$ generated by intervals of the form $(-\infty, b]$, $b \in \mathbb{R}$;

(iii) the σ-algebra $\mathscr{B}_3$ generated by intervals of the form $(a, b]$, $a, b \in \mathbb{R}$, $a < b$.

Proof First note that $\mathscr{B}(\mathbb{R})$ contains the σ-algebra $\mathscr{B}_1$ generated by all closed sets, since the complements of all open sets, i.e., all closed sets, are contained in $\mathscr{B}(\mathbb{R})$. Note that the sets of the form $(-\infty, b]$ are closed, so the σ-algebra $\mathscr{B}_2$ generated by these subsets is contained in $\mathscr{B}_1$. Since $(a, b] = (-\infty, b] \cap (\mathbb{R} \setminus (-\infty, a])$ it follows that the σ-algebra $\mathscr{B}_3$ generated by subsets of the form $(a, b]$ is contained in $\mathscr{B}_2$. Finally, note that

$$(a, b) = \cup_{n=1}^{\infty} (a, b - \tfrac{1}{n}].$$

Thus, by Proposition I-2.5.6, it follows that every open set is a countable union of sets, each of which is a countable intersection of generators of $\mathscr{B}_3$. Thus $\mathscr{B}(\mathbb{R}) \subseteq \mathscr{B}_3$. Putting this all together gives

$$\mathscr{B}(\mathbb{R}) \subseteq \mathscr{B}_3 \subseteq \mathscr{B}_2 \subseteq \mathscr{B}_1 \subseteq \mathscr{B}(\mathbb{R}).$$

Thus we must conclude that $\mathscr{B}_1 = \mathscr{B}_2 = \mathscr{B}_3 = \mathscr{B}(\mathbb{R})$. ■

We can then assert that all Borel sets are Lebesgue measurable.

2.4.10 Theorem (Borel sets are Lebesgue measurable) $\mathscr{B}(\mathbb{R}) \subseteq \mathscr{L}(\mathbb{R})$.

Proof The theorem will follow from Proposition 2.4.9 if we can show that any set of the form $(-\infty, b]$ is Lebesgue measurable. Let A be such an interval and note that since

$$\lambda^*(S) \leq \lambda^*(S \cap A) + \lambda^*(S \cap (\mathbb{R} \setminus A))$$

we need only show the opposite inequality to show that A is Lebesgue measurable. If $\lambda^*(S) = \infty$ this is clearly true, so we may as well suppose that $\lambda^*(S) < \infty$. Let $((a_j, b_j))_{j \in \mathbb{Z}_{>0}}$ cover S so that

$$\sum_{j=1}^{\infty} |b_j - a_j| < \lambda^*(S) + \epsilon.$$

For $j \in \mathbb{Z}_{>0}$ choose intervals (c_j, d_j) and (e_j, f_j), possibly empty, for which

$$\begin{aligned}
&(a_j, b_j) \cap A \subseteq (c_j, d_j),\\
&(a_j, b_j) \cap (\mathbb{R} \setminus A) \subseteq (e_j, f_j),\\
&(d_j - c_j) + (f_j - e_j) \leq (b_j - a_j) + \frac{\epsilon}{2^j}.
\end{aligned}$$

Note that the intervals $((c_j, d_j))_{j \in \mathbb{Z}_{>0}}$ cover $S \cap A$ and that the intervals $((e_j, f_j))_{j \in \mathbb{Z}_{>0}}$ cover $\mathbb{R} \setminus A$ so that

$$\lambda^*(S \cap A) \leq \sum_{j=1}^{\infty} |d_j - c_j|, \quad \lambda^*(S \cap (\mathbb{R} \setminus A)) \leq \sum_{j=1}^{\infty} |f_j - e_j|.$$

From this we have

$$\lambda^*(S \cap A) + \lambda^*(S \cap (\mathbb{R} \setminus A)) \leq \sum_{j=1}^{\infty} |b_j - a_j| + \epsilon < \lambda^*(S) + 2\epsilon,$$

using the fact that $\sum_{j=1}^{\infty} \frac{1}{2^j} = 1$ by Example I-2.4.2–1. Since ϵ can be taken arbitrarily small, the inequality

$$\lambda^*(S) \geq \lambda^*(S \cap A) + \lambda^*(S \cap (\mathbb{R} \setminus A))$$

follows, and so too does the result. ■

The next result sharpens the preceding assertion considerably.

2.4.11 Theorem (Lebesgue measurable sets are the completion of the Borel sets) *$(\mathbb{R}, \mathscr{L}(\mathbb{R}), \lambda)$ is the completion of $(\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda|\mathscr{B}(\mathbb{R}))$.*

Proof First, given $A \in \mathscr{L}(\mathbb{R})$, we find $L, U \in \mathscr{B}(\mathbb{R})$ such that $L \subseteq A \subseteq U$ and such that $\lambda(U \setminus L) = 0$. We first suppose that $\lambda(A) < \infty$. Using Theorem 2.4.19 below, let $(U_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of open sets containing A and for which $\lambda(U_j) \leq \lambda(A) + \frac{1}{j}$ and let $(L_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of compact subsets of A for which $\lambda(L_j) \geq \lambda(A) - \frac{1}{j}$. If we take $L = \cup_{j\in\mathbb{Z}_{>0}} L_j$ and $U = \cap_{j\in\mathbb{Z}_{>0}} U_j$ then we have $L \subseteq A \subseteq U$. We also have

$$\lambda(U \setminus L) \leq \lambda(U_j \setminus L_j) = \lambda(U_j \setminus A) + \lambda(A \setminus L_j) \leq \tfrac{1}{2j}.$$

Since this holds for every $j \in \mathbb{Z}_{>0}$, this gives our claim when A has finite measure, since L and U are Borel sets. If $\lambda(A) = \infty$ then we can write $A = \cup_{j\in\mathbb{Z}_{>0}} A_j$ with $A_j = (-j, j) \cap A$. For each $j \in \mathbb{Z}_{>0}$ we may find $L_j, U_j \in \mathscr{B}(\mathbb{R})$ such that $L_j \subseteq A_j \subseteq U_j$ and $\lambda(U_j \setminus L_j)$. Taking $L = \cup_{j\in\mathbb{Z}_{>0}} L_j$ and $U = \cup_{j\in\mathbb{Z}_{>0}}$ gives $L \subseteq A \subseteq U$ and $\lambda(U \setminus L) = 0$.

The above shows that $\mathscr{L}(\mathbb{R}) \subseteq \mathscr{B}_\lambda(\mathbb{R})$. Now let $B \in \mathscr{B}_\lambda(\mathbb{R})$ and take Borel sets L and U for which $L \subseteq B \subseteq U$ and $\lambda(U \setminus L) = 0$. Note that $(B \setminus L) \subseteq (U \setminus L)$. Note also that since $U \setminus L \in \mathscr{B}(\mathbb{R})$ we have $U \setminus L \in \mathscr{L}(\mathbb{R})$ and $\lambda(U \setminus L) = 0$. By completeness of the Lebesgue measure this implies that $B \setminus L \in \mathscr{L}(\mathbb{R})$. Since $B = (B \setminus L) \cup L$ this implies that $B \in \mathscr{L}(\mathbb{R})$. ■

The following corollary indicates that Borel sets closely approximate Lebesgue measurable sets.

2.4.12 Corollary (Borel approximations to Lebesgue measurable sets) *If $\mathrm{A} \in \mathscr{L}(\mathbb{R})$ then there exists a Borel set B and a set Z of measure zero such that $\mathrm{A} = \mathrm{B} \cup \mathrm{Z}$.*

Proof This follows directly from Theorem 2.4.11 and the definition of the completion. ■

The preceding result looks like good news in that, except for seemingly irrelevant sets of measure zero, Lebesgue measurable sets agree with Borel sets. The problem is that there are lots of sets of measure zero. The following result indicates that this is reflected by a big difference in the number of Lebesgue measurable sets versus the number of Borel sets.

2.4.13 Proposition (The cardinalities of Borel and Lebesgue measurable sets) *We have* $\mathrm{card}(\mathscr{B}(\mathbb{R})) = \mathrm{card}(\mathbb{R})$ *and* $\mathrm{card}(\mathscr{L}(\mathbb{R})) = \mathrm{card}(\mathbf{2}^{\mathbb{R}})$.

Proof Since $\{x\} \in \mathscr{B}(\mathbb{R})$ for every $x \in \mathbb{R}$ is follows that $\mathrm{card}(\mathscr{B}(\mathbb{R})) \geq \mathrm{card}(\mathbb{R})$. Let $\mathscr{O}_{\mathbb{Q}}$ be the collection of open intervals with rational (or infinite) endpoints. The set $\mathscr{O}_{\mathbb{Q}}$ is a countable union of countable sets and so is countable by Proposition I-1.7.16. Since every open set is a countable union of sets from $\mathscr{O}_{\mathbb{Q}}$ (cf. Proposition I-2.5.6 and see Proposition II-1.2.21) it that if we take $\mathscr{S} = \mathscr{O}_{\mathbb{Q}}$ then, in the notation of Theorem 2.2.14, $\mathscr{S}_1$ includes the collection of open sets. Then it follows that $\mathscr{B}(\mathbb{R})$ is the σ-algebra generated by the countable family $\mathscr{O}_{\mathbb{Q}}$ of subsets of $\mathbb{R}$. By Theorem 2.2.14 it follows that $\mathrm{card}(\mathscr{B}(\mathbb{R})) \leq \aleph_0^{\aleph_0} = \mathrm{card}(\mathbb{R})$, using the computation

$$2^{\aleph_0} \leq \aleph_0^{\aleph_0} \leq (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0},$$

which holds since $2 \leq \aleph_0 \leq 2^{\aleph_0}$ by Example I-1.7.14–3 and Exercise I-1.7.4.

To show that $\mathrm{card}(\mathscr{L}(\mathbb{R})) = \mathrm{card}(\mathbf{2}^{\mathbb{R}})$ first note that $\mathrm{card}(\mathscr{L}(\mathbb{R})) \leq \mathrm{card}(\mathbf{2}^{\mathbb{R}})$. For the opposite inequality, recall from Example I-2.5.39 that the middle-thirds Cantor set $C \subseteq [0,1]$ has the properties (1) $\lambda(C) = 0$ and (2) $\mathrm{card}(C) = \mathrm{card}([0,1]) = \mathrm{card}(\mathbb{R})$. Since the Lebesgue measure is complete, every subset of C is Lebesgue measurable and has Lebesgue measure zero. This shows that $\mathrm{card}(\mathbf{2}^C) = \mathrm{card}(\mathbf{2}^{\mathbb{R}}) \leq \mathrm{card}(\mathscr{L}(\mathbb{R}))$. ■

While the preceding result is interesting in that it tells us that there are many more Lebesgue measurable sets than Borel sets, Corollary 2.4.12 notwithstanding, it does not tell us what a non-Borel Lebesgue measurable set might look like. The following is a concrete example of such a set. Our construction uses some facts about measurable functions that we will not introduce until Section 2.6.

2.4.14 Example (A non-Borel Lebesgue measurable set) Recall from Example I-3.2.27 the construction of the Cantor function $f_C\colon [0,1] \to [0,1]$, and recall that f_C is continuous, monotonically increasing, and satisfies $f_C(0) = 0$ and $f_C(1) = 1$. Thus, by the Intermediate Value Theorem, for each $y \in [0,1]$ there exists $x \in [0,1]$ such that $f_C(x) = y$. We use this fact to define $g_C\colon [0,1] \to [0,1]$ by

$$g_C(y) = \inf\{x \in [0,1] \mid f_C(x) = y\}.$$

Let us prove some facts about g_C.

1 Lemma *We have* $\mathrm{f_C \circ g_C(y) = y}$ *and so* $\mathrm{g_C}$ *is injective.*

Proof Let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $[0,1]$ for which $\lim_{j\to\infty} f_C(x_j) = y$. This sequence contains a convergent subsequence $(x_{j_k})_{k\in\mathbb{Z}_{>0}}$ by the Bolzano–Weierstrass Theorem; let $x = \lim_{k\to\infty} x_{j_k}$. Then, by continuity of f_C, $y = f_C(x)$. We also have $g_C(y) = x$ by definition, and so this gives $f_C \circ g_C(y) = y$, as desired. Injectivity of g_C follows from Proposition I-1.3.9. ▼

2 Lemma *The function g_C is monotonically increasing.*

Proof Let $y_1, y_2 \in [0,1]$ satisfy $y_1 < y_2$ and suppose that $g_C(y_1) > g_C(y_2)$. Then $f_C \circ g_C(y_1) \geq f_C \circ g_C(y_2)$ since f_C is monotonically increasing. From the previous lemma this implies that $y_1 \geq y_2$ which is a contradiction. Thus we must have $g_C(y_1) \leq g_C(y_2)$. ▼

3 Lemma $\operatorname{image}(g_C) \subseteq C$.

Proof For $y \in \mathbb{R}$ the set

$$\{x \in [0,1] \mid f_C(x) = y\}$$

is an interval, possibly with empty interior, on which f_C is constant. The endpoints of the interval are points in C. In particular, $g_C(y) \in C$. ▼

Now let $A \subseteq [0,1]$ be the non-Lebesgue measurable subset of Example 2.4.3 and take $B = g_C(A)$. Then $B \subseteq C$ and so is a subset of a set of measure zero by Example I-2.5.39. Since the Lebesgue measure is complete it follows that B is Lebesgue measurable. Because g_C is monotone, we claim that the preimage of a Borel set under g_C is a Borel set. Since intervals of the form $(-\infty, b]$ generate the σ-algebra of Borel sets, it suffices by Proposition I-1.3.5 to show that $g_C^{-1}((-\infty, b])$ is a Borel set. However, by monotonicity, we can directly verify that $g_C^{-1}((-\infty, b])$ is either empty or is the intersection of some unbounded interval with $[0,1]$, i.e., a Borel set. Therefore, were B to be a Borel set, $g_C^{-1}(B)$ would also be a Borel set. However, injectivity of g_C gives $g_C^{-1}(B) = A$, and A is not Lebesgue measurable, and so certainly not Borel. Thus B is not a Borel set. •

When we come to talk about functions defined on measurable spaces in Section 2.6 we will consider functions taking values in $\overline{\mathbb{R}}$. It will then be occasionally useful to have a notion of a Borel subset of $\overline{\mathbb{R}}$. Let us, therefore, define what these subsets are.

2.4.15 Definition (Borel subsets of $\overline{\mathbb{R}}$) The collection of ***Borel sets*** in $\overline{\mathbb{R}}$ is the σ-algebra generated by the subsets of $\overline{\mathbb{R}}$ having the following form:

$$U, \quad U \cup [-\infty, b), \quad U \cup (a, \infty], \quad U \cup [-\infty, b) \cup (a, \infty], \qquad U \in \mathscr{O}(\mathbb{R}),\ a, b \in \mathbb{R}.$$

We denote by $\mathscr{B}(\overline{\mathbb{R}})$ the Borel sets in $\overline{\mathbb{R}}$. •

The idea of the preceding definition is that $\mathscr{B}(\overline{\mathbb{R}})$ is the σ-algebra generated by open subsets of $\overline{\mathbb{R}}$, where open subsets of $\overline{\mathbb{R}}$ are those used in the definition. That these open subsets are indeed the open subsets for a topology on $\overline{\mathbb{R}}$ is argued in Example 1.2.3–4.

The following characterisation of $\mathscr{B}(\overline{\mathbb{R}})$ is useful.

2.4.16 Proposition (Characterisation of $\mathscr{B}(\overline{\mathbb{R}})$) *The σ-algebra $\mathscr{B}(\overline{\mathbb{R}})$ is generated by $\mathscr{B}(\mathbb{R}) \cup \{-\infty\} \cup \{\infty\}$.*

Proof Clearly $\mathscr{B}(\mathbb{R}) \subseteq \mathscr{B}(\overline{\mathbb{R}})$. Since

$$\{\infty\} = \cap_{k \in \mathbb{Z}}(k, \infty], \quad \{-\infty\} = \cap_{k \in \mathbb{Z}}[-\infty, -k),$$

and since $(k, \infty], [-\infty, -k) \in \mathscr{B}(\overline{\mathbb{R}})$ for each $k \in \mathbb{Z}_{>0}$, it follows that $\{-\infty\}, \{\infty\} \in \mathscr{B}(\overline{\mathbb{R}})$. Therefore, the σ-algebra generated by $\mathscr{B}(\mathbb{R}) \cup \{-\infty\} \cup \{\infty\}$ is contained in $\mathscr{B}(\overline{\mathbb{R}})$.

Next we note that $U \in \mathscr{B}(\mathbb{R})$ if $U \in \mathscr{O}(\mathbb{R})$. Also, for $b \in \mathbb{R}$,

$$U \cup [-\infty, b) = U \cup \{-\infty\} \cup (-\infty, b)$$

and so $U \cup [-\infty, b)$ is a union of sets from $\mathscr{B}(\mathbb{R}) \cup \{-\infty\} \cup \{\infty\}$. In similar fashion sets of the form

$$U \cup (a, \infty], \quad U \cup [-\infty, b) \cup (a, \infty]$$

for $a, b \in \mathbb{R}$ are unions of sets from $\mathscr{B}(\mathbb{R}) \cup \{-\infty\} \cup \{\infty\}$. This implies that the generators for the σ-algebra $\mathscr{B}(\overline{\mathbb{R}})$ are contained in the σ-algebra generated by $\mathscr{B}(\mathbb{R}) \cup \{-\infty\} \cup \{\infty\}$. Thus $\mathscr{B}(\overline{\mathbb{R}})$ is contained in the σ-algebra generated by $\mathscr{B}(\mathbb{R}) \cup \{-\infty\} \cup \{\infty\}$. ■

2.4.3 Further properties of the Lebesgue measure on ℝ

In this section we give some additional properties of the Lebesgue measure that (1) illustrate a sort of friendliness of this measure and (2) justify its being in some way natural.

Let us illustrate first an important property of the Lebesgue measure. Let us do this by giving a general definition that creates a little context for this property of Lebesgue measure.

2.4.17 Definition (Regular measure on ℝ) Let $\mathscr{A}$ be a σ-algebra on $\mathbb{R}$ that contains the Borel σ-algebra $\mathscr{B}(\mathbb{R})$. A measure $\mu\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ is ***regular*** if

(i) $\mu(K) < \infty$ for each compact subset $K \subseteq \mathbb{R}$,

(ii) if $A \in \mathscr{A}$ then $\mu(A) = \inf\{\mu(U) \mid U \text{ open and } A \subseteq U\}$, and

(iii) if $U \subseteq \mathbb{R}$ is open then $\mu(U) = \sup\{\mu(K) \mid K \text{ open and } K \subseteq U\}$. •

Before we prove that the Lebesgue measure is regular, let us give some examples that show that irregular measures are possible.

2.4.18 Examples (Regular and irregular measures)

1. For $x \in \mathbb{R}$ the point mass measure $\delta_x\colon \mathscr{B}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$\delta(B) = \begin{cases} 1, & x \in B, \\ 0, & x \notin B \end{cases}$$

is regular, as may be readily verified; see Exercise 2.4.6.

2. One can check that the counting measure $\mu\colon \mathscr{B}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$\mu(B) = \begin{cases} \operatorname{card}(B), & \operatorname{card}(B) < \infty, \\ \infty, & \text{otherwise} \end{cases}$$

is not regular; see Exercise 2.4.7. •

We begin with a theorem that characterises the Lebesgue measure of measurable sets.

2.4.19 Theorem (Regularity of the Lebesgue measure) *The Lebesgue measure $\lambda\colon \mathscr{L}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ is σ-finite and regular. Moreover, for $A \in \mathscr{L}(\mathbb{R})$ we have $\lambda(A) = \sup\{\lambda(K) \mid K \text{ compact and } K \subseteq A\}$.*

Proof To see that λ is σ-finite note that $\mathbb{R} = \cup_{k\in\mathbb{Z}_{>0}}[-k,k]$ with $\lambda([-k,k]) < \infty$.

Next we show that if $A \in \mathscr{L}(\mathbb{R})$ then

$$\lambda(A) = \inf\{\lambda(U) \mid U \text{ open and } A \subseteq U\}.$$

Assume that $\lambda(A) < \infty$ since the result is obvious otherwise. Let $\epsilon \in \mathbb{R}_{>0}$ and let $((a_j,b_j))_{j\in\mathbb{Z}_{>0}}$ be a sequence of open intervals for which $A \subseteq \cup_{j\in\mathbb{Z}_{>0}}(a_j,b_j)$ and for which

$$\sum_{j=1}^{\infty}|b_j - a_j| = \lambda(A) + \epsilon.$$

Now let $U = \cup_{j\in\mathbb{Z}_{>0}}(a_j,b_j)$, noting that U is open and that $A \subseteq U$. By Proposition 2.3.10(iii) and the fact that the measure of an interval is its length we have

$$\lambda(U) \leq \sum_{j=1}^{\infty}|b_j - a_j| = \lambda(A) + \epsilon.$$

Since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary this shows that

$$\lambda(A) \geq \inf\{\lambda(U) \mid U \text{ open and } A \subseteq U\}.$$

Since the other inequality is obvious by the basic properties of a measure, this part of the result follows.

Note that to show that λ is regular it suffices to prove the final assertion of the theorem since open sets are Lebesgue measurable; thus we prove the final assertion of the theorem. First suppose that $A \in \mathscr{L}(\mathbb{R})$ is bounded. Then let $\tilde{K}$ be a compact set containing A. For $\epsilon \in \mathbb{R}_{>0}$ choose U open and containing $\tilde{K} \setminus A$ and for which $\lambda(U) \leq \lambda(\tilde{K} \setminus A) + \epsilon$, this being possible from by the first part of the proof. Note that $K = \tilde{K} \setminus U$ is then a compact set contained in A and that the basic properties of measure then give

$$\lambda(U) \leq \lambda(\tilde{K} \setminus A) + \epsilon \text{ and } \lambda(\tilde{K}) \leq \lambda(K) - \lambda(A) \quad \implies \quad \lambda(K) > \lambda(A) - \epsilon.$$

Since ϵ can be made as small as desired, this gives the second part of the proposition when A is bounded. Define

$$A_j = (-j, j) \cap A,$$

and note that $(A_j)_{j\in\mathbb{Z}_{>0}}$ is an increasing sequence of sets and that $A = \cup_{j\in\mathbb{Z}_{>0}} A_j$. Therefore, by Proposition 2.3.10(iv), $\lambda(A) = \lim_{j\to\infty} \lambda(A_j)$. Then for any $M < \lambda(A)$ there exists $N \in \mathbb{Z}_{>0}$ such that $\lambda(A_N) > M$. We may now find a compact K such that $\lambda(K) > M$ by the fact that we have proved our assertion for bounded sets (as is A_N). Note that $K \subseteq A$ and that $M < \lambda(A)$ is arbitrary, and so the result follows. ■

This result has the following obvious corollary.

2.4.20 Corollary (Approximation of Lebesgue measurable sets by open and compact sets) *If* $\mathrm{A} \in \mathscr{L}(\mathbb{R})$ *satisfies* $\lambda(\mathrm{A}) < \infty$ *and if* $\epsilon \in \mathbb{R}_{>0}$ *then there exists an open set* $\mathrm{U} \subseteq \mathbb{R}$ *and a compact set* $\mathrm{K} \subseteq \mathbb{R}$ *such that*

$$\lambda(\mathrm{U} \setminus \mathrm{A}) < \epsilon, \quad \lambda(\mathrm{A} \setminus \mathrm{K}) < \epsilon.$$

Let us next show that the Lebesgue measure is, in some way, natural. We do this by considering a particular property of the Lebesgue measure, namely that it is "translation-invariant." In order to define what it means for a measure to be translation-invariant, we first need to say what it means for a σ-algebra to be translation-invariant.

2.4.21 Definition (Translation-invariant σ-algebra and measure on $\mathbb{R}$) A σ-algebra $\mathscr{A} \subseteq \mathbf{2}^{\mathbb{R}}$ is ***translation-invariant*** if, for every $A \in \mathscr{A}$ and every $x \in \mathbb{R}$,

$$x + A \triangleq \{x + y \mid y \in A\} \in \mathscr{A}.$$

A ***translation-invariant*** measure on a translation-invariant σ-algebra $\mathscr{A}$ is a map $\mu\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ for which $\mu(x + A) = \mu(A)$ for every $A \in \mathscr{A}$ and $x \in \mathbb{R}$. •

The two σ-algebras we are considering in this section are translation-invariant.

2.4.22 Proposition (Translation-invariance of Borel and Lebesgue measurable sets) *Both* $\mathscr{B}(\mathbb{R})$ *and* $\mathscr{L}(\mathbb{R})$ *are translation-invariant.*

Proof Let us denote

$$\mathscr{B}'(\mathbb{R}) = \{B \mid x + B \in \mathscr{B}(\mathbb{R}) \text{ for every } x \in \mathbb{R}\}.$$

We claim that $\mathscr{B}'(\mathbb{R})$ is a σ-algebra containing the open subsets of $\mathbb{R}$. First of all, if $U \subseteq \mathbb{R}$ is open then $x + U$ is open for every $x \in \mathbb{R}$ (why?) and so $U \in \mathscr{B}'(\mathbb{R})$. To see that $\mathscr{B}'(\mathbb{R})$ is a σ-algebra, first note that $\mathbb{R} = x + \mathbb{R}$ for every $x \in \mathbb{R}$ and so $\mathbb{R} \in \mathscr{B}'(\mathbb{R})$. Next, let $B \in \mathscr{B}'(\mathbb{R})$ and let $x \in \mathbb{R}$. Then

$$\begin{aligned} x + (\mathbb{R} \setminus B) &= \{x + z \mid z \notin B\} = \{y \mid y - x \notin B\} = \{y \mid y \neq x + z,\ z \in B\} \\ &= \{y \mid y \notin (x + B)\} = \mathbb{R} \setminus (x + B) \in \mathscr{B}(\mathbb{R}). \end{aligned}$$

Thus $x + (\mathbb{R} \setminus B) \in \mathscr{B}(\mathbb{R})$ for every $x \in \mathbb{R}$ and so $\mathbb{R} \setminus B \in \mathscr{B}'(\mathbb{R})$. Finally, let $(B_j)_{j\mathbb{Z}_{>0}}$ be a countable collection of subsets from $\mathscr{B}'(\mathbb{R})$. Then, for $x \in \mathbb{R}$ we have

$$x + \cup_{j\in\mathbb{Z}_{>0}} B_j = \cup_{j\in\mathbb{Z}_{>0}} (x + B_j) \in \mathscr{B}(\mathbb{R})$$

and so $\cup_{j\in\mathbb{Z}_{>0}} B_j \in \mathscr{B}'(\mathbb{R})$. Thus $\mathscr{B}'(\mathbb{R})$ is indeed a σ-algebra containing the open sets and so we conclude that $\mathscr{B}(\mathbb{R}) \subseteq \mathscr{B}'(\mathbb{R})$ since $\mathscr{B}(\mathbb{R})$ is the σ-algebra generated by the open sets. This shows that $\mathscr{B}(\mathbb{R})$ is translation-invariant.

Next let us show that $\mathscr{L}(\mathbb{R})$ is translation-invariant. To do this we first show that if $S \subseteq \mathbb{R}$ and if $x \in \mathbb{R}$ then $\lambda^*(x + S) = \lambda^*(S)$. Indeed,

$$\begin{aligned}\lambda^*(x+S) &= \inf\Big\{\sum_{j=1}^{\infty} |b_j - a_j| \;\Big|\; x + S \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} (a_j, b_j)\Big\} \\ &= \inf\Big\{\sum_{j=1}^{\infty} |b_j - a_j| \;\Big|\; x + S \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} (x + a_j, x + b_j)\Big\} \\ &= \inf\Big\{\sum_{j=1}^{\infty} |b_j - a_j| \;\Big|\; S \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} (a_j, b_j)\Big\} = \lambda^*(S).\end{aligned}$$

Now let $A \in \mathscr{L}(\mathbb{R})$ so that, for every subset $S \subseteq \mathbb{R}$,

$$\lambda^*(S) = \lambda^*(S \cap A) + \lambda^*(S \cap (\mathbb{R} \setminus A)).$$

Then, for $x \in \mathbb{R}$ and $S \subseteq \mathbb{R}$,

$$\lambda^*(S \cap (x + A)) = \lambda^*((x + (-x + S)) \cap (x + A)) = \lambda^*((-x + S) \cap A)$$

and, similarly,

$$\lambda^*(S \cap (\mathbb{R} \setminus (x + A))) = \lambda^*((x + (-x + S)) \cap (x + \mathbb{R} \setminus A)) = \lambda^*((-x + S) \cap (\mathbb{R} \setminus A)).$$

Since $\lambda^*(-x + S) = \lambda^*(S)$ this immediately gives

$$\lambda^*(S) = \lambda^*(S \cap (x + A)) + \lambda^*(S \cap (\mathbb{R} \setminus (x + A))),$$

showing that $x + A \in \mathscr{L}(\mathbb{R})$. ■

Now that the σ-algebras are known to be translation-invariant, we can make the following characterisation of the Lebesgue measure.

2.4.23 Theorem (Translation invariance of the Lebesgue measure) *If $\mu\colon \mathscr{B}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ is a nonzero translation-invariant measure for which $\mu(B) < \infty$ for every bounded $B \in \mathscr{B}(\mathbb{R})$, then there exists $c \in \mathbb{R}_{>0}$ such that $\mu(B) = c\lambda(B)$ for every $B \in \mathscr{B}(\mathbb{R})$. Moreover, the Lebesgue measure $\lambda\colon \mathscr{L}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ is translation-invariant.*

Proof That λ is translation-invariant follows from the proof of Proposition 2.4.22 where we showed that $\lambda^*(x + S) = \lambda^*(S)$ for every $S \subseteq \mathbb{R}$ and $x \in \mathbb{R}$. To show that λ is, up to a positive scalar, the only translation-invariant measure we first prove two lemmata.

1 Lemma *If* $U \subseteq \mathbb{R}$ *is a nonempty open set, then there exists a countable collection of disjoint half-open intervals* $(I_j)_{j\in\mathbb{Z}_{>0}}$ *such that* $U = \cup_{j\in\mathbb{Z}_{>0}} I_j$.

Proof For $k \in \mathbb{Z}_{\geq 0}$ define

$$\mathscr{C}_k = \{[j2^{-k}, (j+1)2^{-k}) \mid j \in \mathbb{Z}\}.$$

Note that, for each $k \in \mathbb{Z}_{\geq 0}$, the sets from $\mathscr{C}_k$ form a countable partition of $\mathbb{R}$. Also note that for $k < l$, every interval in $\mathscr{C}_l$ is also an interval in $\mathscr{C}_k$. Now let $U \subseteq \mathbb{R}$ be open. Let $\mathscr{D}_0 = \emptyset$. Let

$$\begin{aligned}
\mathscr{D}_1 &= \{I \in \mathscr{C}_1 \mid I \subseteq U\},\\
\mathscr{D}_2 &= \{I \in \mathscr{C}_2 \mid I \subseteq U,\ I \notin \mathscr{D}_1\},\\
&\ \vdots\\
\mathscr{D}_k &= \{I \in \mathscr{C}_k \mid I \subseteq U,\ I \notin \mathscr{D}_1 \cup \cdots \cup \mathscr{D}_{k-1}\}\\
&\ \vdots
\end{aligned}$$

The result will follow if we can show that each point $x \in U$ is contained in some $\mathscr{D}_k$, $k \in \mathbb{Z}_{>0}$. However, this follows since U is open, and so, for each $x \in U$, one can find a smallest $k \in \mathbb{Z}_{\geq 0}$ with the property that there exists $I \in \mathscr{C}_k$ with $x \in I$ and $I \subseteq U$. ▼

2 Lemma *The Lebesgue measure is the unique measure on* $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ *for which the measure of an interval is its length.*

Proof From Theorem 2.4.2 we know that $\lambda(I) = \ell(I)$ for every interval I. Now suppose that $\mu\colon \mathscr{B}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ is a measure with the property that $\mu(I) = \ell(I)$ for every interval I.

First let $U \subseteq \mathbb{R}$ be open. By Lemma 1 we can write $U = \cup_{j\in\mathbb{Z}_{>0}} I_j$ for a countable family $(I_j)_{j\in\mathbb{Z}_{>0}}$ of disjoint intervals. Therefore, since μ is a measure,

$$\mu(U) = \mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} I_j\Big) = \sum_{j=1}^{\infty} \mu(I_j) = \sum_{j=1}^{\infty} \lambda(I_j) = \lambda(U).$$

Now let B be a bounded Borel set and let U be an open set for which $B \subseteq U$. Then

$$\mu(B) \leq \mu(U) = \lambda(U).$$

Therefore,

$$\mu(B) \leq \inf\{\lambda(U) \mid U \text{ open and } B \subseteq U\} = \lambda(B)$$

by regularity of λ. Therefore, if U is a bounded open set containing B we have

$$\mu(U) = \mu(B) + \mu(U \setminus B) \leq \lambda(B) + \lambda(U \setminus B) = \lambda(U).$$

Since $\mu(U) = \lambda(U)$ it follows that $\mu(B) = \lambda(B)$ and $\mu(U \setminus B) = \lambda(U \setminus B)$.

Finally let B be an unbounded Borel set. We can then write $B = \cup_{j\in\mathbb{Z}} B_j$ where $(B_j)_{j\in\mathbb{Z}_{>0}}$ is the countable family of disjoint Borel sets $B_j = B \cap [j, j+1)$, $j \in \mathbb{Z}$. Then

$$\mu(B) = \sum_{j\in\mathbb{Z}} \mu(B_j) = \sum_{j\in\mathbb{Z}} \lambda(B_j) = \lambda(B),$$

as desired. ▼

To proceed with the proof, let $\mu\colon \mathscr{B}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ be a translation-invariant measure and let $c = \mu([0,1))$. By assumption $c \in \mathbb{R}_{>0}$ since, were $c = 0$,

$$\mu(\mathbb{R}) = \sum_{j=1}^{\infty} \mu([j, j+1)) = 0$$

by translation-invariance of μ. Now let $\mu'\colon \mathscr{B}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ be the measure defined by $\mu'(B) = c^{-1}\mu(B)$. Now, for $k \in \mathbb{Z}_{\geq 0}$ let

$$\mathscr{C}_k = \{[j2^{-k}, (j+1)2^{-k}) \mid j \in \mathbb{Z}\}$$

as in the proof of Lemma 1. Let $I \in \mathscr{C}_k$. We can write $[0,1)$ as a disjoint union of 2^k intervals of the form $x_j + I$. Therefore, by translation-invariance of μ',

$$\mu'([0,1)) = 2^k \mu'(I), \quad \lambda([0,1)) = 2^k \lambda(I).$$

Since $\mu'([0,1)) = \lambda([0,1))$ it follows that $\mu'(I) = \lambda(I)$. Since every interval is a disjoint union of intervals from the sets $\mathscr{C}_k$, $k \in \mathbb{Z}_{\geq 0}$, by Lemma 1 it follows that $\mu'(I) = \lambda(I)$ for every interval I. Thus $\mu' = \lambda$ by Lemma 2 above and so $\mu = c\lambda$, as desired. ■

It is then natural question to ask, "Are there larger σ-algebras than $\mathscr{B}(\mathbb{R})$ which admit a translation-invariant measure?" Obviously one such is the collection $\mathscr{L}(\mathbb{R})$ of Lebesgue measurable sets. But are there larger ones? The following result gives a partial answer, and indicates that the "best possible" construction is impossible.

2.4.24 Theorem (There are no translation-invariant, length-preserving measures on all subsets of $\mathbb{R}$) *There exists no measure space $(\mathbb{R}, \mathscr{A}, \mu)$ having the joint properties that*

(i) $\mathscr{A} = \mathbf{2}^{\mathbb{R}}$,

(ii) $\mu((0,1)) = 1$, *and*

(iii) μ *is translation-invariant.*

Proof Were such a measure to exist, then the non-Lebesgue measurable set $A \subseteq (0,1)$ of Example 2.4.3 would be measurable. But during the course of Example 2.4.3 we showed that $(0,1)$ is a countable disjoint union of translates of A. The dichotomy illustrated in Example 2.4.3 then applies. That is, if $\mu(A) = 0$ then we get $\mu((0,1)) = 0$ and if $\mu(A) \in \mathbb{R}_{>0}$ then $\mu((0,1)) = \infty$, both of which conclusions are false. ■

It is now possible to provide a summary of the "reasonableness" of the Lebesgue measure by providing a natural line of reasoning, the natural terminus of which is the Lebesgue measure. In Figure 2.3 we show a "flow chart" for how one might justify the Lebesgue measure as being the process of some rational line of thought. Note that we are not saying that this actually described the historical development of the Lebesgue measure, but just that, after the fact, it indicates that the Lebesgue measure is not a strange thing to arrive at. It is rare that scientific discovery actually proceeds along the lines that make it most understandable in hindsight.

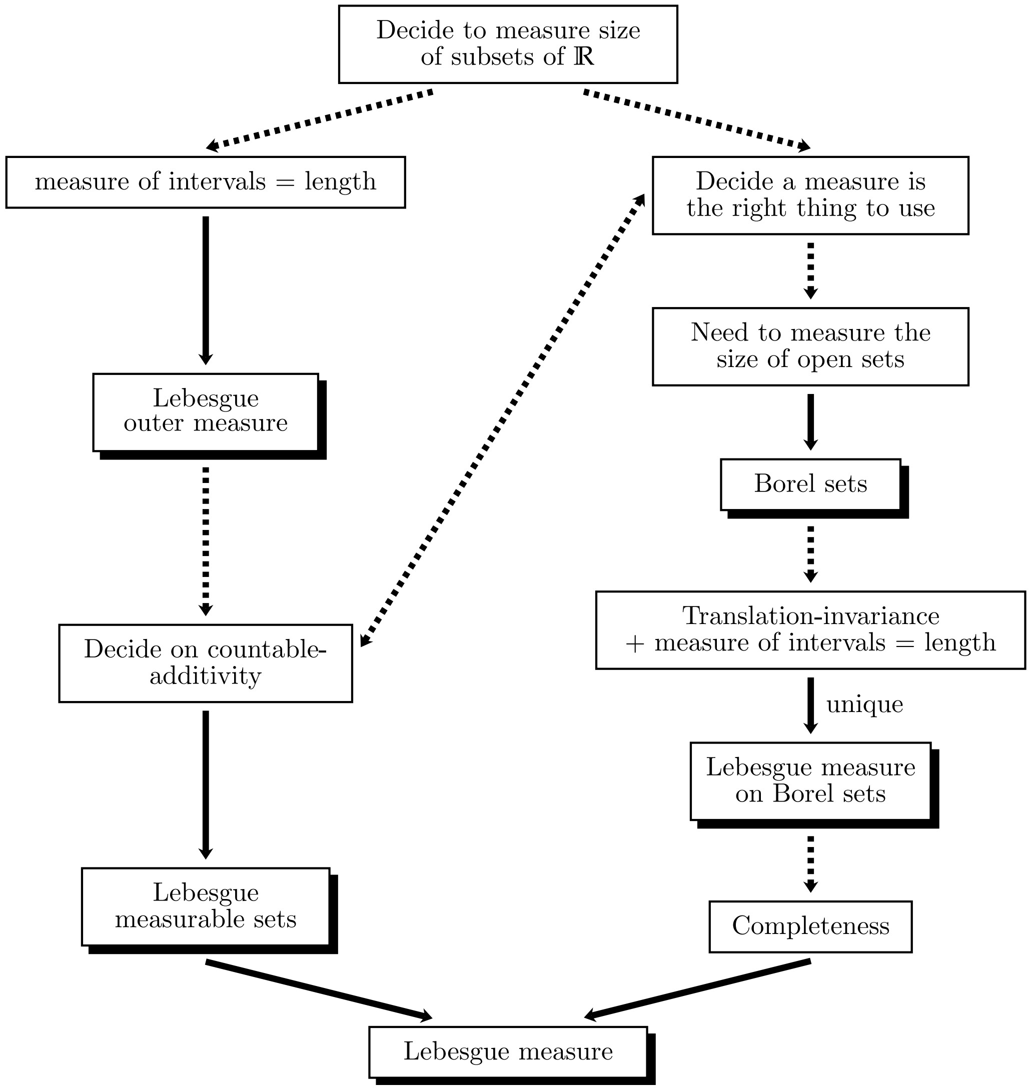


Figure 2.3 Lines of reasoning for arriving at Lebesgue measure. Dashed arrows represent choices that can be made and solid arrows represent conclusions that follow from the preceding decisions

2.4.4 Notes

The construction of the non-Lebesgue measurable subset of Example 2.4.3 is due to Vitali.

Exercises

2.4.1 Using the definition of the Lebesgue measure show that the measure of a singleton is zero.

2.4.2 Let $A \subseteq \mathbb{R}$ be Lebesgue measurable and for $\rho \in \mathbb{R}_{>0}$ define

$$\rho A = \{\rho x \mid x \in A\}.$$

Show that $\lambda(\rho A) = \rho\lambda(A)$.

2.4.3 For the following subsets of $\mathbb{R}$, verify that they are Borel subsets (and therefore measurable sets), and determine their Lebesgue measure:

(a) the bounded, open interval (a, b);
(b) the bounded, open-closed interval $(a, b]$;
(c) the bounded, closed-open interval $[a, b)$;
(d) the singleton $\{x\}$ for any $x \in \mathbb{R}$;
(e) the unbounded closed interval $[a, \infty)$;
(f) the unbounded open interval (a, ∞).

2.4.4 Show that the set $\mathbb{Q}$ of rational numbers is a Borel set.

2.4.5 Show that G_δ's and F_σ's are Borel sets.

2.4.6 Show that for $x \in \mathbb{R}$, the point mass $\delta_x \colon \mathscr{B}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ is regular.

2.4.7 Show that the counting measure $\mu \colon \mathscr{B}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ is not regular.

Section 2.5

Lebesgue measure on $\mathbb{R}^n$

Although we will make most use of the Lebesgue measure on $\mathbb{R}$, we shall certainly have occasion to refer to the Lebesgue measure in higher dimensions, and so in this section we present this. The discussion here mirrors, for the most part, that in Section 2.4, so we will on occasion be a little sparse in our discussion.

Do I need to read this section? The material in this section can be bypassed until it is needed. •

2.5.1 The Lebesgue outer measure and the Lebesgue measure on $\mathbb{R}^n$

As with the Lebesgue measure on $\mathbb{R}$, we construct the Lebesgue measure on $\mathbb{R}^n$ by first defining an outer measure. It is convenient to first define the volume of a rectangle. If $R = I_1 \times \cdots \times I_n$ is a rectangle in $\mathbb{R}^n$ we define its ***volume*** to be

$$\nu(R) = \begin{cases} \prod_{j=1}^n \ell(I_j), & \ell(I_j) < \infty,\ j \in \{1, \dots, n\}, \\ \infty, & \text{otherwise.} \end{cases}$$

With this notation we have the following definition.

2.5.1 Definition (Lebesgue outer measure on $\mathbb{R}^n$) The ***Lebesgue outer measure*** on $\mathbb{R}^n$ is defined by

$$\lambda_n^*(S) = \inf\Big\{\sum_{j=1}^{\infty} \nu(R_j) \ \Big|\ S \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} R_j,\ R_j \text{ an open bounded rectangle},\ j \in \mathbb{Z}_{>0}\Big\}. \quad \bullet$$

The Lebesgue outer measure on $\mathbb{R}^n$ has the same sort of naturality property with respect to volumes of rectangles that the Lebesgue outer measure on $\mathbb{R}$ has with respect to lengths of intervals.

2.5.2 Theorem (Lebesgue outer measure is an outer measure) *The Lebesgue outer measure on $\mathbb{R}^n$ is an outer measure. Furthermore, if $R = I_1 \times \cdots \times I_n$ is a rectangle then $\lambda_n^*(\Lambda) = \nu(\Lambda)$.*

Proof First we show that $\lambda_n^*(\varnothing) = 0$. Indeed, let $(\epsilon_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence converging to zero in $\mathbb{R}_{>0}$ and note that $\varnothing \subseteq (-\epsilon_j, \epsilon_j)^n$, $j \in \mathbb{Z}_{>0}$. Since $\lim_{j\to\infty}|\epsilon_j + \epsilon_j|^n = 0$, our assertion follows.

Next we show monotonicity of λ_n^*. This is clear since if $A \subseteq B \subseteq \mathbb{R}^n$ and if a collection of bounded open rectangles $(R_j)_{j\in\mathbb{Z}_{>0}}$ covers B, then the same collection of intervals covers A.

For countable-subadditivity of λ_n^*, let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a collection of subsets of $\mathbb{R}^n$. If $\sum_{j=1}^{\infty} \lambda_n^*(A_j) = \infty$ then countable-subadditivity follows trivially in this case, so we may as well suppose that $\sum_{j=1}^{\infty} \lambda_n^*(A_j) < \infty$. For $j \in \mathbb{Z}_{>0}$ and $\epsilon \in \mathbb{R}_{>0}$ let $(R_{j,k})_{k\in\mathbb{Z}_{>0}}$ be a collection of bounded open rectangles covering A_j and for which

$$\sum_{k=1}^{\infty} \nu(R_{j,k}) < \lambda_n^*(A_j) + \frac{\epsilon}{2^j}.$$

By Proposition I-1.7.16, $\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$ is countable. Therefore, we may arrange the rectangles $(R_{j,k})_{j,k\in\mathbb{Z}_{>0}}$ into a single sequence $(R_l)_{l\in\mathbb{Z}_{>0}}$ so that

1. $\cup_{j\in\mathbb{Z}_{>0}} A_j \subseteq \cup_{l\in\mathbb{Z}_{>0}} R_l$ and
2. $\displaystyle\sum_{l=1}^{\infty} \nu(R_l) < \sum_{l=1}^{\infty}\Big(\lambda_n^*(A_l) + \frac{\epsilon}{2^l}\Big) = \sum_{l=1}^{\infty} \lambda_n^*(A_j) + \epsilon.$

This shows that

$$\lambda_n^*\Big(\bigcup_{j\in\mathbb{Z}_{>0}} A_j\Big) \leq \sum_{j=1}^{\infty} \lambda_n^*(A_j),$$

giving countable-subadditivity of λ_n^*.

We finally show that $\lambda^*(R) = \nu(R)$ for any rectangle R. We first take R to be compact. Let R_ϵ be an open rectangle containing R and for which $\nu(R_\epsilon) = \nu(R) + \epsilon$. Then

$$R \subseteq R_\epsilon \cup \Big(\cup_{j=2}^{\infty} R_j\Big),$$

where $R_j = \emptyset$, $j \geq 2$. Thus we have $\lambda_n^*(R) < \nu(R) + \epsilon$, and since this holds for every $\epsilon \in \mathbb{R}_{>0}$ it follows that $\nu(R) \leq \lambda_n^*(R)$. Now suppose that $(R_j)_{j\in\mathbb{Z}_{>0}}$ is a family of bounded open rectangles for which $R \subseteq \cup_{j\in\mathbb{Z}_{>0}} R_j$. Since R is compact, there is a finite subset of these rectangles, let us abuse notation slightly and denote them by $(R_1, \dots, R_k)$, such that $R \subseteq \cup_{j=1}^{k} R_j$. Now let $\boldsymbol{P}$ be a partition of R such that each of the subrectangles of $\boldsymbol{P}$ is contained in one of the rectangles $R_1, \dots, R_n$. This is possible since there are only finitely many of the rectangles $R_1, \dots, R_n$. By definition of the volume of a rectangle we have

$$\nu(R) = \sum_{R'\in\boldsymbol{P}} \nu(R') \leq \sum_{j=1}^{k} \nu(R_j) = \sum_{j=1}^{k} \lambda_n^*(R_j).$$

This gives $\nu(R) = \lambda_n^*(R)$, as desired.

Now let R be a bounded rectangle. Since $R \subseteq \mathrm{cl}(R)$ we have $\lambda_n^*(R) \leq \nu(\mathrm{cl}(R)) = \nu(R)$ using monotonicity of λ_n^*. If $\epsilon \in \mathbb{R}_{>0}$ we may find a compact rectangle $R_\epsilon \subseteq R$ for which $\nu(R) \leq \nu(R_\epsilon) + \epsilon$. Since $\lambda_n^*(R_\epsilon) \leq \lambda_n^*(R)$ by monotonicity, it follows that

$$\lambda^*(R) \geq \lambda^*(R_\epsilon) = \nu(R_\epsilon) \geq \nu(R) - \epsilon.$$

Since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary $\lambda_n^*(R) \geq \nu(R)$, showing that $\lambda_n^*(R) = \nu(R)$, as desired.

Finally, if R is unbounded then for any $M \in \mathbb{R}_{>0}$ we may find a compact rectangle $R' \subseteq R$ for which $\lambda_n^*(R') > M$. Since $\lambda_n^*(R) \geq \lambda_n^*(R')$ by monotonicity this means that $\lambda_n^*(R) = \infty$. ■

As with the Lebesgue outer measure on $\mathbb{R}$, there are subsets of $\mathbb{R}^n$ that are not Lebesgue measurable.

2.5.3 Example (A set that is not λ_n^*-measurable) Let $A \subseteq (0,1)$ be the subset of $\mathbb{R}$ constructed in Example 2.4.3 that is not λ^*-measurable. Then define $A_n = A \times (0,1) \times \cdots \times (0,1) \subseteq \mathbb{R}^n$. Then recall from Example 2.4.3 that $(0,1)$ is a countable union of translates of A. Thus $(0,1)^n$ is a countable union of translates of A_n. Since λ_n^* is translation-invariant as we shall show in Theorem 2.5.22, it follows that, if A_n is λ_n^*-measurable, then we have the same dichotomy for A_n as we had for A:

1. if $\lambda_n^*(A_n) = 0$ then $\lambda_n^*((0,1)^n) = 0$;
2. if $\lambda_n^*(A_n) \in \mathbb{R}_{>0}$ then $\lambda_n^*((0,1)^n) = \infty$.

Since both of these conclusions are false, it must be the case that A_n is not λ_n^*-measurable. •

This then leads to the following definition.

2.5.4 Definition (Lebesgue measurable subsets of $\mathbb{R}^n$, Lebesgue measure on $\mathbb{R}^n$) Let λ_n^* be the Lebesgue outer measure on $\mathbb{R}^n$ and denote by $\mathscr{L}(\mathbb{R}^n)$ the set of λ_n^*-measurable subsets of $\mathbb{R}^n$. The sets in $\mathscr{L}(\mathbb{R}^n)$ are called ***Lebesgue measurable***, or merely ***measurable***, and the complete measure $\lambda_n\colon \mathscr{L}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ induced by λ_n^* is called the ***Lebesgue measure*** on $\mathbb{R}^n$. •

As with the Lebesgue measure on $\mathbb{R}$, the Lebesgue measure on $\mathbb{R}^n$ can be restricted to measurable sets.

2.5.5 Proposition (Restriction of Lebesgue measure to measurable sets) *Let* $\mathrm{A} \in \mathscr{L}(\mathbb{R}^n)$ *and denote*

(i) $\mathscr{L}(\mathrm{A}) = \{\mathrm{B} \cap \mathrm{A} \mid \mathrm{B} \in \mathscr{L}(\mathbb{R}^n)\}$ *and*

(ii) $\lambda_{\mathrm{A}}\colon \mathscr{L}(\mathrm{A}) \to \overline{\mathbb{R}}_{\geq 0}$ *given by* $\lambda_{\mathrm{A}}(\mathrm{B} \cap \mathrm{A}) = \lambda(\mathrm{B} \cap \mathrm{A})$.

Then $(\mathrm{A}, \mathscr{L}(\mathrm{A}), \lambda_{\mathrm{A}})$ *is a complete measure space.*

Proof This follows from Propositions 2.2.6, 2.3.18, and 2.3.23. ■

2.5.2 Borel sets in $\mathbb{R}^n$ as examples of Lebesgue measurable sets

Next we turn to the Borel sets in $\mathbb{R}^n$ which provide a large and somewhat comprehensible collection of Lebesgue measurable sets. We denote by $\mathscr{O}(\mathbb{R}^n)$ the open subsets of $\mathbb{R}^n$.

2.5.6 Definition (Borel subsets of $\mathbb{R}^n$) The collection of ***Borel sets*** in $\mathbb{R}^n$ is the σ-algebra generated by $\mathscr{O}(\mathbb{R}^n)$ (see Proposition 2.2.7). We denote by $\mathscr{B}(\mathbb{R}^n)$ the Borel sets in $\mathbb{R}^n$. If $A \in \mathscr{B}(\mathbb{R}^n)$ then we denote

$$\mathscr{B}(A) = \{A \cap B \mid B \in \mathscr{B}(\mathbb{R}^n)\}$$ •

While it is not so easy to come up with a satisfactory description of *all* Borel sets, it is the case that we will only encounter non-Borel sets as examples of things that are peculiar. Thus one can frequently get away with only thinking of Borel sets when one thinks about Lebesgue measurable sets. We shall be a little more precise about just what this means later.

For the moment, let us give a few examples of Borel sets. The following result gives us a ready made and very large class of Borel sets. In the following result we make the natural identification of $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ with $\mathbb{R}^{n_1+n_2}$.

2.5.7 Proposition (Products of Borel sets) *Let $\sigma(\mathscr{B}(\mathbb{R}^{n_1}) \times \mathscr{B}(\mathbb{R}^{n_2}))$ denote the σ-algebra on $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ generated by subsets of the form $B_1 \times B_2$, where $B_1 \in \mathscr{B}(\mathbb{R}^{n_1})$ and $\mathscr{B}(\mathbb{R}^{n_2})$. Then $\mathscr{B}(\mathbb{R}^{n_1+n_2}) = \sigma(\mathscr{B}(\mathbb{R}^{n_1}) \times \mathscr{B}(\mathbb{R}^{n_2}))$.*

what, make sure countability is included here

Proof By it follows that the open sets in $\mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ are countable unions of sets of the form $U_1 \times U_2$ where $U_1 \subseteq \mathbb{R}^{n_1}$ and $U_2 \subseteq \mathbb{R}^{n_2}$ are open. By Exercise 2.2.4 it follows that $\mathscr{B}(\mathbb{R}^{n_1+n_2})$ is generated by subsets from the σ-algebra $\sigma(\mathscr{B}(\mathbb{R}^{n_1}) \times \mathscr{B}(\mathbb{R}^{n_2}))$. Thus $\mathscr{B}(\mathbb{R}^{n_1+n_2}) \subseteq \sigma(\mathscr{B}(\mathbb{R}^{n_1}) \times \mathscr{B}(\mathbb{R}^{n_2}))$.

For the converse inclusion, note that the projections $\mathrm{pr}_1 \colon \mathbb{R}^{n_1+n_2} \to \mathbb{R}^{n_1}$ and $\mathrm{pr}_2 \colon \mathbb{R}^{n_1+n_2} \to \mathbb{R}^{n_2}$ are continuous. From this one can easily show (and this will be shown in Example 2.5.10–3) that $\pi_1^{-1}(B_1), \mathrm{pr}_2^{-1}(B_2) \in \mathscr{B}(\mathbb{R}^{n_1+n_2})$ for $B_1 \in \mathscr{B}(\mathbb{R}^{n_1})$ and $B_2 \in \mathscr{B}(\mathbb{R}^{n_2})$. Therefore,

$$B_1 \cap B_2 = \pi_1^{-1}(B_1) \cap \mathrm{pr}_2^{-1}(B_2) \in \mathscr{B}(\mathbb{R}^{n_1+n_2})$$

for $B_1 \in \mathscr{B}(\mathbb{R}^{n_1})$ and $B_2 \in \mathscr{B}(\mathbb{R}^{n_2})$. Thus $\sigma(\mathscr{B}(\mathbb{R}^{n_1}) \times \mathscr{B}(\mathbb{R}^{n_2})) \subseteq \mathscr{B}(\mathbb{R}^{n_1+n_2})$ since $\sigma(\mathscr{B}(\mathbb{R}^{n_1}) \times \mathscr{B}(\mathbb{R}^{n_2}))$ is the smallest σ-algebra containing products of Borel sets. ■

2.5.8 Remark ($\sigma(\mathscr{L}(\mathbb{R}^{n_1}) \times \mathscr{L}(\mathbb{R}^{n_2})) \neq \mathscr{L}(\mathbb{R}^{n_1+n_2})$) The reader will notice that the result analogous to the preceding one, but for Lebesgue measurable sets was not stated. This is because it is actually not true, as will be seen . This is an instance that illustrates that the mantra "What seems like it should be true is true" should always be verified explicitly. •

where

The following alternative characterisations of Borel sets are sometimes useful.

2.5.9 Proposition (Alternative characterisation of Borel sets) *$\mathscr{B}(\mathbb{R}^n)$ is equal to the following collections of sets:*

(i) the σ-algebra $\mathscr{B}_1$ generated by the closed subsets;

(ii) the σ-algebra $\mathscr{B}_2$ generated by rectangles of the form $(-\infty, b_1] \times \cdots \times (-\infty, b_n]$, $b_1, \dots, b_n \in \mathbb{R}$;

(iii) the σ-algebra $\mathscr{B}_3$ generated by intervals of the form $(a_1, b_1] \times \cdots \times (a_n, b_n]$, $a_j, b_j \in \mathbb{R}$, $a_j < b_n \in \mathbb{R}$, $j \in \{1, \dots, n\}$.

Proof First note that $\mathscr{B}(\mathbb{R}^n)$ contains the σ-algebra $\mathscr{B}_1$ generated by all closed sets, since the complements of all open sets, i.e., all closed sets, are contained in $\mathscr{B}(\mathbb{R}^n)$. Note that the sets of the form $(-\infty, b_1] \times \cdots \times (-\infty, b_n]$ are closed, so the σ-algebra $\mathscr{B}_2$ generated by these subsets is contained in $\mathscr{B}_1$. Since $(a_j, b_j] = (-\infty, b_j] \cap (\mathbb{R} \setminus (-\infty, a_j])$, $j \in \{1, \dots, n\}$, it follows that the σ-algebra $\mathscr{B}_3$ is contained in $\mathscr{B}_2$. Finally, note that

$$(a_j, b_j) = \cup_{k=1}^{\infty} (a_j, b_k - \tfrac{1}{k}], \qquad j \in \{1, \dots, n\}.$$

what? make sure to get countability

Thus, by , each open subset of $\mathbb{R}^n$ is a countable union of sets, each of which is a countable intersection of generators of sets of $\mathscr{B}_3$. Thus $\mathscr{B}(\mathbb{R}^n) \subseteq \mathscr{B}_3$. Putting this all together gives

$$\mathscr{B}(\mathbb{R}^n) \subseteq \mathscr{B}_3 \subseteq \mathscr{B}_2 \subseteq \mathscr{B}_1 \subseteq \mathscr{B}(\mathbb{R}^n).$$

Thus we must conclude that $\mathscr{B}_1 = \mathscr{B}_2 = \mathscr{B}_3 = \mathscr{B}(\mathbb{R}^n)$. ■

We can now give some examples of Borel sets in $\mathbb{R}^n$.

2.5.10 Examples (Borel sets)

1. We claim that if $B_1, \dots, B_n \in \mathscr{B}(\mathbb{R})$ then $B_1 \times \cdots \times B_n \in \mathscr{B}(\mathbb{R}^n)$; this follows by a simple induction from Proposition 2.5.7. This provides us with a large collection of Borel sets, provided we have Borel sets in $\mathbb{R}$.
2. As for Borel sets in $\mathbb{R}$, a set that is a countable intersection of open sets is called a $\mathbf{G}_\delta$ and a set that is a countable union of closed sets is called an $\mathbf{F}_\sigma$.
3. If $B \in \mathscr{B}(\mathbb{R}^n)$ and if $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is continuous, then $f^{-1}(B) \in \mathscr{B}(\mathbb{R}^n)$. To see, by Proposition 2.5.9 it suffices to show that

$$f^{-1}((-\infty, b_1] \times \cdots \times (-\infty, b_n])$$

is closed. If $f(x) = (f_1(x, \dots, f_m(x))$ then

$$f^{-1}((-\infty, b_1] \times \cdots \times (-\infty, b_n]) = f_1^{-1}((-\infty, b_1]) \cap \cdots \cap f_n^{-1}((-\infty, b_n]).$$

Since each of the functions $f_1, \dots, f_n$ are continuous it follows from Corollary II-1.3.4 that $f_j^{-1}((-\infty, b_n])$ is closed for each $j \in \{1, \dots, n\}$. Thus

$$f^{-1}((-\infty, b_1] \times \cdots \times (-\infty, b_n])$$

is closed, being a finite intersection of closed sets. This gives the desired conclusion.

This again gives us a wealth of Borel sets. •

Now that we understand a little of the character of Borel sets, let us provide their relationship with the Lebesgue measurable sets. As with the relationship of $\mathscr{B}(\mathbb{R})$ with $\mathscr{L}(\mathbb{R})$, the correspondence between Borel and Lebesgue measurable sets in $\mathbb{R}^n$ has its nice points and its somewhat deficient aspects.

2.5.11 Theorem (Borel sets are Lebesgue measurable) $\mathscr{B}(\mathbb{R}^n) \subseteq \mathscr{L}(\mathbb{R}^n)$.

Proof The theorem will follow from Proposition 2.5.9 if we can show that any set of the form $(-\infty, b_1] \times \cdots \times (-\infty, b_n]$ is Lebesgue measurable. Let A be such a set and note that since

$$\lambda_n^*(S) \le \lambda_n^*(S \cap A) + \lambda_n^*(S \cap (\mathbb{R}^n \setminus A))$$

we need only show the opposite inequality to show that A is Lebesgue measurable. If $\lambda_n^*(S) = \infty$ this is clearly true, so we may as well suppose that $\lambda_n^*(S) < \infty$. Let $(R_j)_{j\in\mathbb{Z}_{>0}}$ be bounded open rectangles that cover S and be such that

$$\sum_{j=1}^{\infty} \nu(R_j) < \lambda_n^*(S) + \epsilon.$$

For $j \in \mathbb{Z}_{>0}$ choose bounded open rectangles D_j and E_j, possibly empty, for which

$$\begin{aligned} &R_j \cap A \subseteq D_j, \\ &R_j \cap (\mathbb{R}^n \setminus A) \subseteq E_j, \\ &\nu(D_j) + \nu(E_j) \le \nu(R_j) + \frac{\epsilon}{2^j}. \end{aligned}$$

Note that the bounded open rectangles $(D_j)_{j\in\mathbb{Z}_{>0}}$ cover $S \cap A$ and that the bounded open rectangles $(E_j)_{j\in\mathbb{Z}_{>0}}$ cover $\mathbb{R}^n \setminus A$ so that

$$\lambda_n^*(S \cap A) \le \sum_{j=1}^{\infty} \nu(D_j), \quad \lambda_n^*(S \cap (\mathbb{R}^n \setminus A)) \le \sum_{j=1}^{\infty} \nu(E_j).$$

From this we have

$$\lambda_n^*(S \cap A) + \lambda_n^*(S \cap (\mathbb{R}^n \setminus A)) \le \sum_{j=1}^{\infty} \nu(R_j) + \epsilon < \lambda_n^*(S) + 2\epsilon,$$

using the fact that $\sum_{j=1}^{\infty} \frac{1}{2^j} = 1$ by Example I-2.4.2–1. Since ϵ can be taken arbitrarily small, the inequality

$$\lambda_n^*(S) \ge \lambda_n^*(S \cap A) + \lambda_n^*(S \cap (\mathbb{R} \setminus A))$$

follows, and so too does the result. ■

While the preceding result is useful in that it tells us that the large class of (sort of) easily understood Borel sets are Lebesgue measurable, the following result says that much more is true. Namely, up to sets of measure zero, all Lebesgue measurable sets are Borel sets.

2.5.12 Theorem (Lebesgue measurable sets are the completion of the Borel sets) $(\mathbb{R}^n, \mathscr{L}(\mathbb{R}^n), \lambda_n)$ *is the completion of* $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \lambda_n|\mathscr{B}(\mathbb{R}^n))$.

Proof First, given $A \in \mathscr{L}(\mathbb{R}^n)$, we find $L, U \in \mathscr{B}(\mathbb{R}^n)$ such that $L \subseteq A \subseteq U$ and such that $\lambda_n(U \setminus L) = 0$. We first suppose that $\lambda_n(A) < \infty$. Using Theorem 2.5.18 below, let $(U_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of open sets containing A and for which $\lambda_n(U_j) \le \lambda_n(A) + \frac{1}{j}$ and let $(L_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of compact subsets of A for which $\lambda_n(L_j) \ge \lambda_n(A) - \frac{1}{j}$. If we take $L = \cup_{j\in\mathbb{Z}_{>0}} L_j$ and $U = \cap_{j\in\mathbb{Z}_{>0}} U_j$ then we have $L \subseteq A \subseteq U$. We also have

$$\lambda_n(U \setminus L) \le \lambda_n(U_j \setminus L_j) = \lambda_n(U_j \setminus A) + \lambda_n(A \setminus L_j) \le \tfrac{1}{2j}.$$

Since this holds for every $j \in \mathbb{Z}_{>0}$, this gives our claim when A has finite measure, since L and U are Borel sets. If $\lambda_n(A) = \infty$ then we can write $A = \cup_{j\in\mathbb{Z}_{>0}} A_j$ with $A_j = (-j, j)^n \cap A$. For each $j \in \mathbb{Z}_{>0}$ we may find $L_j, U_j \in \mathscr{B}(\mathbb{R})$ such that $L_j \subseteq A_j \subseteq U_j$ and $\lambda_n(U_j \setminus L_j$. Taking $L = \cup_{j\in\mathbb{Z}_{>0}} L_j$ and $U = \cup_{j\in\mathbb{Z}_{>0}}$ gives $L \subseteq A \subseteq U$ and $\lambda_n(U \setminus L) = 0$.

The above shows that $\mathscr{L}(\mathbb{R}^n) \subseteq \mathscr{B}_{\lambda_n}(\mathbb{R}^n)$. Now let $B \in \mathscr{B}_{\lambda_n}(\mathbb{R})$ and take Borel sets L and U for which $L \subseteq B \subseteq U$ and $\lambda_n(U \setminus L) = 0$. Note that $(B \setminus L) \subseteq (U \setminus L)$. Note also that since $U \setminus L \in \mathscr{B}(\mathbb{R})$ we have $U \setminus L \in \mathscr{L}(\mathbb{R})$ and $\lambda_n(U \setminus L) = 0$. By completeness of the Lebesgue measure this implies that $B \setminus L \in \mathscr{L}(\mathbb{R})$. Since $B = (B \setminus L) \cup L$ this implies that $B \in \mathscr{L}(\mathbb{R})$. ■

The theorem has the following corollary which explicitly indicates what it means to approximate a Lebesgue measurable set with a Borel set.

2.5.13 Corollary (Borel approximations to Lebesgue measurable sets) *If* $A \in \mathscr{L}(\mathbb{R}^n)$ *then there exists a Borel set* B *and a set* Z *of measure zero such that* $A = B \cup Z$.

Proof This follows directly from Theorem 2.5.12 and the definition of the completion. ■

As is the case for $\mathscr{B}(\mathbb{R})$ and $\mathscr{L}(\mathbb{R})$, there are many more sets in $\mathscr{L}(\mathbb{R}^n)$ than there are in $\mathscr{B}(\mathbb{R}^n)$, the preceding corollary notwithstanding.

2.5.14 Proposition (The cardinalities of Borel and Lebesgue measurable sets) *We have* $\mathrm{card}(\mathscr{B}(\mathbb{R}^n)) = \mathrm{card}(\mathbb{R})$ *and* $\mathrm{card}(\mathscr{L}(\mathbb{R}^n)) = \mathrm{card}(\mathbf{2}^{\mathbb{R}})$.

Proof Since $\{x\} \in \mathscr{B}(\mathbb{R}^n)$ for every $x \in \mathbb{R}^n$ we obviously have $\mathrm{card}(\mathscr{B}(\mathbb{R}^n)) \geq \mathrm{card}(\mathbb{R}^n) = \mathrm{card}(\mathbb{R})$, the last equality holding by virtue of Theorem I-1.7.17. For the opposite inequality, note that Proposition II-1.2.21 it holds that every open set is a union of open balls with rational radius and whose centres have rational coordinates in $\mathbb{R}^n$. There are countable many such balls by Proposition I-1.7.16. Let $\mathscr{S}$ be the set of such balls and note that, adopting the notation of Theorem 2.2.14, $\mathscr{S}_1$ therefore includes the open subsets of $\mathbb{R}^n$. Thus $\mathscr{B}(\mathbb{R}^n)$ is the σ-algebra generated by $\mathscr{S}$ and so, by Theorem 2.2.14, $\mathrm{card}(\mathscr{B}(\mathbb{R}^n)) \leq \aleph_0^{\aleph_0}$. Since

$$2^{\aleph_0} \leq \aleph_0^{\aleph_0} \leq (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0 \cdot \aleph_0} = 2^{\aleph_0},$$

using the fact that $2 \leq \aleph_0 \leq 2^{\aleph_0}$ by Example I-1.7.14–3 and Exercise I-1.7.4, it follows that $\mathrm{card}(\mathscr{B}(\mathbb{R}^n)) \leq \mathrm{card}(\mathbb{R})$, as desired.

Next, we obviously have

$$\mathrm{card}(\mathscr{L}(\mathbb{R}^n)) \leq \mathrm{card}(\mathbf{2}^{\mathbb{R}^n}) = \mathrm{card}(\mathbf{2}^{\mathbb{R}}),$$

using the fact that $\mathrm{card}(\mathbb{R}^n) = \mathrm{card}(\mathbb{R})$ by Theorem I-1.7.17. For the opposite inequality, we note that the Cantor set $C \subseteq [0,1]$ has Lebesgue measure zero and has the cardinality of $[0,1]$, and thus the cardinality of $\mathbb{R}$. Thus the set $C_n = C \times \mathbb{R}^{n-1} \subseteq \mathbb{R}^n$ also has measure zero (why?), and satisfies

$$\mathrm{card}(C_n) = \mathrm{card}(C) \cdot \mathrm{card}(\mathbb{R}^n) = \mathrm{card}(\mathbb{R})^n = \mathrm{card}(\mathbb{R}),$$

using Theorem I-1.7.17. Since $\mathscr{L}(\mathbb{R}^n)$ is complete it follows that every subset of C_n is Lebesgue measurable, and so

$$\mathrm{card}(\mathscr{L}(\mathbb{R}^n)) \geq \mathrm{card}(\mathbf{2}^{C_n}) = \mathrm{card}(\mathbf{2}^{\mathbb{R}}).$$

Thus $\mathrm{card}(\mathscr{L}(\mathbb{R}^n)) = \mathrm{card}(\mathbf{2}^{\mathbb{R}})$, as desired. ■

Using the fact that this is possible when $n = 1$, it is possible to construct a Lebesgue measurable subset of $\mathbb{R}^n$ that is not Borel.

2.5.15 Example (A non-Borel Lebesgue measurable set) Let $B \subseteq [0,1]$ be the subset defined in Example 2.4.14, recalling that B is Lebesgue measurable but not Borel. We claim that $B_n = B \times \mathbb{R}^{n-1} \subseteq \mathbb{R}^n$ is Lebesgue measurable but not Borel. It is Lebesgue measurable since B is a subset of the Cantor set C which has zero measure, and so $B_n \subseteq C \times \mathbb{R}^{n-1}$ with $C \times \mathbb{R}^{n-1}$ having zero measure. Completeness of $\mathscr{L}(\mathbb{R}^n)$ ensures that B_n is Lebesgue measurable. However, B_n cannot be a Borel set. Indeed, let $i_1 \colon \mathbb{R} \to \mathbb{R}^n$ be the continuous map $i_1(x) = (x, 0, \dots, 0)$. Then one can easily see that $B = i_1^{-1}(B_n)$. Were B_n a Borel set, this would imply that B is a Borel set by Example 2.5.10–3. •

2.5.3 Further properties of the Lebesgue measure on $\mathbb{R}^n$

In this section we shall establish some important properties of the Lebesgue measure. These are intended to show the extent to which the Lebesgue measure is a natural and well-behaved construction.

We begin with an important attribute of measures in general.

2.5.16 Definition (Regular measure on $\mathbb{R}^n$) Let $\mathscr{A}$ be a σ-algebra on $\mathbb{R}^n$ that contains the Borel σ-algebra $\mathscr{B}(\mathbb{R}^n)$. A measure $\mu \colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ is ***regular*** if

(i) $\mu(K) < \infty$ for each compact subset $K \subseteq \mathbb{R}^n$,

(ii) if $A \in \mathscr{A}$ then $\mu(A) = \inf\{\mu(U) \mid U \text{ open and } A \subseteq U\}$, and

(iii) if $U \subseteq \mathbb{R}^n$ is open then $\mu(U) = \sup\{\mu(K) \mid K \text{ open and } K \subseteq U\}$. •

Let us give some simple examples to illustrate what regular means.

2.5.17 Examples (Regular and irregular measures)

1. If $x \in \mathbb{R}^n$, the point mass measure $\delta_x \colon \mathscr{B}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$\delta(B) = \begin{cases} 1, & x \in B, \\ 0, & x \notin B \end{cases}$$

is regular, as may be readily verified; see Exercise 2.5.2.

2. One can check that the counting measure $\mu \colon \mathscr{B}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ defined by

$$\mu(B) = \begin{cases} \operatorname{card}(B), & \operatorname{card}(B) < \infty, \\ \infty, & \text{otherwise} \end{cases}$$

is not regular; see Exercise 2.5.3. •

2.5.18 Theorem (Regularity of the Lebesgue measure) *The Lebesgue measure $\lambda_n\colon \mathscr{L}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ is σ-finite and regular. Moreover, for $A \in \mathscr{L}(\mathbb{R}^n)$ we have $\lambda_n(A) = \sup\{\lambda + n(K) \mid K$ compact and $K \subseteq A\}$.*

Proof To see that λ_n is σ-finite note that $\mathbb{R}^n = \cup_{k\in\mathbb{Z}_{>0}}[-k,k]^n$ with $\lambda_n([-k,k]^n) < \infty$.

Next we show that if $A \in \mathscr{L}(\mathbb{R}^n)$ then

$$\lambda_n(A) = \inf\{\lambda_n(U) \mid U \text{ open and } A \subseteq U\}.$$

Assume that $\lambda_n(A) < \infty$ since the result is obvious otherwise. Let $\epsilon \in \mathbb{R}_{>0}$ and let $(R_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of bounded open rectangles for which $A \subseteq \cup_{j\in\mathbb{Z}_{>0}} R_j$ and for which

$$\sum_{j=1}^{\infty} \nu(R_j) = \lambda_n(A) + \epsilon.$$

Now let $U = \cup_{j\in\mathbb{Z}_{>0}} R_j$, noting that U is open and that $A \subseteq U$. By Proposition 2.3.10(iii) and the fact that the measure of a rectangle is its we have

$$\lambda_n(U) \leq \sum_{j=1}^{\infty} \nu(R_j) = \lambda_n(A) + \epsilon.$$

Since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary this shows that

$$\lambda_n(A) \geq \inf\{\lambda_n(U) \mid U \text{ open and } A \subseteq U\}.$$

Since the other inequality is obvious by the basic properties of a measure, this part of the result follows.

Note that to show that λ_n is regular it suffices to prove the final assertion of the theorem since open sets are Lebesgue measurable; thus we prove the final assertion of the theorem. First suppose that $A \in \mathscr{L}(\mathbb{R}^n)$ is bounded. Then let $\tilde{K}$ be a compact set containing A. For $\epsilon \in \mathbb{R}_{>0}$ choose U open and containing $\tilde{K} \setminus A$ and for which $\lambda_n(U) \leq \lambda_n(\tilde{K} \setminus A) + \epsilon$, this being possible from by the first part of the proof. Note that $K = \tilde{K} \setminus U$ is then a compact set contained in A and that the basic properties of measure then give

$$\lambda_n(U) \leq \lambda_n(\tilde{K} \setminus A) + \epsilon \text{ and } \lambda_n(\tilde{K}) \leq \lambda_n(K) - \lambda_n(A) \quad\Longrightarrow\quad \lambda_n(K) > \lambda_n(A) - \epsilon.$$

Since ϵ can be made as small as desired, this gives the second part of the proposition when A is bounded. Define

$$A_j = (-j, j)^n \cap A,$$

and note that $(A_j)_{j\in\mathbb{Z}_{>0}}$ is an increasing sequence of sets and that $A = \cup_{j\in\mathbb{Z}_{>0}} A_j$. Therefore, by Proposition 2.3.10(iv), $\lambda_n(A) = \lim_{j\to\infty} \lambda_n(A_j)$. Then for any $M < \lambda_n(A)$ there exists $N \in \mathbb{Z}_{>0}$ such that $\lambda_n(A_N) > M$. We may now find a compact K such that $\lambda_n(K) > M$ by the fact that we have proved our assertion for bounded sets (as is A_N). Note that $K \subseteq A$ and that $M < \lambda_n(A)$ is arbitrary, and so the result follows. ∎

The theorem has the following corollary.

2.5.19 Corollary (Approximation of Lebesgue measurable sets by open and compact sets) *If* $\mathrm{A} \in \mathscr{L}(\mathbb{R}^{\mathrm{n}})$ *satisfies* $\lambda_{\mathrm{n}}(\mathrm{A}) < \infty$ *and if* $\epsilon \in \mathbb{R}_{>0}$ *then there exists an open set* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *and a compact set* $\mathrm{K} \subseteq \mathbb{R}^{\mathrm{n}}$ *such that*

$$\lambda_{\mathrm{n}}(\mathrm{U} \setminus \mathrm{A}) < \epsilon, \quad \lambda_{\mathrm{n}}(\mathrm{A} \setminus \mathrm{K}) < \epsilon.$$

Next we show that the Lebesgue measure has the quite natural property of being translation-invariant. First we provide definitions for translation-invariant σ-algebras and measures.

2.5.20 Definition (Translation-invariant σ-algebra and measure on $\mathbb{R}^{\mathrm{n}}$) A σ-algebra $\mathscr{A} \subseteq \mathbf{2}^{\mathbb{R}^n}$ is ***translation-invariant*** if, for every $A \in \mathscr{A}$ and every $\boldsymbol{x} \in \mathbb{R}^n$,

$$\boldsymbol{x} + A \triangleq \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{y} \in A\} \in \mathscr{A}.$$

A ***translation-invariant*** measure on a translation-invariant σ-algebra $\mathscr{A}$ is a map $\mu\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ for which $\mu(\boldsymbol{x} + A) = \mu(A)$ for every $A \in \mathscr{A}$ and $\boldsymbol{x} \in \mathbb{R}^n$. •

The Borel and Lebesgue measurable sets are translation-invariant.

2.5.21 Proposition (Translation-invariance of Borel and Lebesgue measurable sets) *Both* $\mathscr{B}(\mathbb{R}^{\mathrm{n}})$ *and* $\mathscr{L}(\mathbb{R}^{\mathrm{n}})$ *are translation-invariant.*

Proof Let us denote

$$\mathscr{B}'(\mathbb{R}^n) = \{B \mid \boldsymbol{x} + B \in \mathscr{B}(\mathbb{R}^n) \text{ for every } \boldsymbol{x} \in \mathbb{R}^n\}.$$

We claim that $\mathscr{B}'(\mathbb{R}^n)$ is a σ-algebra containing the open subsets of $\mathbb{R}^n$. First of all, if $U \subseteq \mathbb{R}^n$ is open then $\boldsymbol{x} + U$ is open for every $\boldsymbol{x} \in \mathbb{R}^n$ (why?) and so $U \in \mathscr{B}'(\mathbb{R}^n)$. To see that $\mathscr{B}'(\mathbb{R}^n)$ is a σ-algebra, first note that $\mathbb{R}^n = \boldsymbol{x} + \mathbb{R}^n$ for every $\boldsymbol{x} \in \mathbb{R}^n$ and so $\mathbb{R}^n \in \mathscr{B}'(\mathbb{R}^n)$. Next, let $B \in \mathscr{B}'(\mathbb{R}^n)$ and let $\boldsymbol{x} \in \mathbb{R}^n$. Then

$$\begin{aligned}\boldsymbol{x} + (\mathbb{R}^n \setminus B) &= \{\boldsymbol{x} + \boldsymbol{z} \mid \boldsymbol{z} \notin B\} = \{\boldsymbol{y} \mid \boldsymbol{y} - \boldsymbol{x} \notin B\} = \{\boldsymbol{y} \mid \boldsymbol{y} \neq \boldsymbol{x} + \boldsymbol{z},\ \boldsymbol{z} \in B\}\\ &= \{\boldsymbol{y} \mid \boldsymbol{y} \notin (\boldsymbol{x} + B)\} = \mathbb{R}^n \setminus (\boldsymbol{x} + B) \in \mathscr{B}(\mathbb{R}^n).\end{aligned}$$

Thus $\boldsymbol{x} + (\mathbb{R}^n \setminus B) \in \mathscr{B}(\mathbb{R}^n)$ for every $\boldsymbol{x} \in \mathbb{R}^n$ and so $\mathbb{R}^n \setminus B \in \mathscr{B}'(\mathbb{R}^n)$. Finally, let $(B_j)_{j\mathbb{Z}_{>0}}$ be a countable collection of subsets from $\mathscr{B}'(\mathbb{R}^n)$. Then, for $\boldsymbol{x} \in \mathbb{R}^n$ we have

$$\boldsymbol{x} + \cup_{j\in\mathbb{Z}_{>0}} B_j = \cup_{j\in\mathbb{Z}_{>0}} (\boldsymbol{x} + B_j) \in \mathscr{B}(\mathbb{R}^n)$$

and so $\cup_{j\in\mathbb{Z}_{>0}} B_j \in \mathscr{B}'(\mathbb{R}^n)$. Thus $\mathscr{B}'(\mathbb{R}^n)$ is indeed a σ-algebra containing the open sets and so we conclude that $\mathscr{B}(\mathbb{R}^n) \subseteq \mathscr{B}'(\mathbb{R}^n)$ since $\mathscr{B}(\mathbb{R}^n)$ is the σ-algebra generated by the open sets. This shows that $\mathscr{B}(\mathbb{R}^n)$ is translation-invariant.

Next let us show that $\mathscr{L}(\mathbb{R}^n)$ is translation-invariant. To do this we first show that if $S \subseteq \mathbb{R}^n$ and if $\boldsymbol{x} \in \mathbb{R}^n$ then $\lambda_n^*(\boldsymbol{x} + S) = \lambda_n^*(S)$. Indeed,

$$\begin{aligned}\lambda_n^*(\boldsymbol{x} + S) &= \inf\Big\{\sum_{j=1}^{\infty} \nu(R_j) \,\Big|\, \boldsymbol{x} + S \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} R_j\Big\} \\ &= \inf\Big\{\sum_{j=1}^{\infty} \nu(R_j') \,\Big|\, \boldsymbol{x} + S \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} \nu(\boldsymbol{x} + R_j')\Big\} \\ &= \inf\Big\{\sum_{j=1}^{\infty} \nu(R_j') \,\Big|\, S \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} R_j'\Big\} = \lambda_n^*(S),\end{aligned}$$

using the fact that for a rectangle R we have $\nu(R) = \nu(\boldsymbol{x} + R)$. Now let $A \in \mathscr{L}(\mathbb{R}^n)$ so that, for every subset $S \subseteq \mathbb{R}^n$,

$$\lambda_n^*(S) = \lambda_n^*(S \cap A) + \lambda_n^*(S \cap (\mathbb{R}^n \setminus A)).$$

Then, for $\boldsymbol{x} \in \mathbb{R}^n$ and $S \subseteq \mathbb{R}^n$,

$$\lambda_n^*(S \cap (\boldsymbol{x} + A)) = \lambda_n^*((\boldsymbol{x} + (-\boldsymbol{x} + S)) \cap (\boldsymbol{x} + A)) = \lambda_n^*((-\boldsymbol{x} + S) \cap A)$$

and, similarly,

$$\lambda_n^*(S \cap (\mathbb{R}^n \setminus (\boldsymbol{x} + A))) = \lambda_n^*((\boldsymbol{x} + (-\boldsymbol{x} + S)) \cap (\boldsymbol{x} + \mathbb{R}^n \setminus A)) = \lambda_n^*((-\boldsymbol{x} + S) \cap (\mathbb{R}^n \setminus A)).$$

Since $\lambda_n^*(-\boldsymbol{x} + S) = \lambda_n^*(S)$ this immediately gives

$$\lambda_n^*(S) = \lambda_n^*(S \cap (\boldsymbol{x} + A)) + \lambda_n^*(S \cap (\mathbb{R}^n \setminus (\boldsymbol{x} + A))),$$

showing that $\boldsymbol{x} + A \in \mathscr{L}(\mathbb{R}^n)$. ■

We may also show that the Lebesgue measure is translation-invariant, and is, moreover, in some sense unique.

2.5.22 Theorem (Translation invariance of the Lebesgue measure) *If* $\mu\colon \mathscr{B}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ *is a nonzero translation-invariant measure for which* $\mu(B) < \infty$ *for every bounded* $B \in \mathscr{B}(\mathbb{R}^n)$*, then there exists* $c \in \mathbb{R}_{>0}$ *such that* $\mu(B) = c\lambda_n(B)$ *for every* $B \in \mathscr{B}(\mathbb{R}^n)$*. Moreover, the Lebesgue measure* $\lambda_n\colon \mathscr{L}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ *is translation-invariant.*

Proof That λ_n is translation-invariant follows from the proof of Proposition 2.4.22 where we showed that $\lambda_n^*(\boldsymbol{x} + S) = \lambda_n^*(S)$ for every $S \subseteq \mathbb{R}^n$ and $\boldsymbol{x} \in \mathbb{R}^n$. To show that λ_n is, up to a positive scalar, the only translation-invariant measure we first prove two lemmata.

1 Lemma *If* $U \subseteq \mathbb{R}^n$ *is a nonempty open set, then there exists a countable collection of disjoint rectangles* $(R_j)_{j\in\mathbb{Z}_{>0}}$ *of the form*

$$R_j = [a_{j,1}, b_{j,1}) \times \cdots \times [a_{j,n}, b_{j,n})$$

such that $U = \cup_{j\in\mathbb{Z}_{>0}} R_j$.

Proof For $k \in \mathbb{Z}_{\geq 0}$ define

$$\mathscr{C}_k = \{[j_1 2^{-k}, (j_1+1)2^{-k}) \times \cdots \times [j_n 2^{-k}, (j_n+1)2^{-k}) \mid j_1, \dots, j_n \in \mathbb{Z}\}.$$

Note that, for each $k \in \mathbb{Z}_{\geq 0}$, the sets from $\mathscr{C}_k$ form a countable partition of $\mathbb{R}^n$. Also note that for $k < l$, every cube in $\mathscr{C}_l$ is also a cube in $\mathscr{C}_k$. Now let $U \subseteq \mathbb{R}^n$ be open. Let $\mathscr{D}_0 = \emptyset$. Let

$$\begin{aligned}
\mathscr{D}_1 &= \{C \in \mathscr{C}_1 \mid C \subseteq U\},\\
\mathscr{D}_2 &= \{C \in \mathscr{C}_2 \mid C \subseteq U,\ C \notin \mathscr{D}_1\},\\
&\ \vdots\\
\mathscr{D}_k &= \{C \in \mathscr{C}_k \mid C \subseteq U,\ C \notin \mathscr{D}_1 \cup \cdots \cup \mathscr{D}_{k-1}\}\\
&\ \vdots
\end{aligned}$$

The result will follow if we can show that each point $x \in U$ is contained in some $\mathscr{D}_k$, $k \in \mathbb{Z}_{>0}$. However, this follows since U is open, and so, for each $x \in U$, one can find a smallest $k \in \mathbb{Z}_{\geq 0}$ with the property that there exists $C \in \mathscr{C}_k$ with $x \in C$ and $C \subseteq U$. ▼

2 **Lemma** *The Lebesgue measure is the unique measure on* $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ *for which the measure of a rectangle is its volume.*

Proof From Theorem 2.4.2 we know that $\lambda_n(R) = \nu(R)$ for every rectangle R. Now suppose that $\mu\colon \mathscr{B}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ is a measure with the property that $\mu(R) = \nu(R)$ for every rectangle R.

First let $U \subseteq \mathbb{R}^n$ be open. By Lemma 1 we can write $U = \cup_{j \in \mathbb{Z}_{>0}} C_j$ for a countable family $(C_j)_{j \in \mathbb{Z}_{>0}}$ of disjoint bounded cubes. Therefore, since μ is a measure,

$$\mu(U) = \mu\Big(\bigcup_{j \in \mathbb{Z}_{>0}} C_j\Big) = \sum_{j=1}^{\infty} \mu(C_j) = \sum_{j=1}^{\infty} \lambda_n(C_j) = \lambda_n(U).$$

Now let B be a bounded Borel set and let U be an open set for which $B \subseteq U$. Then

$$\mu(B) \leq \mu(U) = \lambda_n(U).$$

Therefore,

$$\mu(B) \leq \inf\{\lambda)n(U) \mid U \text{ open and } B \subseteq U\} = \lambda_n(B)$$

by regularity of λ_n. Therefore, if U is a bounded open set containing B we have

$$\mu(U) = \mu(B) + \mu(U \setminus B) \leq \lambda_n(B) + \lambda_n(U \setminus B) = \lambda_n(U).$$

Since $\mu(U) = \lambda_n(U)$ it follows that $\mu(B) = \lambda_n(B)$ and $\mu(U \setminus B) = \lambda_n(U \setminus B)$.

Finally let B be an unbounded Borel set. We can then write $B = \cup_{j_1, \dots, j_n \in \mathbb{Z}} B_{j_1 \cdots j_n}$ where $(B_{j_1 \cdots j_n})_{j_1, \dots, j_n \in \mathbb{Z}}$ is the (countable by Proposition I-1.7.16) family of disjoint Borel sets

$$B_{j_1 \cdots j_n} = B \cap ([j_1, j_1+1) \times \cdots \times [j_1, j_1+1)), \qquad j_1, \dots, j_n \in \mathbb{Z}.$$

Then

$$\mu(B) = \sum_{j_1, \dots, j_n \in \mathbb{Z}} \mu(B_j) = \sum_{j_1, \dots, j_n \in \mathbb{Z}} \lambda_n(B_j) = \lambda_n(B),$$

as desired. ▼

To proceed with the proof, let $\mu\colon \mathscr{B}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ be a translation-invariant measure and let $c = \mu([0,1)^n)$. By

$$\mu(\mathbb{R}^n) = \sum_{j_1,\dots,j_n \in \mathbb{Z}} \nu([j_1, j_1+1) \times \cdots \times [j_n, j_n+1)) = 0$$

by translation-invariance of μ. Now let $\mu'\colon \mathscr{B}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ be the measure defined by $\mu'(B) = c^{-1}\mu(B)$. Now, for $k \in \mathbb{Z}_{\geq 0}$ let $\mathscr{C}_k$ be as in the proof of Lemma 1. Let $C \in \mathscr{C}_k$. We can write $[0,1)^n$ as a disjoint union of 2^{nk} intervals of the form $\boldsymbol{x}_j + C$. Therefore, by translation-invariance of μ',

$$\mu'([0,1)^n) = 2^{nk}\mu'(C), \quad \lambda_n([0,1)^n) = 2^{nk}\lambda_n(C).$$

Since $\mu'([0,1)^n) = \lambda_n([0,1)^n)$ it follows that $\mu'(C) = \lambda_n(C)$. Since every interval is a disjoint union of intervals from the sets $\mathscr{C}_k$, $k \in \mathbb{Z}_{\geq 0}$, by Lemma 1 it follows that $\mu'(C) = \lambda_n(C)$ for every cube C. Thus $\mu' = \lambda_n$ by Lemma 2 above and so $\mu = c\lambda_n$, as desired. ■

2.5.23 Theorem (There are no translation-invariant, length-preserving measures on all subsets of $\mathbb{R}^n$) *There exists no measure space* $(\mathbb{R}^n, \mathscr{A}, \mu)$ *having the joint properties that*

(i) $\mathscr{A} = 2^{\mathbb{R}^n}$,

(ii) $\mu((0,1)^n) = 1$, *and*

(iii) μ *is translation-invariant.*

Proof Were such a measure to exist, then the non-Lebesgue measurable set $A_n \subseteq (0,1)^n$ of Example 2.5.3 would be measurable. But during the course of Example 2.5.3 we saw that $(0,1)^n$ is a countable disjoint union of translates of A_n. The dichotomy illustrated in Example 2.5.3 then applies. That is, if $\mu(A_n) = 0$ then we get $\mu((0,1)^n) = 0$ and if $\mu(A_n) \in \mathbb{R}_{>0}$ then $\mu((0,1)^n) = \infty$, both of which conclusions are false. ■

Finally in this section, let us record another useful property of the Lebesgue measure, related to its being translation-invariant. From Definition II-1.3.19 the notion of an orthogonal matrix, and the notation $\mathsf{O}(n)$ to denote the set of $n \times n$ orthogonal matrices.

Invariance of measure too

2.5.24 Definition (Rotation-invariant σ-algebra and measure on $\mathbb{R}^n$) A σ-algebra $\mathscr{A} \subseteq 2^{\mathbb{R}^n}$ is ***rotation-invariant*** if, for every $A \in \mathscr{A}$ and every $\boldsymbol{R} \in \mathsf{O}(n)$, $\boldsymbol{R}(A) \in \mathscr{A}$. A ***rotation-invariant*** measure on a rotation-invariant σ-algebra $\mathscr{A}$ is a map $\mu\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ for which $\mu(\boldsymbol{R}(A)) = \mu(A)$ for every $A \in \mathscr{A}$ and $\boldsymbol{R} \in \mathsf{O}(n)$. •

We can then repeat the translation-invariant programme above for rotation-invariance. This begins with the following result.

2.5.25 Proposition (Rotation-invariance of the Borel and Lebesgue measurable sets) *Both $\mathscr{B}(\mathbb{R}^n)$ and $\mathscr{L}(\mathbb{R}^n)$ are rotation-invariant σ-algebras, and, moreover, λ_n is rotation invariant.*

Proof Let us denote

$$\mathscr{B}'(\mathbb{R}^n) = \{B \mid \boldsymbol{R}(B) \in \mathscr{B}(\mathbb{R}^n) \text{ for every } \boldsymbol{R} \in \mathsf{O}(n)\}.$$

We claim that $\mathscr{B}'(\mathbb{R}^n)$ is a σ-algebra containing the open subsets of $\mathbb{R}^n$. First of all, if $U \subseteq \mathbb{R}^n$ is open then $\boldsymbol{R}(U)$ is open for every $\boldsymbol{R} \in \mathsf{O}(n)$ since $\boldsymbol{R}$ is a homeomorphism of $\mathbb{R}^n$. Thus $U \in \mathscr{B}'(\mathbb{R}^n)$. To see that $\mathscr{B}'(\mathbb{R}^n)$ is a σ-algebra, first note that $\mathbb{R}^n = \boldsymbol{R}(\mathbb{R}^n)$ for every $\boldsymbol{R} \in \mathsf{O}(n)$ and so $\mathbb{R}^n \in \mathscr{B}'(\mathbb{R}^n)$. Next, let $B \in \mathscr{B}'(\mathbb{R}^n)$ and let $\boldsymbol{R} \in \mathsf{O}(n)$. Then

$$\begin{aligned}\boldsymbol{R}(\mathbb{R}^n \setminus B) = \{\boldsymbol{R}(\boldsymbol{z}) \mid \boldsymbol{z} \notin B\} &= \{\boldsymbol{y} \mid \boldsymbol{R}^{-1}(\boldsymbol{y}) \notin B\} = \{\boldsymbol{y} \mid \boldsymbol{y} \neq \boldsymbol{R}(\boldsymbol{z}),\ \boldsymbol{z} \in B\}\\ &= \{\boldsymbol{y} \mid \boldsymbol{y} \notin \boldsymbol{R}(B)\} = \mathbb{R}^n \setminus (\boldsymbol{R}(B)) \in \mathscr{B}(\mathbb{R}^n).\end{aligned}$$

Thus $\boldsymbol{R}(\mathbb{R}^n \setminus B) \in \mathscr{B}(\mathbb{R}^n)$ for every $\boldsymbol{R} \in \mathsf{O}(n)$ and so $\mathbb{R}^n \setminus B \in \mathscr{B}'(\mathbb{R}^n)$. Finally, let $(B_j)_{j\mathbb{Z}_{>0}}$ be a countable collection of subsets from $\mathscr{B}'(\mathbb{R}^n)$. Then, for $\boldsymbol{R} \in \mathsf{O}(n)$ we have

$$\boldsymbol{R}(\cup_{j\in\mathbb{Z}_{>0}} B_j) = \cup_{j\in\mathbb{Z}_{>0}} \boldsymbol{R}(B_j) \in \mathscr{B}(\mathbb{R}^n)$$

and so $\cup_{j\in\mathbb{Z}_{>0}} B_j \in \mathscr{B}'(\mathbb{R}^n)$. Thus $\mathscr{B}'(\mathbb{R}^n)$ is indeed a σ-algebra containing the open sets and so we conclude that $\mathscr{B}(\mathbb{R}^n) \subseteq \mathscr{B}'(\mathbb{R}^n)$ since $\mathscr{B}(\mathbb{R}^n)$ is the σ-algebra generated by the open sets. This shows that $\mathscr{B}(\mathbb{R}^n)$ is rotation-invariant.

Next let us show that $\mathscr{L}(\mathbb{R}^n)$ is rotation-invariant. To do this we first show that if $S \subseteq \mathbb{R}^n$ and if $\boldsymbol{R} \in \mathsf{O}(n)$ then $\lambda_n^*(\boldsymbol{R}(S)) = \lambda_n^*(S)$. First note by Theorem II-1.6.38 that $\nu(\boldsymbol{R}(R)) = \nu(R)$ since $\det \boldsymbol{R} \in \{-1, 1\}$ (see Exercise II-1.3.10). Then we compute

$$\begin{aligned}\lambda_n^*(\boldsymbol{R}(S)) &= \inf\Big\{\sum_{j=1}^{\infty} \nu(R_j) \Bigm| \boldsymbol{R}(S) \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} R_j\Big\}\\ &= \inf\Big\{\sum_{j=1}^{\infty} \nu(R'_j) \Bigm| \boldsymbol{R}(S) \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} \nu(\boldsymbol{R}(R'_j))\Big\}\\ &= \inf\Big\{\sum_{j=1}^{\infty} \nu(R'_j) \Bigm| S \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} R'_j\Big\} = \lambda_n^*(S).\end{aligned}$$

Now let $A \in \mathscr{L}(\mathbb{R}^n)$ so that, for every subset $S \subseteq \mathbb{R}^n$,

$$\lambda_n^*(S) = \lambda_n^*(S \cap A) + \lambda_n^*(S \cap (\mathbb{R}^n \setminus A)).$$

Then, for $\boldsymbol{R} \in \mathsf{O}(n)$ and $S \subseteq \mathbb{R}^n$,

$$\lambda_n^*(S \cap (\boldsymbol{R}(A))) = \lambda_n^*((\boldsymbol{R}\boldsymbol{R}^{-1}(S)) \cap (\boldsymbol{R}(A))) = \lambda_n^*((\boldsymbol{R}^{-1}(S)) \cap A)$$

and, similarly,

$$\lambda_n^*(S \cap (\mathbb{R}^n \setminus (\boldsymbol{R}(A)))) = \lambda_n^*((\boldsymbol{R}\boldsymbol{R}^{-1}(S)) \cap (\boldsymbol{R}(\mathbb{R}^n \setminus A))) = \lambda_n^*((\boldsymbol{R}^{-1}(S)) \cap (\mathbb{R}^n \setminus A)).$$

Since $\lambda_n^*(\boldsymbol{R}^{-1}(S)) = \lambda_n^*(S)$ this immediately gives

$$\lambda_n^*(S) = \lambda_n^*(S \cap (\boldsymbol{R}(A))) + \lambda_n^*(S \cap (\mathbb{R}^n \setminus (\boldsymbol{R}(A)))),$$

showing that $\boldsymbol{R}(A) \in \mathscr{L}(\mathbb{R}^n)$.

The final assertion in the statement of the result, that λ_n is rotation-invariant, follows from the fact, proved above, that $\lambda_n^*(S) = \lambda_n^*(\boldsymbol{R}(S))$ for every $S \subseteq \mathbb{R}^n$. ■

The following generalisation of the preceding result is also useful.

2.5.26 Proposition (Lebesgue measure and linear maps) *If* $\mathbf{L} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ *then* $\mathrm{matL(B)} \in \mathscr{B}(\mathbb{R}^m)$ *if* $\mathrm{B} \in \mathscr{B}(\mathbb{R}^n)$ *and* $\mathbf{L}(\mathrm{A}) \in \mathscr{L}(\mathbb{R}^m)$ *if* $\mathrm{A} \in \mathscr{L}(\mathbb{R}^n)$. *Moreover, if* $\mathrm{A} \in \mathscr{L}(\mathbb{R}^n)$ *then* $\lambda_m(\mathbf{L}(\mathrm{A})) = \det \mathbf{L}\lambda_n(\mathrm{A})$.

Proof ■

2.5.4 Lebesgue measure on $\mathbb{R}^n$ as a product measure

The Lebesgue measure on $\mathbb{R}^n$ is *not* the product of the Lebesgue measures on the factors of $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$. The problem, as we shall see, is that the product of the Lebesgue measures is not complete. Fortunately, while the Principle of Desired Commutativity does not apply in its simplest form, it is not too far off since the Lebesgue measure on $\mathbb{R}^n$ is the *completion* of the product measure.

First we consider the relationship between the measure spaces $(\mathbb{R}^n, \sigma(\mathscr{L}(\mathbb{R}) \times \cdots \times \mathscr{L}(\mathbb{R})), \lambda \times \cdots \times \lambda)$ and $(\mathbb{R}^n, \mathscr{L}(\mathbb{R}^n), \lambda_n)$. The first observation is the following.

2.5.27 Proposition ($\sigma(\mathscr{L}(\mathbb{R}) \times \cdots \times \mathscr{L}(\mathbb{R})) \subseteq \mathscr{L}(\mathbb{R}^n)$) *We have* $\sigma(\mathscr{L}(\mathbb{R}) \times \cdots \times \mathscr{L}(\mathbb{R})) \subseteq \mathscr{L}(\mathbb{R}^n)$.

Proof By definition, $\sigma(\mathscr{L}(\mathbb{R}) \times \cdots \times \mathscr{L}(\mathbb{R}))$ is the σ-algebra generated by the measurable rectangles in $\mathbb{R}^n$. It, therefore, suffices to show that measurable rectangles are λ_n-measurable. Thus let $A_1 \times \cdots \times A_n$ be a measurable rectangle and, by Corollary 2.4.12, write $A_j = B_j \cup Z_j$ for $B_j \in \mathscr{B}(\mathbb{R})$ and $Z_j \subseteq \mathbb{R}$ having measure zero. Then $A_1 \times \cdots \times A_n$ is a union of measurable rectangles of the form $S_1 \times \cdots \times S_n$ where $S_j \in \{B_j, Z_j\}$, $j \in \{1, \dots, n\}$. We claim that if $S_j = Z_j$ for some $j \in \{1, \dots, n\}$ then the corresponding measurable rectangle has Lebesgue measure zero, and so in particular is Lebesgue measurable. To see this, consider a measurable rectangle of the form

$$S_1 \times \cdots \times S_{j-1} \times Z_j \times S_{j+1} \times \cdots \times S_n.$$

Let $k \in \mathbb{Z}_{>0}$ and let $C_k = [-k, k]$. Let $\epsilon \in \mathbb{R}_{>0}$. Since Z_j has measure zero, there exists intervals (a_l, b_l), $l \in \mathbb{Z}_{>0}$, such that $Z_j \subseteq \cup_{l \in \mathbb{Z}_{>0}}(a_l, b_l)$ and

$$\sum_{l=1}^{\infty}(b_l - a_l) < \frac{\epsilon}{(2k)^{n-1}}.$$

Therefore,

$$\lambda_n(C_k \cap (S_1 \times \cdots \times S_{j-1} \times Z_j \times S_{j+1} \times \cdots \times S_n)) < (2k)^{n-1}\frac{\epsilon}{(2k)^{n-1}} = \epsilon.$$

Thus

$$\lambda_n(C_k \cap (S_1 \times \cdots \times S_{j-1} \times Z_j \times S_{j+1} \times \cdots \times S_n)) = 0$$

and, since

$$S_1 \times \cdots \times S_{j-1} \times Z_j \times S_{j+1} \times \cdots \times S_n = \cup_{k \in \mathbb{Z}_{>0}} (C_k \cap (S_1 \times \cdots \times S_{j-1} \times Z_j \times S_{j+1} \times \cdots \times S_n)),$$

it follows from Proposition 2.3.3 that

$$\lambda_n(S_1 \times \cdots \times S_{j-1} \times Z_j \times S_{j+1} \times \cdots \times S_n) = 0,$$

as desired. Thus the only measurable rectangle comprising $A_1 \times \cdots \times A_n$ that is possibly not of measure zero is $B_1 \times \cdots \times B_n$. By Proposition 2.5.7 (and its natural generalisation to more than two factors using a trivial induction) it follows that this set will be Borel measurable, and so Lebesgue measurable. Thus $A_1 \times \cdots \times A_n$ is a finite union of Lebesgue measurable sets and so is Lebesgue measurable. ■

An example illustrates that the inclusion from the preceding proposition is strict.

2.5.28 Example ($\sigma(\mathscr{L}(\mathbb{R}) \times \cdots \times \mathscr{L}(\mathbb{R})) \subset \mathscr{L}(\mathbb{R}^n)$) Let

$$A = \mathbb{R} \times \{\mathbf{0}_{n-1}\} \subseteq \mathbb{R}^n$$

and note that by Theorem 2.3.33 we have $\lambda \times \cdots \times \lambda(A) = 0$. Now let $E \subseteq \mathbb{R}$ be a non-Lebesgue measurable set and note that $S \triangleq E \times \{\mathbf{0}_{n-1}\} \subseteq A$ is thus a subset of measure zero, and thus an element of $\mathscr{L}(\mathbb{R}^n)$ by completeness of the n-dimensional Lebesgue measure. We claim that $S \notin \sigma(\mathscr{L}(\mathbb{R}) \times \cdots \times \mathscr{L}(\mathbb{R}))$. Indeed, by Proposition 2.2.18 it follows that if S is measurable then E must be measurable, which it is not. •

Thus the Lebesgue measure on $\mathbb{R}^n$ is *not* the product of the Lebesgue measures on its $\mathbb{R}$ factors. However, all is not lost, as the following result suggests.

2.5.29 Proposition (The Lebesgue measure is the completion of the product measure) *The measure space* $(\mathbb{R}^n, \mathscr{L}(\mathbb{R}^n), \lambda_n)$ *is the completion of the measure space* $(\mathbb{R}^n, \sigma(\mathscr{L}(\mathbb{R}) \times \cdots \times \mathscr{L}(\mathbb{R})), \lambda \times \cdots \times \lambda)$.

Proof Note that open rectangles are in $\mathscr{L}(\mathbb{R}) \times \cdots \times \mathscr{L}(\mathbb{R})$. Thus, since $\mathscr{B}(\mathbb{R}^n)$ is the σ-algebra generated by open rectangles, $\mathscr{B}(\mathbb{R}^n) \subseteq \sigma(\mathscr{L}(\mathbb{R}) \times \cdots \times \mathscr{L}(\mathbb{R}))$. Moreover, for an open rectangle we have $U_1 \times \cdots \times U_n$ we have

$$\lambda \times \cdots \times \lambda(U_1 \times \cdots \times U_n) = \lambda_n(U_1 \times \cdots \times U_n).$$

By Lemma 2 of Theorem 2.5.22 we then have

$$\lambda \times \cdots \times \lambda|\mathscr{B}(\mathbb{R}^n) = \lambda_n|\mathscr{B}(\mathbb{R}^n).$$

Now, by Proposition 2.5.27, we have

$$\mathscr{B}(\mathbb{R}^n) \subseteq \sigma(\mathscr{L}(\mathbb{R}) \times \cdots \times \mathscr{L}(\mathbb{R})) \subseteq \mathscr{L}(\mathbb{R}^n)$$

with $\lambda \times \cdots \times \lambda$ and λ_n agreeing on the left and right sets. By Theorem 2.5.12 the result follows. ■

2.5.5 Coverings of subsets of $\mathbb{R}^n$

It is useful to sometimes be able to cover subsets of $\mathbb{R}^n$ with certain types of sets—say, open balls—and such that the covering has certain desired properties. In this section we give a few such results that are useful and some of which are related to the Lebesgue measure. Various versions of the results here are known as the ***Vitali*[6] *Covering Lemma***.

The most basic such result, and the starting point for other results, is the following.

2.5.30 Lemma (Properties of coverings by balls) *Let* J *be an index set and let* $(\overline{\mathsf{B}}^n(r_j, \mathbf{x}_j))_{j\in\mathbb{Z}_{>0}}$ *be a family of balls such that*

$$\sup\{r_j \mid j \in J\} < \infty.$$

Then there exists a subset $J' \subseteq J$ *with the following properties:*

(i) J' *is countable;*

(ii) the balls $(\overline{\mathsf{B}}^n(r_{j'}, \mathbf{x}_{j'}))_{j'\in J'}$ *are pairwise disjoint;*

(iii) $\cup_{j\in J}\overline{\mathsf{B}}^n(r_j, \mathbf{x}_j) \subseteq \cup_{j'\in J'}\overline{\mathsf{B}}^n(5r_{j'}, \mathbf{x}_{j'})$.

Proof Let us first suppose that $\cup_{j\in J}\overline{\mathsf{B}}^n(r_j, x_j)$ is bounded. We inductively construct a subset J' of J as follows. Let $\rho_1 = \sup\{r_j \mid j \in J\}$ and let $j_1 \in J$ be chosen so that $r_j \geq \frac{1}{2}\rho_1$. Now suppose that $j_1, \dots, j_k$ have been defined and let

$$\rho_{k+1} = \sup\{r_j \mid j \text{ satisfies } \overline{\mathsf{B}}^n(r_j, x_j) \cap \cup_{s=1}^k \overline{\mathsf{B}}^n(r_{j_s}, x_{j_s}) = \emptyset\}.$$

If $\rho_{k+1} = 0$ then take $J' = \{j_1, \dots, j_k\}$. Otherwise define $j_{k+1} \in J \setminus \{j_1, \dots, j_k\}$ such that $r_{j_{k+1}} \geq \frac{1}{2}\rho_{k+1}$. In the case where this inductive procedure does not terminate in finitely many steps, take $J' = \{j_k \mid k \in \mathbb{Z}_{>0}\}$.

The family $(\overline{\mathsf{B}}^n(r_{j'}, x_{j'}))_{j'\in J'}$ so constructed is clearly pairwise disjoint. Moreover, if $x \in \overline{\mathsf{B}}^n(r_j, x_j)$ for some $j \in J$ we have two possibilities.

1. $j \in J'$: In this case we immediately have $x \in \cup_{j'\in J'}\overline{\mathsf{B}}^n(r_{j'}, x_{j'})$.
2. $j \notin J'$: Here we claim that there exists $j' \in J$ such that $x \in \overline{\mathsf{B}}^n(r_{j'}, x_{j'})$. Suppose otherwise. Note that since we are assuming that $\cup_{j\in J}\overline{\mathsf{B}}^n(r_j, x_j)$ is bounded and since $(\overline{\mathsf{B}}^n(r_{j'}, x_{j'}))_{j'\in J'}$ is pairwise disjoint, for every $\epsilon \in \mathbb{R}_{>0}$ we have $\lim_{k\to\infty} r_{j_k} = 0$. Therefore, there must exist $k \in \mathbb{Z}_{>0}$ such that $2r_{j_k} < r_j$. This, however, contradicts the definition of r_k, and so we must have $x \in \overline{\mathsf{B}}^n(r_{j'}, x_{j'})$ for some $j' \in J'$.

To complete the proof in this case we prove a simple geometrical lemma.

1 Sublemma *Let* $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$, *let* $r_1, r_2 \in \mathbb{R}_{>0}$ *satisfy* $r_1 \geq \frac{1}{2}r_2$, *and suppose that* $\overline{\mathsf{B}}^n(r_1, \mathbf{x}_1) \cap \overline{\mathsf{B}}^n(r_2, \mathbf{x}_2) \neq \emptyset$. *Then* $\overline{\mathsf{B}}^n(r_2, \mathbf{x}_2) \subseteq \overline{\mathsf{B}}^n(5r_1, \mathbf{x}_1)$.

[6] Giuseppe Vitali (1875–1932) was an Italian Mathematician who made important contributions to analysis.

Proof Let $x \in \overline{\mathsf{B}}^n(r_2, x_2)$ and let $y \in \overline{\mathsf{B}}^n(r_1, x_1) \cap \overline{\mathsf{B}}^n(r_2, x_2)$. Multiple applications of the triangle inequality gives

$$\|x - x_1\|_{\mathbb{R}^n} \leq \|x - x_2\|_{\mathbb{R}^n} + \|y - x_2\|_{\mathbb{R}^n} + \|y - x_1\|_{\mathbb{R}^n} \leq 5r_1,$$

as desired. ▼

From the sublemma and since we have shown that, for each $j \in J$, $\overline{\mathsf{B}}^n(r_j, x_j)$ intersects at least one of the balls $\overline{\mathsf{B}}^n(r_{j'}, x_{j'})$, $j' \in J'$, it follows that

$$\cup_{j\in J}\overline{\mathsf{B}}^n(r_j, x_j) \subseteq \cup_{j'\in J'}\overline{\mathsf{B}}^n(r_{j'}, x_{j'}),$$

as claimed.

Next we consider the case where $\cup_{j\in J}\overline{\mathsf{B}}^n(r_j, x_j)$ is not bounded. Let $\rho = \sup\{r_j \mid j \in J\}$. We inductively define J'_k, $k \in \mathbb{Z}_{>0}$, of J as follows. Define

$$J_1 = \{j \in J \mid \overline{\mathsf{B}}^n(r_j, x_j) \cap \overline{\mathsf{B}}^n(4\rho, \mathbf{0}_n) \neq \varnothing\}$$

and note that $\cup_{j\in J_1}\overline{\mathsf{B}}^n(r_j, x_j)$ is bounded since ρ is finite. Let $J''_1 \subseteq J_1$ be defined by the applying the procedure from the first part of the proof to the set J_1. Then denote

$$J'_1 = \{j \in J''_1 \mid \overline{\mathsf{B}}^n(r_j, x_j) \cap \overline{\mathsf{B}}^n(\rho, \mathbf{0}_n) \neq \varnothing\}.$$

Note that

1. $(\overline{\mathsf{B}}^n(r_{j'}, x_{j'}))_{j'\in J'_1}$ are pairwise disjoint and that
2. $\cup\{\overline{\mathsf{B}}^n(r_j, x_j) \mid \overline{\mathsf{B}}^n(r_j, x_j) \cap \overline{\mathsf{B}}^n(\rho, \mathbf{0}_n)\} \subseteq \cup_{j'\in J'_1}\overline{\mathsf{B}}^n(5r_{j'}, x_{j'})$.

Next define

$$J_2 = J_1 \cup \{j \in J \mid \overline{\mathsf{B}}^n(r_j, x_j) \cap (\overline{\mathsf{B}}^n(5\rho, \mathbf{0}_n) \setminus \overline{\mathsf{B}}^n(4\rho, \mathbf{0}_n)) \neq \varnothing\}.$$

Also take $J''_2 \subseteq J_2$ to be the subset constructed as in the first part of the proof. Then define

$$J'_2 = \{j \in J''_2 \mid \overline{\mathsf{B}}^n(r_j, x_j) \cap \overline{\mathsf{B}}^n(2\rho, \mathbf{0}_n) \neq \varnothing\}.$$

Note that since the only balls added to J_1 in forming J_2 do not intersect $\overline{\mathsf{B}}^n(\rho, \mathbf{0}_n)$, it follows that $J''_1 \subseteq J''_2$, and thus that $J'_1 \subseteq J'_2$. Moreover, note that

1. $(\overline{\mathsf{B}}^n(r_{j'}, x_{j'}))_{j'\in J'_2}$ are pairwise disjoint and that
2. $\cup\{\overline{\mathsf{B}}^n(r_j, x_j) \mid \overline{\mathsf{B}}^n(r_j, x_j) \cap \overline{\mathsf{B}}^n(2\rho, \mathbf{0}_n)\} \subseteq \cup_{j'\in J'_2}\overline{\mathsf{B}}^n(5r_{j'}, x_{j'})$.

Proceeding in this way we define $J'_1 \subseteq \cdots \subseteq J'_k \subseteq \cdots$. Then take $J' = \cup_{k\in\mathbb{Z}_{>0}}J'_k$. By Proposition I-1.7.16 it follows that J' is countable. If $j'_1, j'_2 \in J'$ then, by construction, there exists $k \in \mathbb{Z}_{>0}$ such that $j'_1, j'_2 \in J'_k$. It thus follows that $\overline{\mathsf{B}}^n(r_{j'_1}, x_{j'_1})$ and $\overline{\mathsf{B}}^n(r_{j'_2}, x_{j'_2})$ are disjoint. If $x \in \cup_{j\in J}\overline{\mathsf{B}}^n(r_j, x_j)$ we have $x \in \overline{\mathsf{B}}^n(r_{j_0}, x_{j_0})$ where $\overline{\mathsf{B}}^n(r_{j_0}, x_{j_0}) \cap \overline{\mathsf{B}}^n(k\rho, \mathbf{0}_m) \neq \varnothing$ for some $k \in \mathbb{Z}_{>0}$. Then $x \in \cup_{j'\in J'_k}\overline{\mathsf{B}}^n(5r_{j'}, x_{j'})$. This gives the lemma. ■

2.5.31 Remark (The Vitali Covering Lemma for finite coverings) If the index set J is finite in the preceding result, then one can strengthen the conclusions to assert that

$$\cup_{j\in J}\overline{\mathsf{B}}^n(r_j, \boldsymbol{x}_j) \subseteq \cup_{j'\in J'}\overline{\mathsf{B}}^n(3r_{j'}, \boldsymbol{x}_{j'}).$$

This is achieved merely by noting that one can choose the numbers $j_1, \dots, j_k$ so that $r_s = \rho_s$ for $s \in \{1, \dots, k\}$. In this case, the factor of $\frac{1}{2}$ in the sublemma can be removed with the resulting change of the factor 5 to 3. •

There is a similar such result for, not balls, but cubes. Recall from Section II-1.2.3 that a cube in $\mathbb{R}^n$ is a rectangle, all of whose sides have the same length. We shall denote by

$$\overline{\mathsf{C}}(r, \boldsymbol{x}) = [x_1 - r, x_1 + r] \times \cdots \times [x_n - r, x_n + r]$$

the cube centred at $\boldsymbol{x} \in \mathbb{R}^n$ and with sides of length $2r$.

2.5.32 Lemma (Properties of coverings by cubes) *Let* $\mathrm{J} \in \{\{1, \dots, \mathrm{m}\}, \mathbb{Z}_{>0}\}$ *and let* $(\overline{\mathsf{C}}(\mathrm{r_j}, \mathbf{x}_\mathrm{j}))_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *be a countable family of cubes. Then there exists a subset* $\mathrm{J}' \subseteq \mathrm{J}$ *with the following properties:*

(i) *the cubes* $(\overline{\mathsf{C}}(\mathrm{r_{j'}}, \mathbf{x}_{\mathrm{j'}}))_{\mathrm{j'}\in\mathrm{J'}}$ *are pairwise disjoint;*

(ii) $\cup_{\mathrm{j}\in\mathrm{J}}\overline{\mathsf{C}}(\mathrm{r_j}, \mathbf{x}_\mathrm{j}) \subseteq \cup_{\mathrm{j'}\in\mathrm{J'}}\overline{\mathsf{C}}(5\mathrm{r_{j'}}, \mathbf{x}_{\mathrm{j'}})$.

Proof The result follows easily after making some observations about cubes, relying on the general notion of a norm that we will introduce and discuss in Section 3.1. If we define

$$\|(x_1, \dots, x_n)\|_\infty = \max\{|x_1|, \dots, |x_n|\},$$

then we note from Example 3.1.3–4 that this defines a norm on $\mathbb{R}^n$. Moreover, the balls in this norm are cubes, as can be easily verified. A review of the proof of Lemma 2.5.30 shows that it is the norm properties of $\|\cdot\|_{\mathbb{R}^n}$ that are used, along with the fact that the balls in the norm $\|\cdot\|_{\mathbb{R}^n}$ are the balls in the usual sense. Thus the entire proof of Lemma 2.5.30 carries over, replacing $\overline{\mathsf{B}}^n(r, \boldsymbol{x})$ with $\overline{\mathsf{C}}(r, \boldsymbol{x})$ and $\|\cdot\|_{\mathbb{R}^n}$ with $\|\cdot\|_\infty$. ■

The importance of the preceding results is not so readily seen at a first glance. To illustrate the essence of the result, consider the following observation. Using the notation of Lemma 2.5.30, suppose that $\cup_{j\in J}\overline{\mathsf{B}}^n(r_j, \boldsymbol{x}_j)$ is bounded. The preceding result says that there is a countable *disjoint* subset these balls that covers at least $\frac{1}{5^n}$ of the volume of region covered by the complete collection of balls. The main point is that the volume fraction covered by the disjoint balls is bounded below by a quantity, $\frac{1}{5^n}$, that is independent of the covering. It is this property that will make these preceding lemmata useful in the subsequent discussion.

First we give a definition for a sort of covering which we shall show has useful properties. In the following definition, let us call the number $2r$ the ***diameter*** of a ball $\overline{\mathsf{B}}^n(r, \boldsymbol{x})$ or a cube $\overline{\mathsf{C}}(r, \boldsymbol{x})$. We denote the diameter of a ball or cube B by $\mathrm{diam}(B)$.

2.5.33 Definition (Vitali covering) Let J be an index set, let $A \subseteq \mathbb{R}^n$, and let $(B_j)_{j\in J}$ be a family of either closed balls or closed cubes, i.e., either (1) B_j is a closed ball for every $j \in J$ or (2) B_j is a closed cube for every $j \in J$. Suppose that $\mathrm{int}(B_j) \neq \emptyset$ for each $j \in J$. This family of balls or cubes is a ***Vitali covering*** of A if, for every $\epsilon \in \mathbb{R}_{>0}$ and for every $x \in A$ there exists $j \in J$ such that $x \in B_j$ and such that $\mathrm{diam}(B_j) < \epsilon$. •

Before giving the essential theorem about Vitali coverings, let us give an example of a Vitali covering.

2.5.34 Example (Vitali covering) If $A \subseteq \mathbb{R}^n$, define

$$\mathscr{C}_A = \{\overline{\mathsf{B}}^n(r,x) \subseteq \mathbb{R}^n \mid r \in \mathbb{R}_{>0},\ A \cap \overline{\mathsf{B}}^n(r,x) \neq \emptyset\}.$$

If $\epsilon \in \mathbb{R}_{>0}$ and $x \in A$ then $\overline{\mathsf{B}}^n(\epsilon,x)$ contains x and is in $\mathscr{C}_A$. This implies that $\mathscr{C}_A$ is a Vitali covering. •

As the definition implies and the above example illustrates, one might expect that a Vitali covering of a set will involve a plentiful, rather than a barely sufficient, collection of balls or cubes.

The following theorem will be useful for us in a few different places in the text.

2.5.35 Theorem (Property of Vitali coverings) *Let $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ be nonempty and let $(\mathrm{B_j})_{\mathrm{j}\in\mathrm{J}}$ be a Vitali covering of A by cubes or balls. Then there exists a countable subset $\mathrm{J}' \subseteq \mathrm{J}$ such that*

(i) the sets $(\mathrm{B_{j'}})_{\mathrm{j'}\in\mathrm{J'}}$ are pairwise disjoint and

(ii) $\lambda^*_{\mathrm{n}}(\mathrm{A} - \cup_{\mathrm{j'}\in\mathrm{J}}\mathrm{B_{j'}}) = 0$.

Proof First we suppose that A is bounded and let U be a bounded open set such that $A \subseteq U$ and define

$$J'' = \{j \in J \mid B_j \subseteq U\}$$

and note that $(B_j)_{j\in J''}$ is a Vitali cover of A (why?). We now apply the construction of either of Lemma 2.5.30 or 2.5.32 as appropriate to arrive at a countable subset $J' \subseteq J''$. For the remainder of the proof, for concreteness let us suppose that J' is infinite and write $J' = \{j_k\}_{k\in\mathbb{Z}_{>0}}$. We also recall from the proof of Lemma 2.5.30 the sequence $(\rho_k)_{k\in\mathbb{Z}_{>0}}$ of positive numbers.

Now let $N \in \mathbb{Z}_{>0}$ and let $x \in A - \cup_{k=1}^N B_{j_k}$. Since the set $\cup_{k=1}^N B_{j_k}$ is closed by Proposition II-1.2.19 and since $(B_j)_{j\in J''}$ is a Vitali covering of A, there exists $j \in J''$ such that $x \in B_j$ and $B_j \cap (\cup_{k=1}^N B_{j_k}) = \emptyset$. Suppose $m \in \mathbb{Z}_{>0}$ is such that $B_j \cap (\cup_{k=1}^m B_{j_k} = \emptyset$. Then $\mathrm{diam}(B_j) \leq \rho_{k+1}$ by definition of ρ_{k+1}. Since $\lim_{k\to\infty} \rho_k = 0$ (see the proof of Lemma 2.5.30) it must therefore be the case that there exists $m_0 \in \mathbb{Z}_{>0}$ such that $B_j \cap (\cup_{k=1}^m B_{j_k}) \neq \emptyset$ for all $m \geq m_0$. Thus $\mathrm{diam}(B_j) \leq \rho_{m_0}$ and so $\mathrm{diam}(B_j) \leq 2\mathrm{diam}(B_{m_0})$ since $\mathrm{diam}(B_{m_0}) \geq \frac{1}{2}\rho_{m_0}$. Since $B_j \cap (\cup_{k=1}^{m_0-1} B_{j_k}) = \emptyset$ we must have $B_j \cap C_{m_0} \neq \emptyset$. For $j \in J$ let B'_j be the ball or cube whose centre agrees with that of B_j but for which $\mathrm{diam}(B'_j) = 5\mathrm{diam}(B_j)$. The lemma from the proof of Lemma 2.5.30 then gives $B_j \subseteq B'_{m_0}$. Since $m_0 \geq N+1$ by virtue of the fact that $B_j \cap (\cup_{k=1}^N B_{j_k}) = \emptyset$, we then have

$$x \in B_j \subseteq B'_{m_0} \subseteq \cup_{k=N+1} B'_{j_k}.$$

This shows that
$$A - \cup_{k=1}^{N} B_{j_k} \subseteq \cup_{k=N+1} B'_{j_k}.$$

Now note that $\sum_{k=1}^{\infty} \lambda_n(B_{j_k}) < \infty$, as was shown during the proof of Lemma 2.5.30. An application of Exercise 2.5.1 then gives $\sum_{k=1}^{\infty} \lambda_n(B'_{j_k}) < \infty$. Let $\epsilon \in \mathbb{R}_{>0}$. By Proposition I-2.4.7 it follows that there exists $N \in \mathbb{Z}_{>0}$ sufficiently large that $\sum_{k=N+1}^{\infty} \lambda_n(B'_{j_k}) < \epsilon$. Therefore,
$$\lambda_n^*\Big(A - \cup_{k=1}^{N} B_{j_k}\Big) \leq \lambda_n^*\Big(\cup_{k=N+1} B'_{j_k}\Big) = \sum_{k=N+1}^{\infty} \lambda_n^*(B'_{j_k}) < \epsilon,$$
using monotonicity and subadditivity of the Lebesgue outer measure. Monotonicity of the Lebesgue outer measure shows that
$$\lambda_n^*\Big(A - \cup_{k=1}^{\infty} B_{j_k}\Big) \leq \lambda_n^*\Big(A - \cup_{k=1}^{N} B_{j_k}\Big) < \epsilon,$$
which completes the proof in the case that A is bounded.

If A is unbounded, proceed as follows. Let $(U_k)_{k\in\mathbb{Z}_{>0}}$ be a countable collection of pairwise disjoint bounded open sets for which
$$\lambda_n(\mathbb{R}^n \setminus \cup_{k=1}^{\infty} U_k) = 0.$$
Let $A_k = U_k \cap A$. For every $k \in \mathbb{Z}_{>0}$ for which $A_k \neq \emptyset$ the first part of the proof yields a countable subset $J'_k \subseteq J$ such that the family $(B_{j'_k})_{j'_k\in J'_K}$ is pairwise disjoint and such that
$$\lambda_n^*\Big(A_k - \cup_{j'_k\in J'_k} B_{j'_k}\Big) = 0.$$
Let us define $J' = \cup_{k=1}^{\infty} J'_k$ and note that, by virtue of the constructions in the first part of the proof, $(B_{j'})_{j'\in J'}$ is pairwise disjoint. Moreover,
$$A = \cup_{k=1}^{\infty} A_k \cup (A \cap (\mathbb{R}^n \setminus \cup_{l=1}^{\infty} U_l))$$
from which we conclude that
$$A - \cup_{j'\in J'} B_{j'} = (\cup_{k=1}^{\infty} A_k - \cup_{j_k=1}^{\infty} B_{j'_k}).$$
Note that J' is countable by Proposition I-1.7.16. Thus $A - \cup_{j'\in J'} B_{j'}$ is a countable union of sets of measure zero, and so is a set of measure zero. ■

2.5.6 The Banach–Tarski Paradox

In this section we give an "elementary" proof of an (in)famous result regarding the strangeness of sets that are not Lebesgue measurable. Let us state the result first and then provide some discussion. After this we will devote the remainder of the section to the proof of the theorem.

To state the result we first introduce some language to organise the statement. We recall from Definition II-1.3.17 the definition of an isometry and from Theorem II-1.3.20 the characterisation of characterisation of isometries. The group of isometries is denoted in Definition II-1.3.21 by $\mathsf{E}(n)$.

2.5.36 Definition (Piecewise congruent) Subsets $X, Y \subseteq \mathbb{R}^n$ are ***piecewise congruent*** if there exists

(i) $N \in \mathbb{Z}_{>0}$,
(ii) a partition $(X_1, \dots, X_N)$ of X, and
(iii) $\rho_1, \dots, \rho_N \in \mathsf{E}(n)$

such that $(\rho_1(X_1), \dots, \rho_N(X_N))$ is a partition of Y. •

Piecewise congruence should be viewed as follows. The set X is chopped up into N bits, and these bits are rearranged without distortion to give Y. An illustration of this in a simple case is given in Figure 2.4. The idea seems innocuous enough,

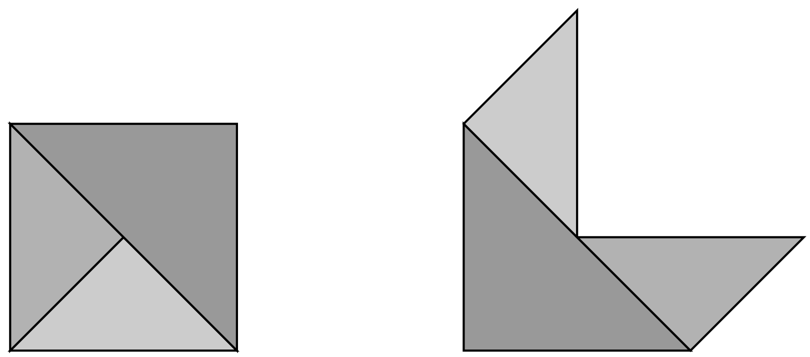


Figure 2.4 Piecewise congruent sets

but the Banach–Tarski Paradox tells us that some unexpected sets can be piecewise congruent.

2.5.37 Theorem (Banach–Tarski Paradox) *If* $X, Y \subseteq \mathbb{R}^3$ *are bounded sets with nonempty interiors then they are piecewise congruent.*

For example, the result says that one can cut up a set the size a pea into a finite number of disjoint components and reassemble these into a set the size of Jupiter. A common first reaction to this is that it is obviously false. But one should take care to understand that the theorem does not say this is true in the physical world, only in the mathematical world. In the mathematical world, or at least the one with the Axiom of Choice, there are sets whose volume does not behave as one normally expects volume to behave. It is this sort of set into which the set X is being partitioned in the theorem. For example, one should consider the set A of Example 2.4.3 that is not Lebesgue measurable. The main idea in showing that A is not Lebesgue measurable consists of showing that $(0, 1)$ can be written as a *countable* disjoint union of translates of A. This led us directly to contradictory conclusions that, if the volume of A is well-behaved, then $(0, 1)$ has either zero or infinite volume. Well, the subsets into which the sets of the Banach–Tarski Paradox are partitioned are non-Lebesgue measurable too. Thus we should not expect that the volumes of these sets behave in a decent way.

Let us now prove the Banach–Tarski Paradox. The proof is involved, but elementary. In the proof we denote by $\mathbb{S}^2$ the boundary of $\overline{\mathsf{B}}^3(1,\mathbf{0})$, i.e., $\mathbb{S}^2$ is the sphere of radius 1 in $\mathbb{R}^3$.

Our proof begins with some algebraic constructions. Define $\boldsymbol{A},\boldsymbol{B}\in\mathsf{O}(3)$ by

$$\boldsymbol{A}=\begin{bmatrix}-\cos\theta & 0 & \sin\theta\\ 0 & -1 & 0\\ \sin\theta & 0 & \cos\theta\end{bmatrix},\quad \boldsymbol{B}=\begin{bmatrix}-\frac{1}{2} & -\frac{\sqrt{3}}{2} & 0\\ \frac{\sqrt{3}}{2} & -\frac{1}{2} & 0\\ 0 & 0 & 1\end{bmatrix}.$$

The value of $\theta\in\mathbb{R}$ will be chosen shortly. One verifies directly that

$$\boldsymbol{B}^2=\begin{bmatrix}-\frac{1}{2} & \frac{\sqrt{3}}{2} & 0\\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} & 0\\ 0 & 0 & 1\end{bmatrix},\quad \boldsymbol{B}^3=\boldsymbol{A}^2=\boldsymbol{I}_3.$$

Thus $\boldsymbol{A}^{-1}=\boldsymbol{A}$ and $\boldsymbol{B}^{-1}=\boldsymbol{B}^2$. It then follows that if we define

$$\mathsf{G}=\{\boldsymbol{R}_1\cdots\boldsymbol{R}_k \mid k\in\mathbb{Z}_{>0},\ \boldsymbol{R}_j\in\{\boldsymbol{A},\boldsymbol{B}\},\ j\in\{1,\dots,k\}\}\cup\{\boldsymbol{I}_3\},$$

then G is a subgroup of $\mathsf{O}(3)$. Note that it is possible that

$$\boldsymbol{R}_1\cdots\boldsymbol{R}_k=\boldsymbol{R}'_1,\dots,\boldsymbol{R}'_{k'},$$

i.e., that different products will actually agree. We wish to eliminate this ambiguity. First of all, note that the relations $\boldsymbol{A}^3=\boldsymbol{B}^2=\boldsymbol{I}_3$ ensure that if $\boldsymbol{R}_1,\dots,\boldsymbol{R}_k\in\{\boldsymbol{A},\boldsymbol{B}\}$ then we can write

$$\boldsymbol{R}_1\cdots\boldsymbol{R}_k=\boldsymbol{R}'_1\cdots\boldsymbol{R}'_{k'}$$

for $\boldsymbol{R}'_j\in\{\boldsymbol{A},\boldsymbol{B},\boldsymbol{B}^2\}$, $j\in\{1,\dots,k'\}$. Next we claim that if $\boldsymbol{R}_1,\dots,\boldsymbol{R}_k\in\{\boldsymbol{A},\boldsymbol{B},\boldsymbol{B}^2\}$ for $k\geq 2$ then

$$\boldsymbol{R}\triangleq\boldsymbol{R}_1\cdots\boldsymbol{R}_k=\boldsymbol{R}'_1\cdots\boldsymbol{R}'_k$$

where $\boldsymbol{R}'_j\in\{\boldsymbol{B}\boldsymbol{A},\boldsymbol{B}^2\boldsymbol{A}\}$. This, however, follows from the fact that the relations $\boldsymbol{A}^3=\boldsymbol{B}^2=\boldsymbol{I}_3$ ensure that at least one of the following four possibilities hold:

1. $\boldsymbol{R}=\boldsymbol{B}^{r_1}\boldsymbol{A}\boldsymbol{B}^{r_2}\boldsymbol{A}\cdots\boldsymbol{B}^{r_m}\boldsymbol{A}$;
2. $\boldsymbol{R}=\boldsymbol{A}\boldsymbol{B}^{r_1}\boldsymbol{A}\boldsymbol{B}^{r_2}\cdots\boldsymbol{A}\boldsymbol{B}^{r_m}$;
3. $\boldsymbol{R}=\boldsymbol{B}^{r_1}\boldsymbol{A}\boldsymbol{B}^{r_2}\boldsymbol{A}\cdots\boldsymbol{B}^{r_{m-1}}\boldsymbol{A}\boldsymbol{B}^{r_m}$;
4. $\boldsymbol{R}=\boldsymbol{A}\boldsymbol{B}^{r_1}\boldsymbol{A}\boldsymbol{B}^{r_2}\cdots\boldsymbol{A}\boldsymbol{B}^{r_m}\boldsymbol{A}$.

where $r_1,\dots,r_m\in\{1,2\}$. This gives the desired conclusion. We shall call any one of these four representations a ***reduced representation***. It is still possible, after reduction to a product in $\{\boldsymbol{B}\boldsymbol{A},\boldsymbol{B}^2\boldsymbol{A}\}$, that the representation as such a product will not be unique. For example, of $\theta=\pi$ then we have $\boldsymbol{B}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{B}^2$. The following result gives a condition under which this lack of uniqueness cannot happen.

2.5.38 Lemma *If* $\cos\theta$ *is transcendental, i.e., it is not a root of any polynomial with rational coefficients, then for* $\mathbf{R} \in \mathsf{G} \setminus \{\mathbf{I}_3\}$ *there exists a unique reduced representation*

$$\mathbf{R} = \mathbf{R}_1 \cdots \mathbf{R}_k$$

for $k \in \mathbb{Z}_{>0}$ *and* $\mathbf{R}_j \in \{\mathbf{A}, \mathbf{B}, \mathbf{B}^2\}$, $j \in \{1, \dots, k\}$.

Proof The existence of the representation follows from the fact that $\boldsymbol{A}^{-1} = \boldsymbol{A}$ and $\boldsymbol{B}^{-1} = \boldsymbol{B}^2$. Thus we need only show uniqueness. It suffices to show that it is not possible to write

$$\boldsymbol{I}_3 = \boldsymbol{R}_1 \cdots \boldsymbol{R}_k$$

for $k \in \mathbb{Z}_{>0}$ and $\boldsymbol{R}_j \in \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{B}^2\}$, $j \in \{1, \dots, k\}$. Indeed, if we have

$$\boldsymbol{R}_1 \cdots \boldsymbol{R}_k = \boldsymbol{R}'_1 \cdots \boldsymbol{R}'_{k'}$$

with the factors in the products not being identical on the left and right, then

$$\boldsymbol{I}_3 = \boldsymbol{R}_k^{-1} \cdots \boldsymbol{R}_1^{-1} \boldsymbol{R}'_1 \cdots \boldsymbol{R}'_{k'},$$

giving $\boldsymbol{I}_3$ as a product in the factors $\{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{B}^2\}$.

It is clear that $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{B}^2 \neq \boldsymbol{I}_3$.

Now let $\boldsymbol{R}$ be one of the first of the four reduced representations given preceding the statement of the lemma. Thus $\boldsymbol{R} = \boldsymbol{R}_1 \cdots \boldsymbol{R}_k$ with $\boldsymbol{R}_j \in \{\boldsymbol{BA}, \boldsymbol{B}^2\boldsymbol{A}\}$, $j \in \{1, \dots, k\}$. By an elementary inductive computation on k one can check that the third component of the vector $\boldsymbol{Re}_3$ is a polynomial in $\cos\theta$ whose coefficients are rational. Since $\cos\theta$ is transcendental is cannot hold that $\boldsymbol{Re}_3 = \boldsymbol{e}_3$ and so $\boldsymbol{R} \neq \boldsymbol{I}_3$.

If $\boldsymbol{R}$ has the second of the four reduced representations then $\boldsymbol{R}' = \boldsymbol{ARA}$ has the first of the four forms, and so cannot be equal to $\boldsymbol{I}_3$. Therefore, $\boldsymbol{R} \neq \boldsymbol{I}_3$ since, if it did, we would have $\boldsymbol{R}' = \boldsymbol{A}^2 = \boldsymbol{I}_3$.

Next let $\boldsymbol{R}$ be of the third of the reduced representations, assuming that m has been chosen to be the smallest positive integer for which the representation is possible; note that we must have $m > 1$. Suppose that $\boldsymbol{R} = \boldsymbol{I}_3$. Note that

$$\boldsymbol{I}_3 = \boldsymbol{B}^{-r_1} \boldsymbol{R} \boldsymbol{B}^{r_1} = \boldsymbol{A}\boldsymbol{B}^{r_2} \cdots \boldsymbol{A}\boldsymbol{B}^{r_1 + r_m}.$$

If $r_1 = r_m$ then $\boldsymbol{B}^{r_1 + r_m} \in \{\boldsymbol{B}^2, \boldsymbol{B}^4 = \boldsymbol{B}\}$ and so this gives $\boldsymbol{I}_3$ as a reduced representation in the second of the four forms. This cannot be, so we cannot have $r_1 = r_m$. Therefore, the only other possibility is $r_1 + r_m = 3$. In this case, if $m > 3$ we have

$$\boldsymbol{I}_3 = \boldsymbol{A}\boldsymbol{B}^{r_m} \boldsymbol{R} \boldsymbol{B}^{r_1} \boldsymbol{A} = \boldsymbol{B}^{r_2} \boldsymbol{A} \cdots \boldsymbol{A} \boldsymbol{B}^{r_{m-1}},$$

contradicting our assumption that m is the smallest positive integer giving the reduced representation. Thus we must have $m \in \{2, 3\}$. For $m = 2$ we then have

$$\boldsymbol{I}_3 = \boldsymbol{B}^{r_2} \boldsymbol{R} \boldsymbol{B}^{r_1} = \boldsymbol{B}$$

and if $m = 3$ we then have

$$\boldsymbol{I}_3 = \boldsymbol{A}\boldsymbol{B}^{r_3} \boldsymbol{R} \boldsymbol{B}^{r_1} \boldsymbol{A} = \boldsymbol{B}^{r_2}.$$

Both of these conclusions are not possible, and so we cannot have $\boldsymbol{R} = \boldsymbol{I}_3$.

Finally, we consider $\boldsymbol{R}$ to have the fourth of the four reduced representations. In this case, if $\boldsymbol{R} = \boldsymbol{I}_3$ then $\boldsymbol{I}_3 = \boldsymbol{ARA}$ has the third reduced representation, giving a contradiction. ■

We now fix θ such that $\cos\theta$ is transcendental; this is possible since only a countable subset of numbers are not transcendental and since $\mathrm{image}(\cos) = [-1, 1]$. If $\boldsymbol{R} \in \mathsf{G}$, by the preceding lemma we can write

$$\boldsymbol{R} = \boldsymbol{R}_1 \cdots \boldsymbol{R}_k$$

for $\boldsymbol{R}_1, \dots, \boldsymbol{R}_k \in \{\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{B}^2\}$ with this representation being unique when it is reduced. In this case we call k the ***length*** of $\boldsymbol{R}$ which we denote by $\ell(\boldsymbol{R})$. The following lemma now uses the preceding lemma to give an essential decomposition of G.

2.5.39 Lemma *The group* G *has a partition* $(\mathsf{G}_1, \mathsf{G}_2, \mathsf{G}_3)$ *into three nonempty subsets such that*

(i) $\mathbf{R} \in \mathsf{G}_1$ *if and only if* $\mathbf{AR} \in \mathsf{G}_2 \cup \mathsf{G}_3$*;*

(ii) $\mathbf{R} \in \mathsf{G}_1$ *if and only if* $\mathbf{BR} \in \mathsf{G}_2$*;*

(iii) $\mathbf{R} \in \mathsf{G}_1$ *if and only if* $\mathbf{B}^2\mathbf{R} \in \mathsf{G}_3$*.*

Proof We define the partitions inductively by the length of their elements. For $\ell(\boldsymbol{R}) = 1$ we assign

$$\boldsymbol{I}_3 \in \mathsf{G}_1,\ \boldsymbol{A} \in \mathsf{G}_2,\ \boldsymbol{B} \in \mathsf{G}_2,\ \boldsymbol{B}^2 \in \mathsf{G}_3. \tag{2.9}$$

Now suppose that all elements $\boldsymbol{R} \in \mathsf{G}$ for which $\ell(\boldsymbol{R}) = m$ have been assigned to G_1, G_2, or G_3. If $\ell(\boldsymbol{R}) = m+1$ then write the reduced representation of $\boldsymbol{R}$ as $\boldsymbol{R} = \boldsymbol{R}_1 \cdots \boldsymbol{R}_{m+1}$. Let $\boldsymbol{R}' = \boldsymbol{R}_2 \cdots \boldsymbol{R}_{m+1}$ so that $\ell(\boldsymbol{R}') = m$. We then assign $\boldsymbol{R}$ to either G_1, G_2, or G_3 as follows:

$$\begin{aligned}
&\boldsymbol{R}_1 = \boldsymbol{A},\ \boldsymbol{R}_2 \in \{\boldsymbol{B}, \boldsymbol{B}^2\},\ \boldsymbol{R}' \in \mathsf{G}_1 \implies \boldsymbol{R} \in \mathsf{G}_2,\\
&\boldsymbol{R}_1 = \boldsymbol{A},\ \boldsymbol{R}_2 \in \{\boldsymbol{B}, \boldsymbol{B}^2\},\ \boldsymbol{R}' \in \mathsf{G}_2 \cup \mathsf{G}_3 \implies \boldsymbol{R} \in \mathsf{G}_1, && (2.10)\\
&\boldsymbol{R}_1 = \boldsymbol{B},\ \boldsymbol{R}_2 = \boldsymbol{A},\ \boldsymbol{R}' \in \mathsf{G}_j, \implies \boldsymbol{R} \in \mathsf{G}_{j+1}, && (2.11)\\
&\boldsymbol{R}_1 = \boldsymbol{B}^2,\ \boldsymbol{R}_2 = \boldsymbol{A},\ \boldsymbol{R}' \in \mathsf{G}_j, \implies \boldsymbol{R} \in \mathsf{G}_{j+2}, && (2.12)
\end{aligned}$$

where we adopt the notational convention that $\mathsf{G}_4 = \mathsf{G}_1$ and $\mathsf{G}_5 = \mathsf{G}_2$. Doing this for each m gives subsets G_1, G_2, and G_3 of G whose union equals G. Moreover, one can check that our inductive construction is unambiguous and so assigns each $\boldsymbol{R} \in \mathsf{G}$ to a unique component G_1, G_2, or G_3. It remains to show that the partition defined has the desired properties (i)–(iii).

We do this by induction on the length of the elements of G. It is obviously true for elements of length 1, using the rules prescribed above for forming the partitions. Now suppose that if $\boldsymbol{R} \in \mathsf{G}$ has length less than $m \in \mathbb{Z}_{>0}$ we have verified properties (i)–(iii). We then let $\boldsymbol{R} \in \mathsf{G}$ with $\boldsymbol{R} = \boldsymbol{R}_1 \cdots \boldsymbol{R}_m$ the unique reduced representation. We denote $\boldsymbol{R}' = \boldsymbol{R}_2 \cdots \boldsymbol{R}_m$. We consider various cases.

1. $\boldsymbol{R}_1 = \boldsymbol{A}$: We have $\boldsymbol{AR} = \boldsymbol{R}'$. Thus $\ell(\boldsymbol{AR}) = m - 1$ and so the induction hypothesis can be applied to $\boldsymbol{AR}$. Doing so yields

$$\begin{aligned}
\boldsymbol{R} \notin \mathsf{G}_1 &\iff \boldsymbol{A}^2\boldsymbol{R} \notin \mathsf{G}_1 \iff \boldsymbol{A}(\boldsymbol{AR}) \in \mathsf{G}_2 \cup \mathsf{G}_3\\
&\iff \boldsymbol{AR} \in \mathsf{G}_1 \iff \boldsymbol{AR} \notin \mathsf{G}_2 \cup \mathsf{G}_3.
\end{aligned}$$

Thus $\boldsymbol{R} \in \mathsf{G}_1$ if and only if $\boldsymbol{AR} \in \mathsf{G}_2 \cup \mathsf{G}_3$ and so (i) holds. Moreover, (2.11) and (2.12) give

$$\boldsymbol{BR} \in \mathsf{G}_2 \iff \boldsymbol{R} \in \mathsf{G}_1, \quad \boldsymbol{B}^2\boldsymbol{R} \in \mathsf{G}_3 \iff \boldsymbol{R} \in \mathsf{G}_1,$$

which gives properties (ii) and (iii).

2. $\boldsymbol{R}_1 = \boldsymbol{B}$: In this case, (2.10) immediately gives

$$\boldsymbol{A}\boldsymbol{R} \in \mathsf{G}_2 \cup \mathsf{G}_3 \iff \boldsymbol{R} \in \mathsf{G}_1,$$

which gives condition (i). We also have $\boldsymbol{B}\boldsymbol{R} = \boldsymbol{B}^2\boldsymbol{R}'$ with $\ell(\boldsymbol{R}') = m - 1$ and with $\boldsymbol{R}_2 = \boldsymbol{A}$. Thus we can apply (2.11) and (2.12) to get

$$\begin{aligned} \boldsymbol{B}\boldsymbol{R} \in \mathsf{G}_2 &\iff \boldsymbol{B}^2\boldsymbol{R}' \in \mathsf{G}_2 \iff \boldsymbol{B}^2\boldsymbol{R}' \in \mathsf{G}_5 \\ &\iff \boldsymbol{R}' \in \mathsf{G}_3, \iff \boldsymbol{B}\boldsymbol{R}' \in \mathsf{G}_4 \\ &\iff \boldsymbol{B}\boldsymbol{R}' \in \mathsf{G}_1 \iff \boldsymbol{R} \in \mathsf{G}_1 \end{aligned}$$

which gives condition (ii). We also immediately have, borrowing an implication from the preceding line,

$$\boldsymbol{B}^2\boldsymbol{R} \in \mathsf{G}_3 \iff \boldsymbol{R}' \in \mathsf{G}_3 \iff \boldsymbol{R} \in \mathsf{G}_1$$

giving condition (iii).

3. $\boldsymbol{R}_1 = \boldsymbol{B}^2$: From (2.10) we have

$$\boldsymbol{A}\boldsymbol{R} \in \mathsf{G}_2 \cup \mathsf{G}_3 \iff \boldsymbol{R} \in \mathsf{G}_1,$$

which gives condition (i). We have $\boldsymbol{B}\boldsymbol{R} = \boldsymbol{R}'$ with $\ell(\boldsymbol{R}') = m - 1$ and with $\boldsymbol{R}_2 = \boldsymbol{A}$. We then have, using (2.11) and (2.12),

$$\begin{aligned} \boldsymbol{B}\boldsymbol{R} \in \mathsf{G}_2 &\iff \boldsymbol{R}' \in \mathsf{G}_2 \iff \boldsymbol{B}^2\boldsymbol{R}' \in \mathsf{G}_4 \\ &\iff \boldsymbol{B}^2\boldsymbol{R}' \in \mathsf{G}_1 \iff \boldsymbol{R} \in \mathsf{G}_1 \end{aligned}$$

which gives condition (ii). Finally, we have

$$\boldsymbol{B}^2\boldsymbol{R} \in \mathsf{G}_3 \iff \boldsymbol{B}\boldsymbol{R}' \in \mathsf{G}_3 \iff \boldsymbol{R}' \in \mathsf{G}_2 \iff \boldsymbol{R} \in \mathsf{G}_1,$$

borrowing an implication from the preceding line. Thus we also have condition (iii). ■

Now we state the result on which the entire proof hinges. It relates the algebraic constructions thus far seen in the proof to conclusions about subsets of $\mathbb{S}^2$. It is here that we employ the Axiom of Choice in an essential way.

2.5.40 Lemma *There exists a partition* $(\mathrm{P}, \mathrm{S}_1, \mathrm{S}_2, \mathrm{S}_3)$ *of* $\mathbb{S}^2$ *for which*

(i) P *is countable,*

(ii) $\mathbf{A}(\mathrm{S}_1) = \mathrm{S}_2 \cup \mathrm{S}_3$,

(iii) $\mathbf{B}(\mathrm{S}_1) = \mathrm{S}_2$, *and*

(iv) $\mathbf{B}^2(\mathrm{S}_2) = \mathrm{S}_3$.

Proof Define

$$P = \{\boldsymbol{x} \in \mathbb{S}^2 \mid \boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{x},\ \boldsymbol{R} \in \mathsf{G} \setminus \{\boldsymbol{I}_3\}\}.$$

Since G is countable and since $\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{x}$ for two point $\boldsymbol{x} \in \mathbb{S}^2$ by Exercise II-1.3.8, it follows that P is countable. If $\boldsymbol{x} \in \mathbb{S}^2 \setminus P$ denote

$$\mathsf{G}\boldsymbol{x} = \{\boldsymbol{R}(\boldsymbol{x}) \mid \boldsymbol{R} \in \mathsf{G}\}.$$

We claim that $\mathsf{G}\boldsymbol{x} \subseteq \mathbb{S}^2 \setminus P$. Indeed, suppose otherwise. Then there exists $\boldsymbol{R} \in \mathsf{G}$ such that $\boldsymbol{R}(\boldsymbol{x}) \in P$. Then $\boldsymbol{S}\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{R}(\boldsymbol{x})$ for $\boldsymbol{S} \in \mathsf{G} \setminus \{\boldsymbol{I}_3\}$. Then $\boldsymbol{R}^{-1}\boldsymbol{S}\boldsymbol{R}(\boldsymbol{x}) = \boldsymbol{x}$ with $\boldsymbol{R}^{-1}\boldsymbol{S}\boldsymbol{R} \neq \boldsymbol{I}_3$, contradicting the assumption that $\boldsymbol{x} \notin P$. If $\boldsymbol{x}, \boldsymbol{y} \in P$, we claim that either $\mathsf{G}\boldsymbol{x} = \mathsf{G}\boldsymbol{y}$ or $\mathsf{G}\boldsymbol{x} \cap \mathsf{G}\boldsymbol{y} = \emptyset$. Indeed, suppose that $\boldsymbol{z} \in \mathsf{G}\boldsymbol{x} \cap \mathsf{G}\boldsymbol{y}$ so that $\boldsymbol{z} = \boldsymbol{R}\boldsymbol{x} = \boldsymbol{S}\boldsymbol{y}$ for $\boldsymbol{R}, \boldsymbol{S} \in \mathsf{G}$. Then let $\boldsymbol{z} \in \mathsf{G}\boldsymbol{x}$ with $\boldsymbol{z} = \boldsymbol{T}\boldsymbol{x}$. Then $\boldsymbol{z} = \boldsymbol{T}\boldsymbol{R}^{-1}\boldsymbol{S}\boldsymbol{y}$ and so $\boldsymbol{z} \in \mathsf{G}\boldsymbol{y}$. Thus

$$\{\mathsf{G}\boldsymbol{x} \mid \boldsymbol{x} \in \mathbb{S}^2 \setminus P\} \tag{2.13}$$

is a partition of $\mathbb{S}^2 \setminus P$. Let $C \subseteq \mathbb{S}^2 \setminus P$ be chosen so that it contains exactly one element of each component of this partition, using the Axiom of Choice. Now define

$$S_j = \{\boldsymbol{R}\boldsymbol{x} \mid \boldsymbol{x} \in C,\ \boldsymbol{R} \in \mathsf{G}_j\}, \qquad j \in \{1, 2, 3\}.$$

We claim that $\mathbb{S}^2 \setminus P = S_1 \cup S_2 \cup S_3$. Indeed, let $\boldsymbol{x} \in \mathbb{S}^2 \setminus P$. Then $\boldsymbol{x} = \boldsymbol{R}(\boldsymbol{x}')$ for some $\boldsymbol{x}' \in C$ and for some $\boldsymbol{R} \in \mathsf{G}$. Since $\mathsf{G} = \mathsf{G}_1 \cup \mathsf{G}_2 \cup \mathsf{G}_3$ it follows that $\boldsymbol{x} \in S_j$ for some $j \in \{1, 2, 3\}$. We also claim that $S_j \cap S_k = \emptyset$ for $j \neq k$. Indeed, suppose that $\boldsymbol{x} \in S_j \cap S_k$. Then $\boldsymbol{x} = \boldsymbol{R}_j(\boldsymbol{x}_j) = \boldsymbol{R}_k(\boldsymbol{x}_k)$ for some $\boldsymbol{R}_j \in \mathsf{G}_j$, $\boldsymbol{R}_k \in \mathsf{G}_k$, $\boldsymbol{x}_j, \boldsymbol{x}_k \in C$. Since C contains exactly one element from each component in the partition (2.13), it follows from the fact that $\boldsymbol{x}_j = \boldsymbol{R}_j^{-1}\boldsymbol{R}_k\boldsymbol{x}_k$ that $\boldsymbol{x}_j$ and $\boldsymbol{x}_k$ are in the same component of the partition and so are equal. Since $C \subseteq \mathbb{S}^2 \setminus P$ it follows that $\boldsymbol{R}_j^{-1}\boldsymbol{R}_k = \boldsymbol{I}_3$ and so $\boldsymbol{R}_j = \boldsymbol{R}_k$. Thus $j = k$. This shows that (P, S_1, S_2, S_3) is indeed a partition of $\mathbb{S}^2$.

Moreover, we compute

$$\begin{aligned}
\boldsymbol{A}(S_1) &= \{\boldsymbol{A}\boldsymbol{R}(\boldsymbol{x}) \mid \boldsymbol{R} \in \mathsf{G}_1,\ \boldsymbol{x} \in C\} = \{\boldsymbol{T}\boldsymbol{x} \mid \boldsymbol{T} \in \mathsf{G}_2 \cup \mathsf{G}_3,\ \boldsymbol{x} \in C\} = S_2 \cup S_3,\\
\boldsymbol{B}(S_1) &= \{\boldsymbol{B}\boldsymbol{R}(\boldsymbol{x}) \mid \boldsymbol{R} \in \mathsf{G}_1,\ \boldsymbol{x} \in C\} = \{\boldsymbol{T}\boldsymbol{x} \mid \boldsymbol{T} \in \mathsf{G}_2,\ \boldsymbol{x} \in C\} = S_2,\\
\boldsymbol{B}^2(S_2) &= \{\boldsymbol{B}^2\boldsymbol{R}(\boldsymbol{x}) \mid \boldsymbol{R} \in \mathsf{G}_1,\ \boldsymbol{x} \in C\} = \{\boldsymbol{T}\boldsymbol{x} \mid \boldsymbol{T} \in \mathsf{G}_3,\ \boldsymbol{x} \in C\} = S_3,
\end{aligned}$$

giving conditions (ii)–(iv). ■

The following rather technical lemma will be crucial to our proof.

2.5.41 Lemma *If* $\mathrm{P} \subseteq \mathbb{S}^2$ *is countable then there exists* $\mathrm{Q} \subseteq \mathbb{S}^2$ *countable and* $\mathbf{T} \in \mathsf{O}(3)$ *such that* $\mathrm{P} \subseteq \mathrm{Q}$ *and such that* $\mathbf{T}(\mathrm{Q}) = \mathrm{Q} - \mathrm{P}$.

Proof Let $\boldsymbol{v} = (v_1, v_2, 0) \in \mathbb{S}^1$ be such that $\boldsymbol{v}, -\boldsymbol{v} \notin P$; since P is countable this is possible. Define

$$\boldsymbol{T}_0 = \begin{bmatrix} v_1 & v_2 & 0 \\ -v_2 & v_1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and note that $\boldsymbol{T}_0 \in \mathsf{O}(3)$ since $v_1^2 + v_2^2 = 1$. Note that $\boldsymbol{T}_0(\boldsymbol{v}) = \boldsymbol{e}_1$ and that $\boldsymbol{e}_1, -\boldsymbol{e}_1 \notin \boldsymbol{T}_0(P)$. For $t \in \mathbb{R}$ define the orthogonal matrix

$$\boldsymbol{U}_t = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos t & -\sin t \\ 0 & \sin t & \cos t \end{bmatrix}.$$

Note that

$$\boldsymbol{U}_t^k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(kt) & -\sin(kt) \\ 0 & \sin(kt) & \cos(kt) \end{bmatrix}, \qquad k \in \mathbb{Z}_{>0}.$$

For $\boldsymbol{x}, \boldsymbol{y} \in P$ and for $k \in \mathbb{Z}_{>0}$ consider the equation $\boldsymbol{U}_t^k(\boldsymbol{x}) = \boldsymbol{y}$. In components this equation reads

$$x_1 = y_1,\ \cos(kt)x_2 - \sin(kt)x_3 = y_2,\ \sin(kt)x_2 + \cos(kt)x_3 = y_3.$$

If $y_1 \neq x_1$ then these equations have no solution in t. If $y_1 = x_1$ then there are infinitely many solutions in t, all satisfying $\cos(kt) = 1$ and $\sin(kt) = 0$. In particular, in $[0, 2\pi)$ there are exactly k solutions if $y_1 = x_1$. Therefore, since the set $P \times P \times \mathbb{Z}_{>0}$ is countable by Proposition I-1.7.16, it follows that the complement to the set

$$\left\{ t \in \mathbb{R} \;\middle|\; \boldsymbol{T}_0(P) \cap \left(\cup_{k\in\mathbb{Z}_{>0}} \boldsymbol{U}_t^k(\boldsymbol{T}_0(P))\right) = \varnothing \right\} \tag{2.14}$$

is countable. Thus choose a t in the set (2.14) and denote $\boldsymbol{U} = \boldsymbol{U}_t$. Then define $\boldsymbol{T} = \boldsymbol{T}_0^{-1}\boldsymbol{U}\boldsymbol{T}_0$ and

$$Q = P \cup \left(\cup_{k\in\mathbb{Z}_{>0}} \boldsymbol{T}^k(P)\right).$$

One can directly check that $\boldsymbol{U}^k\boldsymbol{T}_0 = \boldsymbol{T}_0\boldsymbol{T}^k$ for $k \in \mathbb{Z}_{>0}$. Thus, using the fact that $\boldsymbol{U}$ is defined by t satisfying (2.14), we have

$$\boldsymbol{T}_0(P \cap \boldsymbol{T}(Q)) = \boldsymbol{T}_0\Big(P \cap \left(\cup_{k\in\mathbb{Z}_{>0}} \boldsymbol{T}^k(P)\right)\Big) = \varnothing.$$

We then conclude that $P \cap \boldsymbol{T}(Q) = \varnothing$, and since $Q = P \cup \boldsymbol{T}(Q)$, as follows from the definition of Q, it follows that $\boldsymbol{T}(Q) = Q - P$. ■

Using the previous lemma, we now make a decomposition of $\mathbb{S}^2$.

2.5.42 Lemma *There exists a partition $(\mathrm{T_j})_{1\leq j\leq 10}$ of $\mathbb{S}^2$ and isometries $\boldsymbol{\sigma}_1, \dots, \boldsymbol{\sigma}_{10} \in \mathsf{E}(n)$ such that $(\boldsymbol{\sigma}_j(\mathrm{T_j}))_{1\leq j\leq 6}$ and $(\boldsymbol{\sigma}_j(\mathrm{T_j}))_{7\leq j\leq 10}$ are both partitions of $\mathbb{S}^2$.*

Proof We use the partition (P, S_1, S_2, S_3) from Lemma 2.5.40. We define

$$U_1 = \boldsymbol{A}(S_2),\ U_2 = \boldsymbol{B}\boldsymbol{A}(S_2),\ U_3 = \boldsymbol{B}^2\boldsymbol{A}(S_2),$$
$$V_1 = \boldsymbol{A}(S_3),\ U_2 = \boldsymbol{B}\boldsymbol{A}(S_3),\ U_3 = \boldsymbol{B}^2\boldsymbol{A}(S_3).$$

By Lemma 2.5.40 we see that (U_j, V_j) is a partition of S_j for each $j \in \{1, 2, 3\}$. Now define

$$T_7 = U_1,\ T_8 = U_2,\ T_9 = U_3,\ T_{10} = P,$$
$$\boldsymbol{\sigma}_7 = \boldsymbol{B}^2\boldsymbol{A},\ \boldsymbol{\sigma}_8 = \boldsymbol{A}\boldsymbol{B}^2,\ \boldsymbol{\rho}_9 = \boldsymbol{B}\boldsymbol{A}\boldsymbol{B},\ \boldsymbol{\rho}_{10} = \boldsymbol{I}_3.$$

We can then check that $\boldsymbol{\sigma}_{10}(T_{10}) = P$ and $\boldsymbol{\sigma}_j(T_j) = S_{j-6}$ for $j \in \{7, 8, 9\}$. Thus $(\boldsymbol{\sigma}_j(T_j))_{7\leq j\leq 10}$ is a partition of $\mathbb{S}^2$. Next note that

$$\mathbb{S}^2 \setminus (T_7 \cup T_8 \cup T_9 \cup T_{10}) = V_1 \cup V_2 \cup V_3.$$

Let $Q \subseteq \mathbb{S}^2$ and $\boldsymbol{T} \in \mathsf{O}(3)$ be as in Lemma 2.5.41 and define

$$T_1 = \boldsymbol{\sigma}_8(S_1 \cap Q),\ T_2 = \boldsymbol{\sigma}_9(S_2 \cap Q),\ T_3 = \boldsymbol{\sigma}_7(S_3 \cap Q),$$
$$T_1 = \boldsymbol{\sigma}_8(S_1 \setminus Q),\ T_2 = \boldsymbol{\sigma}_9(S_2 \setminus Q),\ T_3 = \boldsymbol{\sigma}_7(S_3 \setminus Q).$$

Then (T_1, T_4) partitions $\boldsymbol{\rho}_8(S_1) = V_1$, (T_2, T_5) partitions $\boldsymbol{\rho}_9(S_2) = V_2$, and (T_3, T_6) partitions $\boldsymbol{\rho}_7(S_3) = V_3$. Therefore, $(T_j)_{1\le j\le 10}$ is a partition of $\mathbb{S}^2$. Finally, define

$$\boldsymbol{\sigma}_4 = \boldsymbol{\sigma}_8^{-1},\ \boldsymbol{\sigma}_5 = \boldsymbol{\sigma}_9^{-1},\ \boldsymbol{\sigma}_6 = \boldsymbol{\sigma}_7^{-1},\ \boldsymbol{\sigma}_j = \boldsymbol{R}\boldsymbol{\sigma}_{j+3}, \qquad j \in \{1, 2, 3\}.$$

One can then directly check that $\boldsymbol{\sigma}_{j+3}(T_{j+3}) = S_j \setminus Q$, $j \in \{1, 2, 3\}$ so that

$$\cup_{j=1}^{3} \boldsymbol{\sigma}_{j+3}(T_{j+3}) = \mathbb{S}^2 \setminus Q$$

by virtue of the fact that $P \subseteq Q$. Moreover,

$$\boldsymbol{\sigma}_j(T_j) = \boldsymbol{R}^{-1}\boldsymbol{\sigma}_{j+3}(T_j) = \boldsymbol{R}^{-1}(S_j \cap Q), \qquad j \in \{1, 2, 3\},$$

which shows that $\boldsymbol{\sigma}_j(T_j) \cap \boldsymbol{\sigma}_k(T_k) = \varnothing$ if $j \neq k$. This also shows that

$$\cup_{j=1}^{3} \boldsymbol{\sigma}_j(T_j) = \boldsymbol{R}^{-1}(Q - P) = Q.$$

Thus $(\boldsymbol{\sigma}_j(T_j))_{1\le j\le 6}$ is a partition of $\mathbb{S}^2$, completing the proof. ■

2.5.43 Lemma *For* $r \in \mathbb{R}_{>0}$ *and* $\mathbf{x}_0 \in \mathbb{R}^3$ *there exists a partition* $(B_j)_{1\le j\le 40}$ *of* $\overline{\mathsf{B}}^3(r, \mathbf{x}_0)$ *and isometries* $\boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_{40} \in \mathsf{E}(n)$ *such that* $(\boldsymbol{\rho}_j(B_j))_{1\le j\le 24}$ *and* $(\boldsymbol{\rho}_j(B_j))_{25\le j\le 40}$ *are both partitions of* $\overline{\mathsf{B}}^3(r, \mathbf{x}_0)$.

Proof Let us first prove the result when $r = 1$ and $\boldsymbol{x}_0 = \boldsymbol{0}$ in which case $\mathrm{bd}(\overline{\mathsf{B}}^3(1, \boldsymbol{0})) = \mathbb{S}^2$. For $S \subseteq \mathbb{S}^2$ let us denote

$$\hat{S} = \{\lambda \boldsymbol{x} \mid \lambda \in (0, 1],\ \boldsymbol{x} \in S\}.$$

Thus, for example, $\hat{\mathbb{S}}^2 = \overline{\mathsf{B}}^3(1, \boldsymbol{0}) \setminus \{\boldsymbol{0}\}$.

Let $P = \{\boldsymbol{e}_1\}$ and, by Lemma 2.5.41, let $Q \subseteq \mathbb{S}^2$ and $\boldsymbol{R}_0 \in \mathsf{O}(3)$ be such that Q is countable, $P \subseteq Q$, and $\boldsymbol{R}_0(Q) = Q \setminus P$. Define

$$N_1 = \left\{\tfrac{1}{2}(\boldsymbol{x} - \boldsymbol{e}_1) \;\middle|\; \boldsymbol{x} \in Q\right\}$$

and define $\boldsymbol{\rho}_0 \in \mathsf{E}(3)$ by

$$\boldsymbol{\rho}_0(\boldsymbol{x}) = \boldsymbol{R}_0(\boldsymbol{x} + \tfrac{1}{2}\boldsymbol{e}_1) - \tfrac{1}{2}\boldsymbol{e}_1;$$

thus $\boldsymbol{\rho}_0$ is a rotation about $\frac{1}{2}\boldsymbol{e}_1$. Note that $\boldsymbol{0} \in N_1$ and that $\boldsymbol{\rho}_0(N_1) = N_1 \setminus \{\boldsymbol{0}\}$. Denote

$$N_2 = \overline{\mathsf{B}}^3(1, \boldsymbol{0}) \setminus N_1,\ \boldsymbol{f}_1 = \boldsymbol{\rho}_0,\ \boldsymbol{f}_2 = \boldsymbol{I}_3,\ M_k = \boldsymbol{f}_k(N_k), \qquad j \in \{1, 2\}.$$

Then we have (N_1, N_2) as a partition of $\overline{\mathsf{B}}^3(1, \boldsymbol{0})$ and (M_1, M_2) as a partition of $\hat{\mathbb{S}}^2$. Let $(T_j)_{1\le j\le 10}$ and $(\boldsymbol{\sigma}_j)_{1\le j\le 10}$ be as in Lemma 2.5.42, noting that $(\hat{T}_j)_{1\le j\le 10}$ is a partition of $\hat{\mathbb{S}}^2$.

Note that

$$(M_k \cap \hat{T}_j \cap \boldsymbol{\sigma}_j(M_l) \mid \ k, l \in \{1, 2\}, \ j \in \{1, \dots, 10\})$$

is a partition of $\hat{\mathbb{S}}^2$ into forty components. Moreover, if we define

$$B_{klj} = \boldsymbol{f}_k^{-1}(M_k \cap \hat{T}_j \cap \boldsymbol{\sigma}_j^{-1}(M_l)), \qquad k, l \in \{1, 2\}, \ j \in \{1, \dots, 10\},$$

then these sets partition $\overline{\mathsf{B}}^3(1, \mathbf{0})$. Moreover, for fixed $j \in \{1, \dots, 10\}$, the sets

$$\boldsymbol{\sigma}_j \circ \boldsymbol{f}_k(B_{klj}) = M_l \cap \boldsymbol{\sigma}_j(M_k \cap \hat{T}_j), \qquad k, l \in \{1, 2\},$$

partition $\boldsymbol{\sigma}_j(\hat{T}_j)$. By Lemma 2.5.42 we have that

$$(\boldsymbol{\sigma}_j \circ \boldsymbol{f}_k(B_{klj}) \mid \ k, l \in \{1, 2\}, \ j \in \{1, \dots, 6\}),$$
$$(\boldsymbol{\sigma}_j \circ \boldsymbol{f}_k(B_{klj}) \mid \ k, l \in \{1, 2\}, \ j \in \{7, \dots, 10\})$$

each partition $\hat{\mathbb{S}}^2$. Therefore, if we define $\boldsymbol{\rho}_{klj} = \boldsymbol{f}_l^{-1} \circ \boldsymbol{\sigma}_j \circ \boldsymbol{f}_k$ we see that

$$(\boldsymbol{\rho}_{klj}(B_{klj}) \mid \ k, l \in \{1, 2\}, \ j \in \{1, \dots, 6\}),$$
$$(\boldsymbol{\rho}_{klj}(B_{klj}) \mid \ k, l \in \{1, 2\}, \ j \in \{7, \dots, 10\})$$

each partition $\overline{\mathsf{B}}^3(1, \mathbf{0})$. This proves the lemma for $r = 1$ and $\boldsymbol{x}_0 = \mathbf{0}$.

In general, we define $B'_{klj} \subseteq \overline{\mathsf{B}}^3(r, \boldsymbol{x}_0)$ and $\boldsymbol{\rho}'_{klj} \in \mathsf{E}(3)$, $k, l \in \{1, 2\}$, $j \in \{1, \dots, 10\}$, by

$$B'_{klj} = \{r\boldsymbol{x} + \boldsymbol{x}_0 \mid \ \boldsymbol{x} \in B_{klj}\}$$

and

$$\boldsymbol{\rho}'_{klj}(\boldsymbol{x}) = r\boldsymbol{\rho}_{klj}(r^{-1}(\boldsymbol{x} - \boldsymbol{x}_0)) + \boldsymbol{x}_0$$

and then directly verify that

$$(\boldsymbol{\rho}'_{klj}(B'_{klj}) \mid \ k, l \in \{1, 2\}, \ j \in \{1, \dots, 6\}),$$
$$(\boldsymbol{\rho}'_{klj}(B'_{klj}) \mid \ k, l \in \{1, 2\}, \ j \in \{7, \dots, 10\})$$

each partition $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_0)$. ■

Now let us introduce some notation that will be convenient in the remainder of the proof. If set $X, Y \subseteq \mathbb{R}^n$ are piecewise congruent then we write $X \sim Y$. If X is piecewise congruent to a subset of Y then we write $X \precsim Y$. The following lemma records some useful facts about these relations.

2.5.44 Lemma *For $X, Y, Z \subseteq \mathbb{R}^n$ the following statements hold:*

(i) $X \sim X$;

(ii) *if* $X \sim Y$ *then* $Y \sim X$;

(iii) *if* $X \sim Y$ *and* $Y \sim Z$ *then* $X \sim Z$;

(iv) *if* $X \sim Y$ *then* $X \precsim Y$;

(v) *if* $X \precsim Y$ *and* $Y \precsim Z$ *then* $X \precsim Z$*;*

(vi) *if* $X \subseteq Y$ *then* $X \precsim Y$*;*

(vii) *if* $X \precsim Y$ *and* $Y \precsim X$ *then* $X \sim Y$.

Proof (i) This is obvious.

(ii) This follows since if ρ is an isometry then it is invertible and ρ^{-1} is an isometry.

(iii) Let $(X_1, \dots, X_N)$ and $(Y_1, \dots, Y_M)$ be partitions of X and Y, respectively, with $\rho_1, \dots, \rho_N \in \mathsf{E}(n)$ and $\sigma_1, \dots, \sigma_M \in \mathsf{E}(n)$ such that $(\rho_j(X_j))_{j\in\{1,\dots,N\}}$ and $(\sigma_k(Y_k))_{k\in\{1,\dots,M\}}$ are partitions of Y and Z, respectively. Then, for $j \in \{1, \dots, N\}$ and $k \in \{1, \dots, M\}$, define $A_{jk} = X_j \cap \rho_j^{-1}(Y_k)$, noting that the sets A_{jk}, $j \in \{1, \dots, N\}$, $k \in \{1, \dots, M\}$, form a partition of X. Thus the sets $\rho_j(A_{jk}) = Y_k \cap \rho_j(X_j)$, $j \in \{1, \dots, N\}$, $k \in \{1, \dots, M\}$, form a partition of Y and the sets $\sigma_k \circ \rho_j(A_{ij})$, $j \in \{1, \dots, N\}$, $k \in \{1, \dots, M\}$, form a partition of Z. Since $\sigma_k \circ \rho_j \in \mathsf{E}(n)$ for each $j \in \{1, \dots, N\}$ and $k \in \{1, \dots, M\}$, it follows that $X \sim Z$, as desired.

(iv) This follows because $Y \subseteq Y$.

(v) Let $(X_1, \dots, X_N)$ and $(Y_1, \dots, Y_M)$ be partitions of X and Y, respectively, with $\rho_1, \dots, \rho_N \in \mathsf{E}(n)$ and $\sigma_1, \dots, \sigma_M \in \mathsf{E}(n)$ such that $(\rho_j(X_j))_{j\in\{1,\dots,N\}}$ and $(\sigma_k(Y_k))_{k\in\{1,\dots,M\}}$ are partitions of $Y' \subseteq Y$ and $Z' \subseteq Z$, respectively. Then, for $j \in \{1, \dots, N\}$ and $k \in \{1, \dots, M\}$, define $A_{jk} = X_j \cap \rho_j^{-1}(Y_k)$, noting that the sets A_{jk}, $j \in \{1, \dots, N\}$, $k \in \{1, \dots, M\}$, form a partition of X. Thus, for fixed $k \in \{1, \dots, M\}$, the sets $\rho_j(A_{jk}) = Y_k \cap \rho_j(X_j)$, $j \in \{1, \dots, N\}$, form a partition of $Y_k \cap Y'$ and the sets $\sigma_k \circ \rho_j(A_{ij})$, $j \in \{1, \dots, N\}$, $k \in \{1, \dots, M\}$, then form a partition for some subset $Z'' \subseteq Z$. Since $\sigma_k \circ \rho_j \in \mathsf{E}(n)$ for each $j \in \{1, \dots, N\}$ and $k \in \{1, \dots, M\}$, it follows that $X \precsim Z$, as desired.

(vi) This is obvious.

(vii) Suppose that $X \sim Y'$ and $Y \sim X'$ for $X' \subseteq X$ and $Y' \subseteq Y$. Let $(X_1, \dots, X_N)$ and $(Y_1, \dots, Y_M)$ be partitions of X and Y, respectively, with $\rho_1, \dots, \rho_N \in \mathsf{E}(n)$ and $\sigma_1, \dots, \sigma_M \in \mathsf{E}(n)$ such that $(\rho_j(X_j))_{j\in\{1,\dots,N\}}$ and $(\sigma_k(Y_k))_{k\in\{1,\dots,M\}}$ are partitions of $Y' \subseteq Y$ and $X' \subseteq X$, respectively. Define bijections $\rho\colon X \to Y'$ and $\sigma\colon Y \to X'$ by asking that $\rho|X_j = \rho_j$ and that $\sigma|Y_k = \sigma_k$ for $j \in \{1, \dots, N\}$ and $k \in \{1, \dots, M\}$. If $A \subseteq X$ denote

$$\tilde{A} = X \setminus \sigma(Y \setminus \rho(A)).$$

It is easy to verify that of $A \subseteq B \subseteq X$ that $\tilde{A} \subseteq \tilde{B}$. Now define $\mathscr{S} = \{A \subseteq X \mid A \subseteq \tilde{A}\}$. Since $\varnothing \in \mathscr{S}$, $\mathscr{S}$ is not empty. Let $S = \cup\mathscr{S}$. If $A \in \mathscr{S}$ then $A \subseteq S$ and so $\tilde{A} \subseteq \tilde{S}$. Therefore, $S \subseteq \tilde{S}$ and so $\tilde{S} \subseteq \tilde{\tilde{S}}$. Thus $\tilde{S} \in \mathscr{S}$ and so $\tilde{S} \subseteq S$. Therefore, $S = \tilde{S}$. Therefore, by definition of $\tilde{\cdot}$,

$$S = X \setminus \sigma(Y \setminus \rho(S)) \quad\Longrightarrow\quad X \setminus S = \sigma(Y \setminus \rho(S)).$$

This implies that $X \setminus S \in X'$. Now let $l \in \{1, \dots, N + M\}$ and define

$$A_l = \begin{cases} S \cap X_l, & l \in \{1, \dots, N\}, \\ \sigma_{l-N}(Y_{l-N} \setminus \rho(S)), & l \in \{N + 1, \dots, N + M\} \end{cases}$$

and

$$\tau_l = \begin{cases} \rho_l, & l \in \{1, \dots, N\}, \\ \sigma_{l-N}^{-1}, & l \in \{N + 1, \dots, N + M\}. \end{cases}$$

One then verifies that $(A_1, \dots, A_N)$ partitions S, $(A_{N+1}, \dots, A_{N+M})$ partitions $X \setminus S$, $(\tau_1(A_1), \dots, \tau_N(A_N))$ partitions $\rho(S)$, and $(\tau_{N+1}(A_{N+1}), \dots, \tau_{N+M}(A_{N+M}))$ partitions $Y \setminus \rho(S)$. This gives $X \sim Y$. ■

Next we state a lemma about piecewise congruence of identical balls with finite unions of the same sized balls.

2.5.45 Lemma *If* $\mathrm{r} \in \mathbb{R}_{>0}$ *and* $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_\mathrm{k} \in \mathbb{R}^\mathrm{n}$ *then* $\overline{\mathsf{B}}^3(\mathrm{r}, \mathbf{x}_0) \sim \cup_{\mathrm{j}=1}^{\mathrm{k}} \overline{\mathsf{B}}^3(\mathrm{r}, \mathbf{x}_\mathrm{j})$*.*

Proof We first prove the lemma in the case of $k = 2$, assuming that $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathbb{R}^3} > 2\epsilon$, i.e., assuming that $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_1)$ and $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_2)$ do not intersect. Let $(B_j)_{1\le j\le 40}$ be the partition of $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_0)$ and let $(\boldsymbol{\rho}_j)_{1\le j\le 40}$ be the isometries given by Lemma 2.5.43. Then define $\boldsymbol{\sigma}_j \in \mathsf{E}(n)$, $j \in \{1, \dots, 40\}$, by

$$\boldsymbol{\sigma}_j(\boldsymbol{x}) = \begin{cases} \boldsymbol{\rho}_j(\boldsymbol{x}) - \boldsymbol{x}_0 + \boldsymbol{x}_1, & j \in \{1, \dots, 24\}, \\ \boldsymbol{\rho}_j(\boldsymbol{x}) - \boldsymbol{x}_0 + \boldsymbol{x}_2, & j \in \{25, \dots, 40\}. \end{cases}$$

Then $(\boldsymbol{\sigma}_j(B_j))_{1\le j\le 24}$ is a partition of $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_1)$ and $(\boldsymbol{\sigma}_j(B_j))_{25\le j\le 40}$ is a partition of $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_2)$ by Lemma 2.5.43. Thus we have $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_0) \sim \overline{\mathsf{B}}^3(r, \boldsymbol{x}_1) \cup \overline{\mathsf{B}}^3(r, \boldsymbol{x}_2)$ under the stated hypotheses.

Now we prove the lemma by induction on k. The result is clear for $k = 1$: one need only translate $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_0)$ to $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_1)$ by the isometry $\boldsymbol{x} \mapsto \boldsymbol{x} - \boldsymbol{x}_0 + \boldsymbol{x}_1$. Suppose the result holds for $k = m - 1$ and consider $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$. Choose an arbitrary $\boldsymbol{x}_0' \in \mathbb{R}^n$ such that $\|\boldsymbol{x}_0 - \boldsymbol{x}_0'\|_{\mathbb{R}^3} > 2\epsilon$. By the induction hypothesis we have $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_0) \sim \cup_{j=1}^{m-1} \overline{\mathsf{B}}^3(r, \boldsymbol{x}_j)$. Note that

$$\overline{\mathsf{B}}^3(r, \boldsymbol{x}_m) \setminus \left(\cup_{j=1}^{m-1} \overline{\mathsf{B}}^3(r, \boldsymbol{x}_j)\right) \subseteq \overline{\mathsf{B}}^3(r, \boldsymbol{x}_m) \sim \overline{\mathsf{B}}^3(r, \boldsymbol{x}_0'),$$

and so

$$\overline{\mathsf{B}}^3(r, \boldsymbol{x}_m) \setminus \left(\cup_{j=1}^{m-1} \overline{\mathsf{B}}^3(r, \boldsymbol{x}_j)\right) \precsim \overline{\mathsf{B}}^3(r, \boldsymbol{x}_0').$$

From this we conclude that

$$\overline{\mathsf{B}}^3(r, \boldsymbol{x}_0) \precsim \cup_{j=1}^{m} \overline{\mathsf{B}}^3(r, \boldsymbol{x}_j) \precsim \overline{\mathsf{B}}^3(r, \boldsymbol{x}_0) \cup \overline{\mathsf{B}}^3(r, \boldsymbol{x}_0') \precsim \overline{\mathsf{B}}^3(r, \boldsymbol{x}_0)$$

from the first part of the proof. From part (vii) of Lemma 2.5.44 we deduce that $\overline{\mathsf{B}}^3(r, \boldsymbol{x}_0) \sim \cup_{j=1}^{m} \overline{\mathsf{B}}^3(r, \boldsymbol{x}_j)$ as desired. ■

Now we may conclude the proof of the Banach–Tarski Paradox. Let X and Y be as in the statement of the theorem and let $\boldsymbol{x} \in \operatorname{int}(X)$ and $\boldsymbol{y} \in \operatorname{int}(Y)$. Choose $\epsilon \in \mathbb{R}_{>0}$ such that $\overline{\mathsf{B}}^3(\epsilon, \boldsymbol{x}) \subseteq \operatorname{int}(X)$ and $\overline{\mathsf{B}}^3(\epsilon, \boldsymbol{y}) \subseteq \operatorname{int}(Y)$. Boundedness of X ensures that there exists $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathbb{R}^3$ such that $X \subseteq \cup_{j=1}^{k} \overline{\mathsf{B}}^3(\epsilon, \boldsymbol{x}_j)$. By Lemma 2.5.45 and part (vi) of Lemma 2.5.44 we have

$$\overline{\mathsf{B}}^3(\epsilon, \mathbf{0}) \precsim X \subseteq \cup_{j=1}^{k} \overline{\mathsf{B}}^3(\epsilon, \boldsymbol{x}_j) \precsim \overline{\mathsf{B}}^3(\epsilon, \mathbf{0}).$$

By part (vii) of Lemma 2.5.44 it follows that $X \sim A$. Similarly we show that $Y \sim A$. By part (iii) of Lemma 2.5.44 it follows that $X \sim Y$, which is the result.

2.5.7 Notes

The Banach–Tarski Paradox is due to none other than Banach and Tarski [1924]. The proof we give follows that of Stromberg [1979]. The number 40 used in Lemma 2.5.43 is not optimal. Indeed, Dekker and de Groot [1956] show that one can decompose a ball into five disjoint components which can then be rearranged into two balls of the same size.

Exercises

2.5.1 Let $A \subseteq \mathbb{R}^n$ be Lebesgue measurable and for $\rho \in \mathbb{R}_{>0}$ define

$$\rho A = \{\rho \boldsymbol{x} \mid \boldsymbol{x} \in A\}.$$

Show that $\lambda_n(\rho A) = \rho^n \lambda_n(A)$.

2.5.2 Show that for $\boldsymbol{x} \in \mathbb{R}^n$, the point mass $\delta_{\boldsymbol{x}} : \mathscr{B}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ is regular.

2.5.3 Show that the counting measure $\mu : \mathscr{B}(\mathbb{R}^n) \to \overline{\mathbb{R}}_{\geq 0}$ is not regular.

Section 2.6

Measurable functions

In order to define the Lebesgue integral, one first defines functions for which it is *possible* to define the Lebesgue integral. What results is a quite general class of functions, certainly general enough to capture any function one is likely to encounter in that fantastic place called "The Real World."

Our approach, as with basic measure theory, is to start with generalities, and then proceed to particular aspects of Lebesgue measurable functions.

Do I need to read this section? If you are wanting to learn about integration in general, and the Lebesgue integral in particular, then this section is essential to this. •

2.6.1 General measurable maps and functions

We begin with a rather general definition of a measurable map between measurable spaces. The reader will observe that this definition harkens one back to the
what? definition of continuity (cf.), and so can perhaps be seen as natural, provided you
what? are comfortable with the naturality of continuity as in .

2.6.1 Definition (Measurable map) Let $(X, \mathscr{A})$ and $(Y, \mathscr{B})$ be measurable spaces. A map $f: X \to Y$ is ***$(\mathscr{A}, \mathscr{B})$-measurable*** if $f^{-1}(B) \in \mathscr{A}$ for every $B \in \mathscr{B}$. The set of $(\mathscr{A}, \mathscr{B})$-measurable maps is denoted by $\mathsf{L}^{(0)}((X, \mathscr{A}); (Y, \mathscr{B}))$, or simply by $\mathsf{L}^{(0)}(X; Y)$, with the understanding that the σ-algebras $\mathscr{A}$ and $\mathscr{B}$ are implicit. •

We shall not often consider maps between general measure spaces. However, the above general definition is useful because it gives some context for the particular definitions to follow.

It is often useful to be able to check measurability of a map by using generators for the σ-algebra involved. The following result is helpful for doing this.

2.6.2 Proposition (Measurability of maps using generators for σ-algebras) *Let* X *and* Y *be sets, let* $\mathscr{S} \subseteq \mathbf{2}^{\mathrm{X}}$ *and* $\mathscr{T} \subseteq \mathbf{2}^{\mathrm{Y}}$*, and let* $\mathscr{A}_{\mathscr{S}}$ *and* $\mathscr{A}_{\mathscr{T}}$ *be the σ-algebras generated by* $\mathscr{S}$ *and* $\mathscr{T}$*, respectively. If* $\mathrm{f}: \mathrm{X} \to \mathrm{Y}$ *is a map and if*

$$\mathscr{T} \subseteq \{\mathrm{T} \subseteq \mathrm{Y} \mid \mathrm{f}^{-1}(\mathrm{T}) \in \mathscr{S}\}$$

then f *is* $(\mathscr{A}_{\mathscr{S}}, \mathscr{A}_{\mathscr{T}})$*-measurable.*

Proof Let us denote

$$\mathscr{A}' = \{T \subseteq Y \mid f^{-1}(T) \in \mathscr{A}_{\mathscr{S}}\}.$$

We claim that $\mathscr{A}'$ is a σ-algebra containing $\mathscr{T}$. To see that it is a σ-algebra, first note that $f^{-1}(Y) = X \in \mathscr{A}_{\mathscr{S}}$ and so $Y \in \mathscr{A}'$. If $T \in \mathscr{A}'$ then

$$f^{-1}(Y \setminus T) = X \setminus f^{-1}(T)$$

by Exercise I-1.3.3. Since $X \setminus f^{-1}(T) \in \mathscr{A}_{\mathscr{S}}$ by virtue of $\mathscr{A}_{\mathscr{S}}$ being a σ-algebra, it follows that $Y \setminus T \in \mathscr{A}'$. Finally, suppose that $(T_j)_{j\in\mathbb{Z}_{>0}}$ is a countable family of sets in $\mathscr{A}'$. Then, by Proposition I-1.3.5 we have

$$f^{-1}\Big(\bigcup_{j\in\mathbb{Z}_{>0}} T_j\Big) = \bigcup_{j\in\mathbb{Z}_{>0}} f^{-1}(T_j) \in \mathscr{A}_{\mathscr{S}}$$

since $\mathscr{A}_{\mathscr{S}}$ is a σ-algebra. We thus conclude that $\cup_{j\in\mathbb{Z}_{>0}} T_j \in \mathscr{A}'$. This shows that $\mathscr{A}'$ is a σ-algebra. By hypothesis, if

$$B \in \mathscr{T} \subseteq \{T \subseteq Y \mid f^{-1}(T) \in \mathscr{S}\}$$

then $f^{-1}(B) \in \mathscr{S} \subseteq \mathscr{A}_{\mathscr{S}}$. Thus $\mathscr{T} \subseteq \mathscr{A}'$ and so $\mathscr{A}_{\mathscr{T}} \subseteq \mathscr{A}'$ since $\mathscr{A}_{\mathscr{T}}$ is the smallest σ-algebra containing $\mathscr{T}$. It therefore follows that if $B \in \mathscr{A}_{\mathscr{T}}$ then $f^{-1}(B) \in \mathscr{A}_{\mathscr{S}}$, i.e., that f is $(\mathscr{A}_{\mathscr{S}}, \mathscr{A}_{\mathscr{T}})$-measurable. ■

We can give an application of the preceding result that gives an important class of measurable maps.

2.6.3 Example (Continuous maps are Borel-measurable) We claim that if $\boldsymbol{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ is continuous then it is $(\mathscr{B}(\mathbb{R}^n), \mathscr{B}(\mathbb{R}^m))$-measurable. Indeed, $\mathscr{B}(\mathbb{R}^n)$ and $\mathscr{B}(\mathbb{R}^m)$ are the σ-algebras generated by the collections $\mathscr{O}(\mathbb{R}^n)$ and $\mathscr{O}(\mathbb{R}^m)$ of open subsets of $\mathbb{R}^n$ and $\mathbb{R}^m$. Since $\boldsymbol{f}$ is continuous it follows from Corollary II-1.3.4 that

$$\mathscr{O}(\mathbb{R}^m) \subseteq \{U \subseteq \mathbb{R}^m \mid \boldsymbol{f}^{-1}(U) \in \mathscr{O}(\mathbb{R}^n)\}.$$

From Proposition 2.6.2 we conclude that $\boldsymbol{f}$ is $(\mathscr{B}(\mathbb{R}^n), \mathscr{B}(\mathbb{R}^m))$-measurable. •

What we are really interested in in this section are $\mathbb{R}$-valued functions. It turns out to be interesting to consider $\overline{\mathbb{R}}$-valued functions. The reason for this degree of generality is not that we are interested in infinite-valued functions *per se*, but that we are interested in sequences of $\mathbb{R}$-valued functions that turn out to have infinite limits. The reader will want to be familiar with the order relations on $\overline{\mathbb{R}}$ defined in Section I-2.2.5.

In any case, we now turn our attention to functions $f\colon X \to \overline{\mathbb{R}}$ defined on a measurable space $(X, \mathscr{A})$. For such functions we have the following equivalent properties.

2.6.4 Proposition (Characterisations of measurable functions) *For a measurable space* $(X, \mathscr{A})$ *and a map* $f\colon X \to [-\infty, \infty]$*, the following statements are equivalent:*

(i) *for each* $b \in \mathbb{R}$ *the set* $f^{-1}([-\infty, b]) = \{x \in X \mid f(x) \le b\}$ *is measurable;*

(ii) *for each* $b \in \mathbb{R}$ *the set* $f^{-1}([-\infty, b)) = \{x \in X \mid f(x) < b\}$ *is measurable;*

(iii) *for each* $a \in \mathbb{R}$ *the set* $f^{-1}([a, \infty]) = \{x \in X \mid f(x) \ge a\}$ *is measurable;*

(iv) *for each* $a \in \mathbb{R}$ *the set* $f^{-1}((a, \infty]) = \{x \in X \mid f(x) > a\}$ *is measurable.*

Proof (i) $\implies$ (ii) We write

$$f^{-1}([-\infty,b)) = f^{-1}(\cup_{k\in\mathbb{Z}_{>0}} f^{-1}([-\infty,b-\tfrac{1}{k}]) = \cup_{k\in\mathbb{Z}_{>0}} f^{-1}([-\infty,b-\tfrac{1}{k}])$$

by Proposition I-1.3.5. Since $f^{-1}([-\infty,b-\frac{1}{k}] \in \mathscr{A}$ by assumption and since $\mathscr{A}$ is a σ-algebra, we conclude that $f^{-1}([-\infty,b)) \in \mathscr{A}$.

(ii) $\implies$ (iii) Here we note that

$$f^{-1}([a,\infty]) = X \setminus f^{-1}([-\infty,a))$$

by Exercise I-1.3.3. Since $\mathscr{A}$ is a σ-algebra and since $f^{-1}([-\infty,a)) \in \mathscr{A}$ by assumption, it follows that $f^{-1}([a,\infty]) \in \mathscr{A}$.

(iii) $\implies$ (iv) Here we write

$$f^{-1}((a,\infty]) = \cup_{k\in\mathbb{Z}_{>0}} f^{-1}([a+\tfrac{1}{k},\infty]) =$$

by Proposition I-1.3.5. As in the first part of the proof we conclude that $f^{-1}((a,\infty]) \in \mathscr{A}$.

(iv) $\implies$ (i) Here we note that

$$f^{-1}([-\infty,b]) = X \setminus f^{-1}((b,\infty])$$

by Exercise I-1.3.3 and then argue as in the second part of the proof that $f^{-1}([-\infty,b]) \in \mathscr{A}$. ■

With this result at hand, the following definition makes sense.

2.6.5 Definition (Measurable function) For a measurable space $(X,\mathscr{A})$ a function $f\colon X\to\overline{\mathbb{R}}$ satisfying any one of the four equivalent conditions of Proposition 2.6.4 is an ***$\mathscr{A}$-measurable*** function. We shall frequently just say that f is ***measurable*** if $\mathscr{A}$ is understood. For any subset $I\subseteq\overline{\mathbb{R}}$ (typically we will be concerned with $I\in\{\mathbb{R},\overline{\mathbb{R}}_{\geq 0}\}$) we denote the set of measurable I-valued maps by $\mathsf{L}^{(0)}((X,\mathscr{A});I)$, or by $\mathsf{L}^{(0)}(X;I)$, with the understanding that the σ-algebra $\mathscr{A}$ is implicit. •

The relationship of this notion of measurability with that of Definition 2.6.1 is perhaps not immediately clear. So let us make this clear, recalling from Definition 2.4.15 the definition of the σ-algebra $\mathscr{B}(\overline{\mathbb{R}})$ on $\overline{\mathbb{R}}$.

2.6.6 Proposition (Characterisation of measurable functions) *For a measurable space $(X,\mathscr{A})$ and a map $f\colon X\to\overline{\mathbb{R}}$, the following statements are equivalent:*

(i) $f\in\mathsf{L}^{(0)}((X,\mathscr{A});\overline{\mathbb{R}})$*;*

(ii) the sets $\{x\in X \mid f(x)=-\infty\}$ *and* $\{x\in X \mid f(x)=\infty\}$ *are measurable and* $f^{-1}(B)\in\mathscr{A}$ *for every* $B\in\mathscr{B}(\mathbb{R})$*;*

(iii) f *is* $(\mathscr{A},\mathscr{B}(\overline{\mathbb{R}}))$*-measurable.*

Proof (i) $\implies$ (ii) We have

$$f^{-1}(-\infty) = f^{-1}(\cap_{k\in\mathbb{Z}_{>0}}[-\infty,-k]) = \cap_{k\in\mathbb{Z}} f^{-1}([-\infty,-k]),$$
$$f^{-1}(\infty) = f^{-1}(\cap_{k\in\mathbb{Z}_{>0}}[k,\infty]) = \cap_{k\in\mathbb{Z}} f^{-1}([k,\infty]),$$

by Proposition I-1.3.5. Thus $f^{-1}(-\infty)$ and $f^{-1}(\infty)$ are countable intersections of measurable sets and so themselves measurable. We must also show that $f^{-1}(B)$ is measurable for a Borel set B. To prove this, we denote

$$\mathscr{B}'(\mathbb{R}) = \{S \subseteq \mathbb{R} \mid f^{-1}(S) \in \mathscr{A}\}.$$

We claim that $\mathscr{B}'(\mathbb{R})$ is a σ-algebra containing $\mathscr{B}(\mathbb{R})$. Certainly $\mathbb{R} \in \mathscr{B}'(\mathbb{R})$ since $f^{-1}(\mathbb{R}) = X \in \mathscr{A}$. If $(S_j)_{j\in\mathbb{Z}_{>0}}$ is a countable collection of subsets from $\mathscr{B}'(\mathbb{R})$ we have

$$f^{-1}\Big(\bigcup_{j\in\mathbb{Z}_{>0}} S_j\Big) = \bigcup_{j\in\mathbb{Z}_{>0}} f^{-1}(S_j) \in \mathscr{A},$$

where we have used Proposition I-1.3.5. Thus $\cup_{j\in\mathbb{Z}_{>0}} S_j \in \mathscr{B}'(\mathbb{R})$. Also, by Exercise I-1.3.3, if $S \in \mathscr{B}'(\mathbb{R})$ then

$$f^{-1}(\mathbb{R} \setminus S) = X \setminus f^{-1}(S) \in \mathscr{A}$$

and so $\mathbb{R} \setminus S \in \mathscr{B}'(\mathbb{R})$. Thus $\mathscr{B}'(\mathbb{R})$ is a σ-algebra. By hypothesis we have $(-\infty, b] \in \mathscr{B}'(\mathbb{R})$ for every $b \in \mathbb{R}$. Thus $\mathscr{B}'(\mathbb{R})$ contains the σ-algebra generated by sets of the form $(-\infty, b]$ for $b \in \mathbb{R}$. By Proposition 2.4.9 this means that $\mathscr{B}(\mathbb{R}) \subseteq \mathscr{B}'(\mathbb{R})$, as claimed. This proves that $f^{-1}(B) \in \mathscr{A}$ for $B \in \mathscr{B}(\mathbb{R})$.

(ii) $\implies$ (iii) Let

$$\mathscr{B}'(\overline{\mathbb{R}}) = \{T \subseteq \overline{\mathbb{R}} \mid f^{-1}(T) \in \mathscr{A}\},$$

and note that, by hypothesis, $\mathscr{B}(\mathbb{R}) \cup \{-\infty\} \cup \{\infty\} \subseteq \mathscr{B}'(\overline{\mathbb{R}})$. By Proposition 2.4.16 it follows that f is $(\mathscr{A}, \mathscr{B}(\overline{\mathbb{R}}))$-measurable.

(iii) $\implies$ (i) For $a \in \mathbb{R}$ we have

$$f^{-1}((a,\infty]) = f^{-1}((a,\infty) \cup \{\infty\}) = f^{-1}((a,\infty)) \cup f^{-1}(\{\infty\})$$

by Proposition I-1.3.5. Since (a,∞) is open it is a Borel set and so in $\mathscr{B}(\overline{\mathbb{R}})$ by Proposition 2.4.16. Thus $f^{-1}((a,\infty)) \in \mathscr{A}$ by hypothesis. Also, $\{\infty\} \in \mathscr{B}(\overline{\mathbb{R}})$ by Proposition 2.4.16 and so $f^{-1}(\{\infty\}) \in \mathscr{A}$. Therefore, $f^{-1}((a,\infty])$ is a union of measurable sets and so is measurable. Thus f is $\mathscr{A}$-measurable. ■

For functions that are $\mathbb{R}$-valued this gives the following result.

2.6.7 Corollary (Measurability of $\mathbb{R}$-valued functions) *For a measurable space $(X, \mathscr{A})$, a function $f\colon X \to \mathbb{R}$ is measurable if and only if it is $(\mathscr{A}, \mathscr{B}(\mathbb{R}))$-measurable.*

It is often fairly easy to apply Definition 2.6.5 to ascertain whether a given function is measurable (as opposed to employing the equivalent characterisation of Proposition 2.6.6).

2.6.8 Examples (Measurable functions)

1. For a measurable space $(X, \mathscr{A})$ and for $\alpha \in \overline{\mathbb{R}}$, we claim that the constant function $f_\alpha\colon x \mapsto \alpha$ is $\mathscr{A}$-measurable. To see this we let $b \in \mathbb{R}$ and determine that
$$f_\alpha^{-1}([-\infty, b)) = \begin{cases} \varnothing, & b \le \alpha, \\ X, & b > \alpha, \end{cases}$$
provided that $\alpha \neq -\infty$. If $\alpha = -\infty$ then $f_\alpha^{-1}([-\infty, b)) = X$ for every $b \in \mathbb{R}$. In any case, $f_\alpha^{-1}([-\infty, b)) \in \mathscr{A}$ for all $b \in \mathbb{R}$ and so f_α is $\mathscr{A}$-measurable.
2. Let $(X, \mathscr{A})$ be a measurable space and let $A \in \mathscr{A}$. We claim that the characteristic function $\chi_A\colon X \to \mathbb{R}$ is $\mathscr{A}$-measurable. Indeed,
$$\chi_A^{-1}([a, \infty]) = \begin{cases} X, & a \le 0, \\ A, & a \in (0, 1], \\ \varnothing, & a > 1. \end{cases}$$
Since $X, A, \varnothing \in \mathscr{A}$ it follows that χ_A is indeed $\mathscr{A}$-measurable.
Note that the same argument shows that, if $A \notin \mathscr{A}$, then χ_A is not $\mathscr{A}$-measurable.
3. Let $A \in \mathscr{L}(\mathbb{R}^n)$ and let $f\colon A \to \mathbb{R}$ be continuous. We claim that f is $\mathscr{L}(\mathbb{R}^n)$-measurable. Indeed, for $a \in \mathbb{R}$ the set $f^{-1}((a, \infty))$ is open in A by Corollary II-1.3.4. Thus there exists an open subset $U_a \subseteq \mathbb{R}^n$ such that $f^{-1}((a, \infty)) = U_a \cap A$. Since $U_a \in \mathscr{L}(\mathbb{R}^n)$ (open sets are Borel sets and so are Lebesgue measurable) we have $f^{-1}((a, \infty)) \in \mathscr{L}(\mathbb{R}^n)$ and so is a measurable subset of A. •

Let $(X, \mathscr{A})$ be a measurable space. By Corollary 2.6.7, measurability of $f\colon X \to \mathbb{R}$ is equivalent to $(\mathscr{A}, \mathscr{B}(\mathbb{R}))$-measurability of f. A natural question to ask is: "Why use the σ-algebra of *Borel* sets on $\mathbb{R}$ to define measurability of a function? Why not use the σ-algebra of *Lebesgue* measurable sets?" The answer to this question perhaps cannot be divined immediately. The reason for using the Borel measurable sets is answered by answering the question, "What is it we are trying to achieve with our definition of a measurable function?" We shall not address this here, but instead refer ahead to Section 2.6.5. For now, let us simply illustrate that $(\mathscr{A}, \mathscr{B}(\mathbb{R}))$-measurability and $(\mathscr{A}, \mathscr{L}(\mathbb{R}))$-measurability are not equivalent.

more to say here?

2.6.9 Example ($(\mathscr{A}, \mathscr{B}(\mathbb{R}))$- and $(\mathscr{A}, \mathscr{L}(\mathbb{R}))$-measurability are different) Since $\mathscr{B}(\mathbb{R}) \subseteq \mathscr{L}(\mathbb{R})$ it follows that $f\colon X \to \mathbb{R}$ is $(\mathscr{A}, \mathscr{B}(\mathbb{R}))$-measurable if it is $(\mathscr{A}, \mathscr{L}(\mathbb{R}))$-measurable. The converse implication is not generally true, however. We illustrate this with an example. We take $X = [0, 1]$ and $\mathscr{A} = \mathscr{L}([0, 1])$. We define a function $f\colon [0, 1] \to \mathbb{R}$ that is $(\mathscr{L}([0, 1]), \mathscr{B}(\mathbb{R}))$-measurable but not $(\mathscr{L}([0, 1]), \mathscr{L}(\mathbb{R}))$-measurable. Our construction relies on the reader understanding the construction of the sets C_ϵ and C from Examples I-2.5.42 and I-2.5.39.

Let $\epsilon \in \mathbb{R}_{>0}$ and let $C_\epsilon \subseteq [0, 1]$ be the "fat" Cantor set of Example I-2.5.42. Let $C \subseteq [0, 1]$ be the standard middle-thirds Cantor set of Example I-2.5.39. Recall that the inductive construction of these sets is the same in that they are defined by, at step k, removing 2^k open intervals from the set defined at step $k-1$. This defines countable collections $(I_{\epsilon,k})_{k\in\mathbb{Z}_{>0}}$ and $(I_k)_{k\in\mathbb{Z}_{>0}}$ of disjoint open intervals such that

$$C_\epsilon = [0,1] \setminus \cup_{j\in\mathbb{Z}_{>0}} I_{\epsilon,j}, \quad C = [0,1] \setminus \cup_{j\in\mathbb{Z}_{>0}} I_j.$$

Moreover, since the constructions of C_ϵ and C proceed in the same way, the intervals $(I_{\epsilon,j})_{j\in\mathbb{Z}_{>0}}$ and $(I_j)_{j\in\mathbb{Z}_{>0}}$ can be enumerated consistently such $I_{\epsilon,1}$ and I_1 are the intervals removed in the first step in the inductive constructions of C_ϵ and C, $I_{\epsilon,2}$ and $I_{\epsilon,3}$, and I_2 and I_3 are the intervals, ordered from left to right, removed in the second step in the inductive constructions of C_ϵ and C, and so on. We then define $f\colon [0,1] \to \mathbb{R}$ by asking that $f|I_{\epsilon,j}$ maps $I_{\epsilon,j}$ linearly onto the interval I_j, mapping the left (resp. right) endpoint of $I_{\epsilon,j}$ to the left (resp. right) endpoint of I_j. Note that since $\mathrm{cl}([0,1] \setminus C_\epsilon) = [0,1]$, it follows that this definition of f on $[0,1] \setminus C_\epsilon$ extends to a continuous function f from $[0,1]$ to $\mathbb{R}$. By Example 2.6.8–3 it follows that f is $(\mathscr{L}([0,1]),\mathscr{B}(\mathbb{R}))$-measurable. Moreover, $f(C_\epsilon) = C$ since, by construction, the points in C_ϵ and C are the endpoints of intervals from $(I_{\epsilon,j})_{j\in\mathbb{Z}_{>0}}$ and $(I_j)_{j\in\mathbb{Z}_{>0}}$. Moreover, we claim that f is strictly monotonically increasing. It is obviously monotonically increasing. To see that it is strictly monotonically increasing, suppose that $x_1, x_2 \in [0,1]$ satisfy $x_1 < x_2$ and $f(x_1) = f(x_2)$. This means that $f|[x_1,x_2]$ is constant which, by construction of f implies that $[x_1,x_2] \subseteq C_\epsilon$, contradicting the fact that $\mathrm{int}(C_\epsilon) = \varnothing$. Thus f is strictly monotonically increasing and so injective by Theorem I-3.1.30. By Theorem 2.4.5 there exists a subset $S \subseteq C_\epsilon$ that is not Lebesgue measurable. Let $T = f(S) \subseteq C$ so that $T \in \mathscr{L}(\mathbb{R})$ since $\lambda(C) = 0$ and since $\mathscr{L}(\mathbb{R})$ is complete. Injectivity of f implies that $f^{-1}(T) = S \notin \mathscr{L}([0,1])$. Thus f is not $(\mathscr{L}([0,1]),\mathscr{L}(\mathbb{R}))$-measurable. •

It is often the case that one is able to draw conclusions about a function only almost everywhere, not everywhere. In such cases, one would like to assert that this almost everywhere knowledge of the function is enough to ensure its measurability. It should not be surprising that completeness plays a rôle here. Note that this is the first time we have used a measure in our discussion of measurable functions. Up to now we have only used measurable spaces.

2.6.10 Proposition (Measurability of almost everywhere known functions) *If* $(\mathsf{X},\mathscr{A},\mu)$ *is a complete measure space, if* $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}})$ *is* $\mathscr{A}$*-measurable, and if* $\mathsf{g}\colon \mathsf{X} \to \overline{\mathbb{R}}$ *satisfies*

$$\mu(\{\mathsf{x} \in \mathsf{X} \mid \mathsf{f}(\mathsf{x}) \neq \mathsf{g}(\mathsf{x})\}) = 0,$$

then $\mathsf{g} \in \mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}})$.

Proof Let

$$A_{f,g} = \{x \in A \mid f(x) = g(x)\}$$

and let $b \in \mathbb{R}$. Then

$$\{x \in X \mid g(x) \le b\} = (\{x \in X \mid f(x) \le b\} \cap A_{f,g}) \cup (\{x \in X \mid g(x) \le b\} \cap (X \setminus A_{f,g})).$$

The set $X \setminus A_{f,g}$ has measure zero and so is measurable. Thus $A_{f,g}$ is measurable and so the set

$$\{x \in X \mid f(x) \leq b\} \cap A_{f,g}$$

is measurable. Since the set

$$\{x \in X \mid g(x) \leq b\} \cap (X \setminus A_{f,g})$$

is a subset of the set $A_{f,g}$ which has measure zero, completeness of $(X, \mathscr{A}, \mu)$ ensures that it has measure zero, and in particular is measurable. Thus

$$\{x \in X \mid g(x) \leq b\}$$

is the intersection of measurable sets, and so is measurable. ■

2.6.2 Measurability and operations on functions

At this point we are still not clear on the significance of measurable functions, and we will continue to postpone this until Section 2.6.5. All we really know at the moment is that the set of measurable functions on $(\mathbb{R}^n, \mathscr{L}(\mathbb{R}^n))$ contains the continuous functions, and so there is a nice subset of measurable functions in this case. It turns out that measurable functions also have nice properties with respect to the natural operations one performs on functions and sequences of functions. In this section we prove these properties.

We begin with the interaction of measurable functions with standard algebraic operations. In order to do this, the reader will wish to recall from Section I-2.2.5 the "algebraic" operations on $\overline{\mathbb{R}}$. This is complicated a little for measurable functions since these are $\overline{\mathbb{R}}$-valued. To properly state the result we need, it is, therefore, convenient to introduce some notation to account for the fact that certain algebraic operations are ill-defined on $\overline{\mathbb{R}}$. If X is a set, if $f: X \to \overline{\mathbb{R}}$, and if $\alpha_-, \alpha_+ \in \overline{\mathbb{R}}$, then we denote by $f_{\alpha_-,\alpha_+}: X \to \overline{\mathbb{R}}$ the function given by

$$f_{\alpha_-,\alpha_+}(x) = \begin{cases} f(x), & f(x) \in \mathbb{R}, \\ \alpha_-, & f(x) = -\infty, \\ \alpha_+, & f(x) = \infty. \end{cases}$$

Similarly, for $\alpha_-, \alpha_+, \alpha_0 \in \mathbb{R}$ we denote by $f_{\alpha_-,\alpha_+,\alpha_0}: X \to \mathbb{R}$ the function given by

$$f_{\alpha_-,\alpha_+,\alpha_0}(x) = \begin{cases} \alpha_-, & f(x) = -\infty, \\ \alpha_+, & f(x) = \infty, \\ \alpha_0, & f(x) = 0, \\ f(x), & \text{otherwise}. \end{cases}$$

Next, for $f, g \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ and for $\alpha \in \overline{\mathbb{R}}$ denote $f +_\alpha g: X \to \overline{\mathbb{R}}$ the function defined by

$$(f +_\alpha g)(x) = \begin{cases} \alpha, & f(x) = \infty,\ g(x) = -\infty \text{ or } f(x) = -\infty,\ g(x) = \infty, \\ f(x) + g(x), & \text{otherwise}. \end{cases}$$

With these tedious bits of notation out of the way, we can now state the desired result.

2.6.11 **Proposition (Algebraic operations on measurable functions)** *Let* $(\mathsf{X}, \mathscr{A})$ *be a measurable space, let* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}})$*, let* $\beta \in \mathbb{R}$*, let* $\beta_-, \beta_+, \beta_0 \in \mathbb{R}^*$*, let* $\alpha, \alpha_-, \alpha_+ \in \overline{\mathbb{R}}$*, let* $\mathsf{p} \in \mathbb{R}_{>0}$*, and let* $\mathsf{k} \in \mathbb{Z}_{>0}$*. Then the following functions are* $\mathscr{A}$*-measurable:*

(i) $\beta\mathsf{f}$;

(ii) $\mathsf{f} +_\alpha \mathsf{g}$;

(iii) fg;

(iv) $\dfrac{\mathsf{f}}{\mathsf{g}_{\beta_-,\beta_+,\beta_0}}$;

(v) $(|\mathsf{f}|^{\mathsf{p}})_{\alpha_-,\alpha_+}$;

(vi) $(\mathsf{f}^{\mathsf{k}})_{\alpha_-,\alpha_+}$.

Proof We shall freely make use of Proposition 2.6.13 below.

(i) Let $\phi_\beta \colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ be defined by $\phi_\beta(y) = \beta y$. Then $\beta f = \phi_\beta \circ f$. Since

$$\phi_\beta^{-1}(U) = \{\beta y \mid y \in U\}$$

it follows that $\phi_\beta^{-1}(U)$ is open for open set U. Also,

$$\phi_\beta^{-1}([-\infty, b)) = \begin{cases} [-\infty, \beta b), & \beta \in \mathbb{R}_{>0}, \\ \{0\}, & \beta = 0, \\ (\beta b, \infty], & \beta \in \mathbb{R}_{<0}, \end{cases}$$

and so $\phi_\beta^{-1}([-\infty, b)) \in \mathscr{B}(\overline{\mathbb{R}})$ for every $\beta, b \in \mathbb{R}$. Similarly, $\phi_\beta^{-1}((a, \infty]) \in \mathscr{B}(\overline{\mathbb{R}})$ for every $\beta, a \in \mathbb{R}$. From this we deduce, using Proposition I-1.3.5, that the preimage by ϕ_β of the generators of the σ-algebra $\mathscr{B}(\overline{\mathbb{R}})$ are in $\mathscr{B}(\overline{\mathbb{R}})$. By Proposition 2.6.2 we conclude that ϕ_β is $\mathscr{B}(\overline{\mathbb{R}})$-measurable. By Proposition 2.6.13 we then conclude that βf is $\mathscr{A}$-measurable.

(ii) Here we use a pair of fairly simple lemmata.

1 **Lemma** *For a measurable space* $(\mathsf{X}, \mathscr{A})$ *and* $\mathscr{A}$*-measurable functions* $\mathsf{f}, \mathsf{g} \colon \mathsf{X} \to \overline{\mathbb{R}}$*, the following sets are measurable:*

(i) $\{\mathsf{x} \in \mathsf{X} \mid \mathsf{f}(\mathsf{x}) > \mathsf{g}(\mathsf{x})\}$;

(ii) $\{\mathsf{x} \in \mathsf{X} \mid \mathsf{f}(\mathsf{x}) \geq \mathsf{g}(\mathsf{x})\}$;

(iii) $\{\mathsf{x} \in \mathsf{X} \mid \mathsf{f}(\mathsf{x}) = \mathsf{g}(\mathsf{x})\}$.

Proof (i) We claim that

$$\{x \in X \mid f(x) > g(x)\} = \bigcup_{q \in \mathbb{Q}} \Big(\{x \in X \mid f(x) > q\} \cap \{x \in X \mid g(x) < q\}\Big).$$

Indeed, let $x \in \{x' \in X \mid f(x') > g(x')\}$. If $f(x) = \infty$ then $g(x) < \infty$. Thus there exists $q \in \mathbb{Q}$ such that $f(x) > q$ and $g(x) < q$. If $f(x) < \infty$ then $f(x) \in \mathbb{R}$ since we cannot have $f(x) = -\infty$. Therefore, there exists $q \in \mathbb{Q}$ such that $f(x) > q$ and $g(x) < q$. This shows that

$$\{x \in X \mid f(x) > g(x)\} \subseteq \bigcup_{q \in \mathbb{Q}} \Big(\{x \in X \mid f(x) > q\} \cap \{x \in X \mid g(x) < q\}\Big).$$

For the converse inclusion, suppose that $x \in X$ has the property that there exists $q \in \mathbb{Q}$ such that $g(x) < q < f(x)$. Clearly $x \in \{x' \in X \mid f(x') > g(x')\}$, giving our claim.

Now, since f and g are $\mathscr{A}$-measurable, the sets

$$\{x \in X \mid f(x) > q\}, \quad \{x \in X \mid g(x) < q\}, \qquad q \in \mathbb{Q},$$

are measurable, and so too then is their intersection. Thus $\{x \in X \mid f(x) > g(x)\}$ is a countable union of measurable sets, which is then measurable.

(ii) Note that

$$\{x \in X \mid f(x) \geq g(x)\} = X \setminus \{x \in X \mid g(x) > f(x)\}.$$

Since $\{x \in X \mid g(x) > f(x)\}$ is measurable by the first part of the lemma it follows that $\{x \in X \mid f(x) \geq g(x)\}$ is also measurable.

(iii) We have

$$\{x \in X \mid f(x) = g(x)\} = \{x \in X \mid f(x) \geq g(x)\} \cap \{x \in X \mid g(x) \geq f(x)\}.$$

The right-hand side is the intersection of two measurable sets by the second part of the lemma, and so is measurable. ▼

2 Lemma *If* $(X, \mathscr{A})$ *is a measurable space, if* $f\colon X \to \overline{\mathbb{R}}$ *is* $\mathscr{A}$*-measurable, and if* $\beta \in \mathbb{R}$*, then the function* $x \mapsto f(x) + \beta$ *is* $\mathscr{A}$*-measurable.*

Proof Define $\phi_\beta\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ by $\phi_\beta(y) = y + \beta$. By Proposition 2.4.22 it follows that $\phi_\beta^{-1}(B) \in \mathscr{B}(\mathbb{R})$ for $B \in \mathscr{B}(\mathbb{R})$. It is clear that $\phi_\beta^{-1}(\{-\infty\}) = \{-\infty\}$ and that $\phi_\beta^{-1}(\{\infty\}) = \{\infty\}$. Therefore, by Propositions 2.4.16 and 2.6.2, it follows that ϕ_β is $\mathscr{B}(\overline{\mathbb{R}})$-measurable. Thus $\phi_\beta \circ f$ is $\mathscr{A}$-measurable by Proposition 2.6.13. ▼

To proceed with the proof, let $a \in \mathbb{R}$ and let

$$\begin{aligned} A_{a,\alpha} = (\{x \in X \mid f(x) = \infty\} \cap \{x \in X \mid g(x) = -\infty\}) \\ \cup (\{x \in X \mid f(x) = -\infty\} \cap \{x \in X \mid g(x) = \infty\}) \end{aligned}$$

if $a < \alpha$ and let $A_{a,\alpha} = \emptyset$ if $a \geq \alpha$. We then have

$$\begin{aligned} (f +_\alpha g)^{-1}((a, \infty]) &= \{x \in X \mid f(x) + g(x) > a\} \cup A_{a,\alpha} \\ &= \{x \in X \mid f(x) > a - g(x)\} \cup A_{a,\alpha}. \end{aligned}$$

By the two lemmata above, the set $\{x \in X \mid f(x) > a - g(x)\}$ is measurable. By Proposition 2.6.6 each of the four sets comprising the definition of $A_{a,\alpha}$ when $a < \alpha$ is measurable. Thus $A_{a,\alpha}$ is measurable and so $(f +_\alpha g)^{-1}((a, \infty])$ is measurable, being a union of measurable sets.

(iii) We denote

$$\begin{aligned} A_{f,-} &= \{x \in X \mid f(x) = -\infty\}, \quad A_{f,+} = \{x \in X \mid f(x) = \infty\}, \\ A_{g,-} &= \{x \in X \mid g(x) = -\infty\}, \quad A_{g,+} = \{x \in X \mid g(x) = \infty\}. \end{aligned}$$

By Proposition 2.6.4 these sets are measurable. For $x \notin A_{f,-} \cup A_{f,+} \cup A_{g,-} \cup A_{g,+}$ we have

$$f(x)g(x) = \tfrac{1}{2}((f(x) + g(x))^2 - f(x)^2 - g(x)^2).$$

If $x \in A_{f,-} \cap A_{g,+}$ or $x \in A_{f,+} \cap A_{g,-}$ then $f(x)g(x) = -\infty$ and if $x \in A_{f,-} \cap A_{g,-}$ or $x \in A_{f,+} \cap A_{g,+}$ then $f(x)g(x) = \infty$. Then, for $a \in \mathbb{R}$ we have

$$\begin{aligned}(fg)^{-1}((a,\infty]) &= \{x \in X \mid f(x)g(x) > a\}\\ &= \{x \in (A_{f,-} \cap A_{g,+}) \cup (A_{f,+} \cap A_{g,-}) \mid f(x)g(x) > a\}\\ &\quad \cup \{x \in (A_{f,-} \cap A_{g,-}) \cup (A_{f,+} \cap A_{g,+}) \mid f(x)g(x) > a\}\\ &\quad \cup \{x \in X \setminus (A_{f,-} \cup A_{f,+} \cup A_{g,-} \cup A_{g,+}) \mid\\ &\quad \tfrac{1}{2}((f(x) + g(x))^2 - f(x)^2 - g(x)^2) > a\}.\end{aligned}$$

The set

$$\{x \in (A_{f,-} \cap A_{g,+}) \cup (A_{f,+} \cap A_{g,-}) \mid f(x)g(x) > a\}$$

is empty, the set

$$\{x \in (A_{f,-} \cap A_{g,-}) \cup (A_{f,+} \cap A_{g,+}) \mid f(x)g(x) > a\}$$

is measurable being a union of measurable sets, and the set

$$\{x \in X \setminus (A_{f,-} \cup A_{f,+} \cup A_{g,-} \cup A_{g,+}) \mid \tfrac{1}{2}((f(x) + g(x))^2 - f(x)^2 - g(x)^2) > a\}$$

is measurable by parts (ii) and (vi). Thus $(fg)^{-1}((a,\infty])$ is a union of three measurable sets and so measurable.

(iv) We first consider the case when $f(x) = 1$ for every $x \in X$. In this case let us define $\phi_{\beta_-,\beta_+,\beta_0}\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ by

$$\phi_{\beta_-,\beta_+,\beta_0}(y) = \begin{cases} \frac{1}{y}, & y \in \mathbb{R},\\ \frac{1}{\beta_0}, & y = 0,\\ \frac{1}{\beta_-}, & y = -\infty,\\ \frac{1}{\beta_+}, & y = \infty. \end{cases}$$

Note that $y \mapsto \frac{1}{y}$ is $(\mathscr{B}(\mathbb{R}), \mathscr{B}(\mathbb{R}))$-measurable by Example 2.6.3. Therefore, by Propositions 2.4.16 and 2.6.2 it is easy to see that $\phi_{\beta_-,\beta_+,\beta_0}$ is $\mathscr{B}(\overline{\mathbb{R}})$-measurable. Since $\frac{1}{g_{\beta_-,\beta_+,\beta_0}} = \phi_{\beta_-,\beta_+,\beta_0} \circ g$ this part of the result follows from Proposition 2.6.13 in the case that $f = 1$. For general f the result follows from the result for $f = 1$ and from part (iii).

(v) Here we define $\phi_{\alpha_-,\alpha_+}\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ by

$$\phi_{\alpha_-,\alpha_+}(y) = \begin{cases} |y|^p, & y \in \mathbb{R},\\ \alpha_-, & y = -\infty,\\ \alpha_+, & y = \infty. \end{cases}$$

It is easy to verify by Propositions 2.4.16 and 2.6.2 that ϕ_{α_-,α_+} is $\mathscr{B}(\overline{\mathbb{R}})$-measurable. Since the function $y \mapsto |y|^p$ is continuous and so $(\mathscr{B}(\mathbb{R}), \mathscr{B}(\mathbb{R}))$-measurable by Example 2.6.3. Thus, since $(|f|^p)_{\alpha_-,\alpha_+} = \phi_{\alpha_-,\alpha_+} \circ f$, this part of the result follows from Proposition 2.6.13.

(vi) If we define $\phi_{\alpha_-,\alpha_+}\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ as in the proof of part (v), this part of the proof is carried out exactly as that for part (v). ■

We now consider the interaction of composition and measurability. First of all, the most general result is false as the following example shows.

2.6.12 Example (Compositions of measurable functions may not be measurable) We recall from Exercise 2.6.9 the construction of a map $f\colon [0,1] \to [0,1]$ that is $(\mathscr{L}([0,1]),\mathscr{B}(\mathbb{R}))$-measurable but not $(\mathscr{L}([0,1]),\mathscr{L}(\mathbb{R}))$-measurable. Let $S,T \subseteq [0,1]$ be the subsets constructed in Exercise 2.6.9 and let $\chi_T\colon [0,1] \to \mathbb{R}$ be the characteristic function. Since T is Lebesgue measurable, as we showed in Exercise 2.6.9, it follows from Example 2.6.8–2 that χ_T is $\mathscr{L}([0,1])$-measurable. However, by construction of f, $\chi_T \circ f = \chi_S$. Since $S \notin \mathscr{L}([0,1])$ by construction, it follows from Example 2.6.8–2 that χ_S is not $\mathscr{L}([0,1])$-measurable. Thus the composition of measurable functions need not be measurable. •

The preceding counterexample notwithstanding, there is a useful result concerning measurability of compositions. The result relies on the notion of the σ-algebra $\mathscr{B}(\overline{\mathbb{R}})$ on $\overline{\mathbb{R}}$ as defined in Definition 2.4.15.

2.6.13 Proposition (Composition and measurable functions) *Let $(\mathsf{X},\mathscr{A})$ be a measurable space, let $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}})$, and let $\phi\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ be $\mathscr{B}(\overline{\mathbb{R}})$-measurable. Then $\phi \circ \mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}})$.*

Proof Let $B \in \mathscr{B}(\overline{\mathbb{R}})$. By assumption and by Proposition 2.6.6 we have $\phi^{-1}(B) \in \mathscr{B}(\overline{\mathbb{R}})$. Thus, using Exercise I-1.3.2,

$$(\phi \circ f)^{-1}(B) = f^{-1}(\phi^{-1}(B)) \in \mathscr{A},$$

and so $\phi \circ f$ is $\mathscr{A}$-measurable, as desired. ■

2.6.14 Corollary (Composition by continuous functions and measurability) *Let $(\mathsf{X},\mathscr{A})$ be a measurable space, let $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}})$, and let $\phi\colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ be continuous. Then $\phi \circ \mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}})$.*

Proof This follows from Proposition 2.6.13, along with Example 2.6.3. ■

In the following result we consider measurability of functions restricted to measurable sets. We recall from Proposition 2.2.6 the definition of the restriction $\mathscr{A}_A$ of a measurable space $(X,\mathscr{A})$ to a measurable subset $A \in \mathscr{A}$.

2.6.15 Proposition (Measurability and restriction) *Let $(\mathsf{X},\mathscr{A})$ be a measurable space, let $\mathsf{f}\colon \mathsf{X} \to \overline{\mathbb{R}}$ be $\mathscr{A}$-measurable, and let $\mathsf{A} \in \mathscr{A}$. Then $\mathsf{f}|\mathsf{A}$ is $\mathscr{A}_\mathsf{A}$-measurable.*

Moreover, if $\mathsf{B} = \mathsf{X} \setminus \mathsf{A}$ and if we have $\mathscr{A}_\mathsf{A}$- and $\mathscr{A}_\mathsf{B}$-measurable functions $\mathsf{f}_\mathsf{A}\colon \mathsf{A} \to \overline{\mathbb{R}}$ and $\mathsf{f}_\mathsf{B}\colon \mathsf{B} \to \overline{\mathbb{R}}$, respectively, then the function $\mathsf{f}\colon \mathsf{X} \to \overline{\mathbb{R}}$ defined by

$$\mathsf{f}(\mathsf{x}) = \begin{cases} \mathsf{f}_\mathsf{A}(\mathsf{x}), & \mathsf{x} \in \mathsf{A}, \\ \mathsf{f}_\mathsf{B}(\mathsf{x}), & \mathsf{x} \in \mathsf{B} \end{cases}$$

is $\mathscr{A}$-measurable.

Proof Let $E \in \mathscr{A}$ so that $A \cap E \in \mathscr{A}_A$. Then, by Proposition I-1.3.5,

$$f^{-1}(A \cap E) = f^{-1}(A) \cap f^{-1}(E),$$

and from this we deduce that $f^{-1}(A \cap E)$ is the intersection of measurable sets, and so measurable.

For the second assertion of the proposition, let $E \in \mathscr{A}$ and write $E = (A \cap E) \cup (B \cap E)$. Then, again by Proposition I-1.3.5,

$$f^{-1}(E) = f^{-1}(A \cap E) \cup f^{-1}(B \cap E) = f_A^{-1}(A \cap E) \cup f_B^{-1}(B \cap E).$$

Since f_A and f_B are $\mathscr{A}_A$- and $\mathscr{A}_B$-measurable, $f_A^{-1}(A \cap E) \in \mathscr{A}_A$ and $f_B^{-1}(B \cap E) \in \mathscr{A}_B$. Since $\mathscr{A}_A, \mathscr{A}_B \subseteq \mathscr{A}$, $f^{-1}(E)$ is the union of $\mathscr{A}$-measurable sets, and so is $\mathscr{A}$-measurable. ■

Let us consider the rôle of measurability with respect to the operations of min and max.

2.6.16 Proposition (Measurability and max and min) *If* $(\mathsf{X}, \mathscr{A})$ *is a measure space and if* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}})$, *then the functions*

$$\mathsf{X} \ni x \mapsto \min\{\mathsf{f}(x), \mathsf{g}(x)\} \in \overline{\mathbb{R}}, \qquad \mathsf{X} \ni x \mapsto \max\{\mathsf{f}(x), \mathsf{g}(x)\} \in \overline{\mathbb{R}}.$$

are $\mathscr{A}$*-measurable.*

Proof Let $a \in \mathbb{R}$ and note that

$$\{x \in X \mid \min\{f(x), g(x)\} \leq a\} = \{x \in X \mid f(x) \leq a\} \cup \{x \in X \mid g(x) \leq a\}$$

and

$$\{x \in X \mid \max\{f(x), g(x)\} \leq a\} = \{x \in X \mid f(x) \leq a\} \cap \{x \in X \mid g(x) \leq b\}.$$

Thus $\{x \in X \mid \min\{f(x), g(x)\} \leq a\}$ and $\{x \in X \mid \max\{f(x), g(x)\} \leq a\}$ are measurable and this gives the result. ■

The previous result has the following obvious corollary that will be useful when we define the integral.

2.6.17 Corollary (Measurability of positive and negative parts of a function) *Let* $(\mathsf{X}, \mathscr{A})$ *be a measurable space and let* $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}})$ *be* $\mathscr{A}$*-measurable. Then the functions* $\mathsf{f}_-, \mathsf{f}_+ \colon \mathsf{X} \to \overline{\mathbb{R}}$ *defined by*

$$\mathsf{f}_+(x) = \max\{\mathsf{f}(x), 0\}, \qquad \mathsf{f}_-(x) = \max\{-\mathsf{f}(x), 0\}$$

are $\mathscr{A}$*-measurable.*

Proof By Example 2.6.8–1 the function $x \mapsto 0$ is $\mathscr{A}$-measurable. The corollary now follows immediately from Proposition 2.6.16. ■

2.6.3 Sequences of measurable functions

In Sections I-3.6 and II-1.7 we considered sequences of continuous functions. on topological spaces? We saw that notions of uniform convergence are important for the preservation of continuity of the limit function. Measurable functions are far more flexible in this regard, and so we are able to assert the measurability of a fairly general collection of operations applied to sequences of measurable functions. First let us define some notation to facilitate the statement of the result. We let $(X, \mathscr{A})$ be a measurable space with $S = (f_j)_{j\in\mathbb{Z}_{>0}}$ a sequence of $\mathscr{A}$-measurable functions. We then define functions $\inf S, \sup S, \liminf S, \limsup S\colon X \to \overline{\mathbb{R}}$ by

$$\inf S(x) = \inf\{f_j(x) \mid j \in \mathbb{Z}_{>0}\}, \quad \sup S(x) = \sup\{f_j(x) \mid j \in \mathbb{Z}_{>0}\},$$
$$\liminf S(x) = \liminf_{j\to\infty} f_j(x), \quad \limsup S(x) = \limsup_{j\to\infty} f_j(x).$$

Note that these four functions are always defined, regardless of the sequence. Let us also define

$$A_S = \{x \in X \mid \liminf S(x) = \limsup S(x)\}$$

and define $\lim S\colon A_S \to \overline{\mathbb{R}}$ by $\lim S(x) = \lim_{j\to\infty} f_j(x)$, noting that this is also a well-defined function. With this notation we have the following result.

2.6.18 Proposition (Limit operations on measurable functions) *Let $(X, \mathscr{A})$ be a measurable space and let $S = (f_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$. Then the following statements hold:*

(i) $\inf S \in L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$;

(ii) $\sup S \in L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$;

(iii) $\liminf S \in L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$;

(iv) $\limsup S \in L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$;

(v) $A_S \in \mathscr{A}$ *and* $\lim S \in L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$.

Proof (i) For $b \in \mathbb{R}$ we have

$$\{x \in X \mid \inf S(x) < b\} = \bigcup_{j\in\mathbb{Z}_{>0}} \{x \in X \mid f_j(x) < b\}.$$

Since the sets of the right are measurable, so is their union.

(ii) For $b \in \mathbb{R}$ we have

$$\{x \in X \mid \sup S(x) \le b\} = \bigcap_{j\in\mathbb{Z}_{>0}} \{x \in X \mid f_j(x) \le b\}.$$

Since the sets on the right are measurable, so is their intersection.

(iii) Define a sequence of functions $(\underline{f}_j)_{j\in\mathbb{Z}_{>0}}$ by $\underline{f}_j(x) = \sup_{k\ge j} f_k(x)$. These functions are $\mathscr{A}$-measurable by part (i). By Proposition I-2.3.16 we have

$$\liminf_{j\to\infty} f_j(x) = \sup\left\{\underline{f}_k(x) \;\middle|\; k \in \mathbb{Z}_{>0}\right\},$$

and so this part of the result follows from part (i).

(iv) Define a sequence of functions $(\overline{f}_j)_{j\in\mathbb{Z}_{>0}}$ by $\overline{f}_j(x) = \sup_{k\geq j} f_k(x)$. These functions are $\mathscr{A}$-measurable by part (ii). By Proposition I-2.3.15 we have

$$\limsup_{j\to\infty} f_j(x) = \inf\left\{\overline{f}_k(x) \;\middle|\; k \in \mathbb{Z}_{>0}\right\},$$

and so this part of the result follows from part (ii).

(v) Measurability of A_S follows from parts (iii) and (iv), along with Lemma 1 from the proof of Proposition 2.6.11. Now let $b \in \mathbb{R}$ and note that

$$\{x \in A_S \mid \lim f(x) \leq b\} = A_S \cap \{x \in X \mid \limsup S(x) \leq b\}.$$

The set on the right is the intersection of measurable sets and so is measurable. This then gives $\mathscr{A}$-measurability of $\lim S$ by Proposition 2.2.6. ■

The following corollary will come up often. Note that this result is unlike most of the results thus far in this section in that it depends on a measure.

2.6.19 Corollary (Measurability of almost everywhere convergent sequences) *Let $(X, \mathscr{A}, \mu)$ be a complete measure space, let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$, and let $f\colon X \to \overline{\mathbb{R}}$ be such that*

$$\mu\Big(X \setminus \Big\{x \in X \;\Big|\; f(x) = \lim_{j\to\infty} f_j(x)\Big\}\Big) = 0.$$

Then f is $\mathscr{A}$-measurable.

Proof Let $S = (f_j)_{j\in\mathbb{Z}_{>0}}$ and define

$$B_S = \left\{x \in X \;\middle|\; f(x) = \lim_{j\to\infty} f_j(x)\right\}.$$

From Proposition 2.6.18 the function $\liminf S$ is $\mathscr{A}$-measurable. Since f and $\liminf S$ agree except on the set B_S which has measure zero, it follows from Proposition 2.6.10 that f is $\mathscr{A}$-measurable since μ is complete. ■

The preceding few results had to do with the measurability of various sorts of limits of measurable functions. Let us now study systematically the various sorts of convergence that may be experienced by sequences of measurable functions. In we described the notions of pointwise and uniform convergence in a general way using topological ideas. These definitions carry over to sequences of functions defined on measure spaces, but there are additional notions arising from the measure theoretic setting, as the following definitions make clear. what

2.6.20 Definition (Modes of convergence for sequences of measurable functions) Let $(X, \mathscr{A}, \mu)$ be a measure space, let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$, and let $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$. The sequence

(i) ***converges pointwise*** to f if $\lim_{j\to\infty} f_j(x) = f(x)$ for every $x \in X$,

(ii) ***converges pointwise almost everywhere*** to f if

$$\mu\Big(X \setminus \Big\{x \in X \;\Big|\; f(x) = \lim_{j\to\infty} f_j(x)\Big\}\Big) = 0,$$

(iii) ***converges uniformly*** to f if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $|f(x) - f_j(x)| < \epsilon$ for every $x \in X$ and for every $j \geq N$,

(iv) ***converges almost uniformly*** to f if, for every $\delta \in \mathbb{R}_{>0}$, there exists a set $E_\delta \subseteq X$ having the following properties:

 (a) $\mu(E_\delta) < \delta$;

 (b) for every $\epsilon \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{Z}_{>0}$ such that $|f(x) - f_j(x)| < \epsilon$ for every $x \in X \setminus E_\delta$ and for every $j \geq N$,

and

(v) ***converges in measure*** to f if, for every $\epsilon \in \mathbb{R}_{>0}$,

$$\lim_{j\to\infty} \mu\big(\{x \in X \mid |f(x) - f_j(x)| > \epsilon\}\big) = 0. \qquad \bullet$$

Some of the relationships between the various notions of convergence are obvious. For example, the implications

$$\text{(iv)} \Longleftarrow \text{(iii)} \implies \text{(i)} \implies \text{(ii)}$$

obviously hold. Moreover, the converse implications of some of the preceding implications fairly obviously do not hold in general. For example, we know from Section I-3.6.2 that generally (i)$\not\Rightarrow$(iii). It is also pretty evident that generally (ii)$\not\Rightarrow$(i); see Exercise 2.6.4. Let us now explore the possibility of other implications. The first result shows that, perhaps a little surprisingly, (ii) implies (iv) when the functions in the sequence and the limit function are $\mathbb{R}$-valued, and when the measure is finite.

2.6.21 Theorem (Egorov's[7] Theorem) *Let* $(X, \mathscr{A}, \mu)$ *be a finite measure space and let* $f_j, f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$, $j \in \mathbb{Z}_{>0}$, *have the following properties:*

(i) the sets $\{x \in X \mid f_j(x) \notin \mathbb{R}\}$, $j \in \mathbb{Z}_{>0}$, *and* $\{x \in X \mid f(x) \notin \mathbb{R}\}$ *have measure zero;*

(ii) $(f_j)_{j\in\mathbb{Z}_{>0}}$ *converges pointwise almost everywhere to* f.

Then $(f_j)_{j\in\mathbb{Z}_{>0}}$ *converges almost uniformly to* f.

Proof First let us suppose that f and f_j, $j \in \mathbb{Z}_{>0}$, are $\mathbb{R}$-valued. For $k, m \in \mathbb{Z}_{>0}$ define

$$E_{km} = \Big\{x \in X \;\Big|\; |f(x) - f_m(x)| < \tfrac{1}{k}\Big\}.$$

Since $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges almost everywhere to f, there exists a set $Z \subseteq X$ such that

1. $\mu(Z) = 0$ and

[7] Dimitri Fedorovich Egorov (1869–1931) was a Russian mathematician whose main mathematical contributions were to differential geometry and analysis.

2. for $k \in \mathbb{Z}_{>0}$ and $x \in X \setminus Z$, there exists $m \in \mathbb{Z}_{>0}$ such that $|f(x) - f_j(x)| < \frac{1}{k}$ for $j \geq m$.

That is to say,

$$X \setminus Z \subseteq \bigcup_{n \in \mathbb{Z}_{>0}} \bigcap_{m \geq n} E_{km} \quad \Longrightarrow \quad Z \subseteq \bigcap_{n \in \mathbb{Z}_{>0}} \bigcup_{m \geq n} X \setminus E_{km}$$

for every $k \in \mathbb{Z}_{>0}$, using De Morgan's Laws. Denote $A_{kn} = \cup_{m \geq n} X \setminus E_{km}$. Note that $A_{kn} \supseteq A_{k(n+1)}$ for every $k, n \in \mathbb{Z}_{>0}$, and that $\cap_{n \in \mathbb{Z}_{>0}} A_{kn} \subseteq Z$ which implies that $\cap_{n \in \mathbb{Z}_{>0}} A_{kn}$ has zero measure, being a subset of a set with zero measure. Let $Z_k = \cap_{n \in \mathbb{Z}_{>0}} A_{kn}$ and note that

$$\lim_{n \to \infty} \mu(A_{kn}) = \lim_{n \to \infty} \mu(A_{kn} \setminus Z_k) = 0$$

using Proposition 2.3.3.

Let $\delta \in \mathbb{R}_{>0}$. For $k \in \mathbb{Z}_{>0}$ let $N_k \in \mathbb{Z}_{>0}$ be such that $\mu(A_{kn}) < \frac{\delta}{2^k}$ for $n \geq N_k$. Define

$$E_\delta = \bigcup_{k \in \mathbb{Z}_{>0}} A_{kN_k}.$$

Then

$$\mu(E_\delta) \leq \sum_{k=1}^{\infty} \mu(A_{kN_k}) < \sum_{k=1}^{\infty} \frac{\delta}{2^k} = \delta$$

by Example I-2.4.2–1.

Now let $\epsilon \in \mathbb{R}_{>0}$ and take $K \in \mathbb{Z}_{>0}$ such that $\frac{1}{K} < \epsilon$. If $x \in X \setminus E_\delta$ we have, by definition of E_δ and De Morgan's Laws,

$$x \in \bigcap_{k \in \mathbb{Z}_{>0}} \bigcap_{m \geq N_k} E_{km},$$

which implies in particular that $x \in E_{Km}$ whenever $m \geq N_K$. That is to say, if $j \geq N_K$ then $|f(x) - f_j(x)| < \epsilon$ for every $x \in X \setminus E_\delta$, as desired.

To conclude the proof, let us relax the assumption made above that f and f_j, $j \in \mathbb{Z}_{>0}$, are $\mathbb{R}$-valued. Define

$$N = \{x \in X \mid f(x) \notin \mathbb{R}\}, \quad N_j = \{x \in X \mid f_j(x) \notin \mathbb{R}\}, \qquad j \in \mathbb{Z}_{>0}.$$

If $Z = N \cup (\cup_{j \in \mathbb{Z}_{>0}} N_j)$ then Z is a measurable set with zero measure, being a countable union of sets with zero measure. The hypotheses from the first part of the proof hold for $X \setminus Z$ and for f and f_j, $j \in \mathbb{Z}_{>0}$, restricted to $X \setminus Z$. That is to say, for every $\delta \in \mathbb{R}_{>0}$ there exists a set $E'_\delta \subseteq (X \setminus Z)$ such that $\mu(E'_\delta) < \delta$ and such that, for every $\epsilon \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{Z}_{>0}$ such that $|f(x) - f_j(x)| < \epsilon$ for every $x \in X \setminus (Z \cup E'_\delta)$ and for every $j \geq N$. Now let $\delta \in \mathbb{R}_{>0}$ and take $E_\delta = E'_\delta \cup Z$. Note that $\mu(E_\delta) = \mu(E'_\delta) < \delta$. Now, for $\epsilon \in \mathbb{R}_{>0}$ let N be chosen as above, so that $|f(x) - f_j(x)| < \epsilon$ for every $x \in X \setminus E_\delta$ and for every $j \geq N$. This gives almost uniform convergence of $(f_j)_{j \in \mathbb{Z}_{>0}}$ to f, as desired. ■

Note that the theorem allows us to immediately conclude that generally (iv)$\not\Rightarrow$(iii) from Definition 2.6.20. Indeed, suppose that $(f_j)_{j \in \mathbb{Z}_{>0}}$ is a sequence of $\mathbb{R}$-valued functions on $[0, 1]$ that converges pointwise, but not uniformly, to a function f. Then the preceding theorem implies that the sequence converges almost uniformly to f.

The next example shows that finiteness of the measure space in Egorov's Theorem is necessary.

2.6.22 Example (Egorov's Theorem generally fails for measure spaces that are not finite) We take $X = \mathbb{R}$, $\mathscr{A} = \mathscr{L}(\mathbb{R})$, and $\mu = \lambda$. We consider the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{L}^{(0)}((\mathbb{R}, \mathscr{L}(\mathbb{R})); \mathbb{R})$ defined by $f_j = \chi_{[j,j+1)}$. We also define $f \in \mathsf{L}^{(0)}((\mathbb{R}, \mathscr{L}(\mathbb{R})); \mathbb{R})$ by $f(x) = 0$ for all $x \in \mathbb{R}$. We claim that $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges pointwise, and so pointwise almost everywhere, to f, but does not converge almost uniformly to f. To verify pointwise convergence, let $x \in \mathbb{R}$ and choose $N \in \mathbb{Z}_{>0}$ such that $N + 1 > x$, Then we have $f_j(x) = 0$ for all $j \geq N$, so verifying pointwise convergence to f. To see that the sequence does not converge almost uniformly, let $\delta, \epsilon \in (0, 1)$ and suppose that $E_\delta \subseteq \mathbb{R}$ is such that there exists $N \in \mathbb{Z}_{>0}$ for which $|f(x) - f_j(x)| < \epsilon$ for $x \in \mathbb{R} \setminus E_\delta$ and for $j \geq N$. This means that $|f_j(x)| \geq \epsilon$ on a set A contained in E_δ for $j \geq N$. But this implies that $[N, N + 1) \subseteq E_\delta$, implying that $\mu(E_\delta) > \delta$. This precludes almost uniform convergence. •

The preceding discussion concerning the relationships between modes of convergence has not involved convergence in measure. Let us now investigate the rôle of convergence in measure relative to the other modes of convergence. The first result establishes that for finite measure spaces we have the implication (ii) $\Longrightarrow$ (v) from Definition 2.6.20.

2.6.23 Proposition (Almost everywhere pointwise convergence sometimes implies convergence in measure) *Let $(X, \mathscr{A}, \mu)$ be a finite measure space. Consider a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ and a function f in $\mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ with the following properties:*

(i) the sets $\{x \in X \mid f(x) \notin \mathbb{R}\}$ and $\{x \in X \mid f_j(x) \notin \mathbb{R}\}$, $j \in \mathbb{Z}_{>0}$, have measure zero;

(ii) $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges pointwise almost everywhere to f.

Then $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges in measure to f.

Proof First suppose that f and f_j, $j \in \mathbb{Z}_{>0}$, are $\mathbb{R}$-valued. Let $\epsilon \in \mathbb{R}_{>0}$ and define

$$A_{\epsilon,j} = \{x \in X \mid |f(x) - f_j(x)| > \epsilon\}, \qquad j \in \mathbb{Z}_{>0}$$

and $B_{\epsilon,k} = \cup_{j=1}^k A_{\epsilon,j}$, $k \in \mathbb{Z}_{>0}$. Then we have $B_{k+1} \supseteq B_k$ for $k \in \mathbb{Z}_{>0}$ and

$$\cap_{k\in\mathbb{Z}_{>0}} B_{\epsilon,k} \subseteq \{x \in X \mid (f_j(x))_{j\in\mathbb{Z}_{>0}} \text{ does not converge to } f(x)\}$$

Therefore, $\mu(\cap_{k\in\mathbb{Z}_{>0}} B_{\epsilon,k}) = 0$ and so, by Proposition 2.3.3, $\lim_{k\to\infty} B_{\epsilon,k} = 0$. Therefore, since $A_{\epsilon,j} \subseteq B_{\epsilon,j}$ for $j \in \mathbb{Z}_{>0}$, we have $\lim_{j\to\infty} A_{\epsilon,j} = 0$. This is exactly the statement that $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges to f in measure.

To complete the proof, suppose that f_j, $j \in \mathbb{Z}_{>0}$, are not necessarily $\mathbb{R}$-valued. Let

$$N = \{x \in X \mid f(x) \notin \mathbb{R}\}, \quad N_j = \{x \in X \mid f_j(x) \notin \mathbb{R}\}, \qquad j \in \mathbb{Z}_{>0}$$

so that $Z = N \cup (\cup_{j\in\mathbb{Z}_{>0}} N_j)$ is a measurable set with zero measure, it being a countable union of sets with zero measure. The first part of the proof then applies for $X \setminus Z$ and for f and f_j, $j \in \mathbb{Z}_{>0}$, restricted to $X \setminus Z$. Thus, for $\epsilon \in \mathbb{R}_{>0}$ we have

$$\lim_{j\to\infty} \mu(\{x \in X \setminus Z \mid |f(x) - f_j(x)| > \epsilon\}) = 0.$$

Since

$$\begin{aligned}\{x \in X \mid |f(x) - f_j(x)| > \epsilon\} &= \{x \in X \setminus Z \mid |f(x) - f_j(x)| > \epsilon\} \cup \{x \in Z \mid |f(x) - f_j(x)| > \epsilon\} \\ &\subseteq \{x \in X \setminus Z \mid |f(x) - f_j(x)| > \epsilon\} \cup Z,\end{aligned}$$

we have

$$\lim_{j\to\infty} \mu(\{x \in X \mid |f(x) - f_j(x)| > \epsilon\}) = 0,$$

giving convergence in measure as desired. ■

The condition that the measure space be finite is generally necessary in the preceding result.

2.6.24 Example (Almost everywhere pointwise convergence does not always imply convergence in measure) Here we take $X = \mathbb{R}$, $\mathscr{A} = \mathscr{L}(\mathbb{R})$, and $\mu = \lambda$. We define a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ and a function f in $\mathsf{L}^{(0)}((\mathbb{R}, \mathscr{L}(\mathbb{R})); \mathbb{R})$ by $f_j = \chi_{[j,j+1)}$ and $f(x) = 0$ for $x \in \mathbb{R}$. We saw in Example 2.6.22 that the sequence converges pointwise to f, and so converges pointwise almost everywhere to f. However, if $\epsilon \in (0, 1)$ then

$$\lambda(\{x \in \mathbb{R} \mid |f(x) - f_j(x)| > \epsilon\}) = 1$$

which clearly precludes the sequence from converging to f in measure. •

Now let us investigate the extent to which convergence in measure implies almost everywhere pointwise convergence. The following example shows that the general implication fails to hold, even for finite measure spaces.

2.6.25 Example (Convergence in measure does not imply almost everywhere pointwise convergence) We take $X = [0, 1)$, $\mathscr{A} = \mathscr{L}([0, 1))$, and $\mu = \lambda_{[0,1)}$. We define a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{L}^{(0)}(([0, 1), \mathscr{L}([0, 1))); \mathbb{R})$ as follows. For $k \in \mathbb{Z}_{\geq 0}$ we define $f_{2^k}, f_{2^k+1}, \dots, f_{2^{k+1}-1}$ by $f_{2^k+j} = \chi_{[j2^{-k},(j+1)2^{-k})}$, $j \in \{0, 1, \dots, 2^k - 1\}$. Thus, for example,

$$\begin{aligned}&f_1 = \chi_{[0,1)},\\ &f_2 = \chi_{[0,\frac{1}{2})},\ f_3 = \chi_{[\frac{1}{2},1)},\\ &f_4 = \chi_{[0,\frac{1}{4})},\ f_5 = \chi_{[\frac{1}{4},\frac{1}{2})},\ f_6 = \chi_{[\frac{1}{2},\frac{3}{4})},\ f_7 = \chi_{[\frac{3}{4},1)}.\end{aligned}$$

We also define $f \in \mathsf{L}^{(0)}(([0, 1), \mathscr{L}([0, 1))); \mathbb{R})$ by $f(x) = 0$ for $x \in [0, 1)$. We claim that this sequence converges in measure to f, but does not converge pointwise almost everywhere to f.

To verify convergence in measure, let $\epsilon \in \mathbb{R}_{>0}$ and note that for any $j \in \mathbb{Z}_{>0}$ we have

$$\{x \in [0, 1) \mid |f(x) - f_j(x)| > \epsilon\} \subseteq \{x \in [0, 1) \mid |f_j(x)| > 0\}.$$

If $j \in \{2^k, 2^k + 1, \dots, 2^{k+1} - 1\}$ then

$$\lambda(\{x \in [0, 1) \mid |f_j(x)| > 0\}) = 2^{-k}.$$

Therefore, it follows that

$$\lim_{j\to\infty} \lambda(\{x \in [0,1) \mid |f_j(x)| > 0\}) = 0,$$

giving convergence in measure.

Now we verify that the sequence does not converge pointwise almost everywhere. Let $x \in [0,1)$ and let $N \in \mathbb{Z}_{>0}$. Choose $k \in \mathbb{Z}_{>0}$ such that $2^k > N$ and choose $j \in \{0,1,\dots,2^k-1\}$ such that $x \in [j2^{-k},(j+1)2^{-k})$. Then, for $m \in \{2^k, 2^k+1,\dots,2^{k+1}-1\}$ we have

$$f_m(x) = \begin{cases} 1, & m = 2^k + j, \\ 0, & \text{otherwise.} \end{cases}$$

Thus, no matter how large we choose N, there are terms beyond the Nth term in the sequence $(f_j(x))_{j\in\mathbb{Z}_{>0}}$ that have value 1 and terms beyond the Nth term in the sequence $(f_j(x))_{j\in\mathbb{Z}_{>0}}$ that have value 0. This precludes pointwise convergence at x. Since this is true for every $x \in [0,1)$ it follows that almost everywhere pointwise convergence is precluded. Indeed, the sequence converges pointwise nowhere. •

The situation is not entirely hopeless, however. Indeed, one has the following result.

2.6.26 Proposition (Convergence in measure implies almost everywhere pointwise convergence of a subsequence) *Let $(\mathsf{X},\mathscr{A},\mu)$ be a measure space, let $(\mathsf{f}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}})$ and let $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}})$ satisfy the following:*

(i) the sets $\{x \in \mathsf{X} \mid \mathsf{f}(x) \notin \mathbb{R}\}$ and $\{x \in \mathsf{X} \mid \mathsf{f}_j(x) \notin \mathbb{R}\}$ have measure zero;

(ii) the sequence $(\mathsf{f}_j)_{j\in\mathbb{Z}_{>0}}$ converges to f in measure.

Then there exists a subsequence of $(\mathsf{f}_j)_{j\in\mathbb{Z}_{>0}}$ which converges pointwise almost everywhere to f.

Proof Define a strictly increasing sequence $(j_k)_{k\in\mathbb{Z}_{>0}}$ in $\mathbb{Z}_{>0}$ as follows. Let j_1 be such that

$$\mu(\{x \in X \mid |f(x) - f_{j_1}(x)| > 1\}) \le \tfrac{1}{2},$$

this being possible by definition of convergence in measure. Then suppose that $j_1,\dots,j_k$ have been defined. Define j_{k+1} such that $j_{k+1} > j_k$ and such that

$$\mu(\{x \in X \mid |f(x) - f_{j_{k+1}}(x)| > \tfrac{1}{k+1}\} \le \tfrac{1}{2^{k+1}},$$

this again being possible by definition of convergence in measure. Now define

$$A_k = \{x \in X \mid |f(x) - f_{j_k}(x)| < \tfrac{1}{k}\}, \qquad k \in \mathbb{Z}_{>0},$$

and $B_m = \cup_{k=m}^{\infty} A_k$, $m \in \mathbb{Z}_{>0}$. Note that $B_{m+1} \supseteq B_m$ for $m \in \mathbb{Z}_{>0}$. Moreover,

$$\mu(B_m) \le \sum_{k=m}^{\infty} \mu(A_k) \le \sum_{k=m}^{\infty} \frac{1}{2^k} = \frac{1}{2^{m-1}} \sum_{k=1}^{\infty} \frac{1}{2^k} = \frac{1}{2^{m-1}}$$

by Example I-2.4.2–1. Therefore, by Proposition 2.3.3,

$$\mu(\cap_{m=1}^{\infty} B_m) = \lim_{m\to\infty} \mu(B_m) \le \lim_{m\to\infty} \frac{1}{2^{m-1}} = 0.$$

Now, if $x \notin \cap_{m=1}^{\infty} B_m$ there exists $m \in \mathbb{Z}_{>0}$ such that $x \notin B_m$. Thus $x \notin \cup_{k=m}^{\infty} A_k$ and so

$$|f(x) - f_{j_k}(x)| < \tfrac{1}{2^k}, \qquad k \ge m.$$

Thus $\lim_{k\to\infty} f_{j_k}(x) = f(x)$. Thus $(f_{j_k})_{k\in\mathbb{Z}_{>0}}$ converges pointwise to f on $X \setminus (\cap_{m=1}^{\infty} B_m)$. This gives almost everywhere pointwise convergence of this subsequence to f. ■

2.6.4 ℂ- and vector-valued measurable functions

It is important to be able to talk about functions taking values in spaces more interesting than $\overline{\mathbb{R}}$. In particular, ℂ-valued functions will be frequently encountered in these volumes. Here we allow this by considering functions taking values in $\mathbb{R}^n$.

First let us define what we mean by a measurable $\mathbb{R}^n$-valued function.

2.6.27 Definition (Measurable vector-valued function) For a measurable space $(X, \mathscr{A})$, a function $\boldsymbol{f}\colon X \to \mathbb{R}^n$ is ***$\mathscr{A}$-measurable*** if its components $f_1, \dots, f_n\colon X \to \mathbb{R}$ are measurable in the sense of Definition 2.6.5. We denote the set of measurable $\mathbb{R}^n$-valued maps by $\mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R}^n)$, or simply by $\mathsf{L}^{(0)}(X; \mathbb{R}^n)$ with the understanding that the σ-algebra $\mathscr{A}$ is implicit. •

Let us relate this notion of measurability to that in Definition 2.6.1.

2.6.28 Proposition (Characterisation of vector-valued measurable functions) *For a measurable space* $(X, \mathscr{A})$ *and for a function* $\mathbf{f}\colon X \to \mathbb{R}^n$ *the following statements are equivalent:*

(i) $\mathbf{f} \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R}^n)$*;*

(ii) $\mathbf{f}$ *is* $(\mathscr{A}, \mathscr{B}(\mathbb{R}^n))$*-measurable.*

Proof Suppose that $\boldsymbol{f} \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R}^n)$. By Propositions 2.4.9 it follows that $f_j^{-1}((-\infty, b_j]) \in \mathscr{A}$ for every $b_j \in \mathbb{R}$ and for $j \in \{1, \dots, n\}$. Now note that

$$\boldsymbol{f}^{-1}((-\infty, b_1] \times \cdots \times (-\infty, b_n]) = f_1^{-1}((-\infty, b_1]) \cap \cdots \cap f_n^{-1}((-\infty, b_n]) \in \mathscr{A}$$

for every $b_1, \dots, b_n \in \mathbb{R}$. By Propositions 2.5.9 and 2.6.2 it follows that $\boldsymbol{f}$ is $(\mathscr{A}, \mathscr{B}(\mathbb{R}^n))$-measurable.

Next suppose that $\boldsymbol{f}$ is $(\mathscr{A}, \mathscr{B}(\mathbb{R}^n))$-measurable. Then, for $j \in \{1, \dots, n\}$ and $b_j \in \mathbb{R}$,

$$\begin{aligned} f_j^{-1}((-\infty, b_j]) &= X \cap \cdots \cap f_j^{-1}((-\infty, b_j]) \cap \cdots \cap X \\ &= f_1^{-1}(\mathbb{R}) \cap \cdots \cap f_j^{-1}((-\infty, b_j]) \cap \cdots \cap f_n^{-1}(\mathbb{R}) \\ &= \boldsymbol{f}^{-1}(\mathbb{R} \times \cdots \times (-\infty, b_j] \times \cdots \times \mathbb{R}). \end{aligned}$$

Since $\mathbb{R} \times \cdots \times (-\infty, b_j] \times \cdots \times \mathbb{R}$ is a Borel set (it is closed), it follows that $f_j^{-1}((-\infty, b_j])$ is a Borel set, and so the result follows from Propositions 2.4.9 and 2.6.2. ■

The definition of measurable $\mathbb{C}$-valued functions follows directly from the preceding constructions. Indeed, we note that $\mathbb{C}$ is isomorphic as a $\mathbb{R}$-vector space to $\mathbb{R}^n$ via the isomorphism $z \mapsto (\mathrm{Re}(z), \mathrm{Im}(z))$. Thus the following definition simply specialises the above general definition.

2.6.29 Definition (Measurable $\mathbb{C}$-valued functions) For a measurable space $(X, \mathscr{A})$, a function $f\colon X \to \mathbb{C}$ is ***$\mathscr{A}$-measurable*** if the $\mathbb{R}$-valued functions

$$\mathrm{Re}(f)\colon x \mapsto \mathrm{Re}(f(x)), \quad \mathrm{Im}(f)\colon x \mapsto \mathrm{Im}(f(x))$$

are measurable in the sense of Definition 2.6.5. We denote the set of measurable $\mathbb{C}$-valued maps by $\mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{C})$, with the understanding that the σ-algebra $\mathscr{A}$ is implicit. •

It is straightforward to adapt the results concerning operations on measurable functions in Section 2.6.2 to vector-valued functions. Let us record this here for $\mathbb{R}^n$-valued functions, noting that these results apply immediately to $\mathbb{C}$-valued functions.

2.6.30 Proposition (Algebraic operations on measurable functions) *Let* $(\mathrm{X}, \mathscr{A})$ *be a measurable space, let* $\mathbf{f}, \mathbf{g} \in \mathsf{L}^{(0)}((\mathrm{X}, \mathscr{A}); \mathbb{R}^{\mathrm{n}})$*, and let* $\mathrm{a} \in \mathbb{R}$*. Then the functions* $\mathbf{f} + \mathbf{g}$ *and* $\mathrm{a}\mathbf{f}$ *defined by*

$$(\mathbf{f} + \mathbf{g})(\mathrm{x}) = \mathbf{f}(\mathrm{x}) + \mathbf{g}(\mathrm{x}), \qquad (\mathrm{a}\mathbf{f})(\mathrm{x}) = \mathrm{a}(\mathbf{f}(\mathrm{x}))$$

are $\mathscr{A}$*-measurable.*

Proof This follows directly from the definition of $\mathscr{A}$-measurable vector-valued functions, the definitions of vector addition and scalar multiplication, and Proposition 2.6.11. ■

Next we consider compositions of measurable functions with functions between Euclidean spaces.

2.6.31 Proposition (Composition and measurable functions) *Let* $(\mathrm{X}, \mathscr{A})$ *be a measurable space, let* $\mathbf{f} \in \mathsf{L}^{(0)}((\mathrm{X}, \mathscr{A}); \mathbb{R}^{\mathrm{n}})$*, and let* $\boldsymbol{\phi}\colon \mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{m}}$ *be* $\mathscr{B}(\overline{\mathbb{R}})$*-measurable. Then* $\boldsymbol{\phi} \circ \mathbf{f} \in \mathsf{L}^{(0)}((\mathrm{X}, \mathscr{A}); \mathbb{R}^{\mathrm{m}})$*.*

Proof Let $B \in \mathscr{B}(\mathbb{R}^m)$. By assumption and by Proposition 2.6.6 we have $\boldsymbol{\phi}^{-1}(B) \in \mathscr{B}(\mathbb{R}^n)$. Thus, using Exercise I-1.3.2,

$$(\boldsymbol{\phi} \circ \boldsymbol{f})^{-1}(B) = \boldsymbol{f}^{-1}(\boldsymbol{\phi}^{-1}(B)) \in \mathscr{A},$$

and so $\boldsymbol{\phi} \circ \boldsymbol{f}$ is $\mathscr{A}$-measurable, as desired. ■

2.6.32 Corollary (Composition by continuous functions and measurability) *Let* $(\mathrm{X}, \mathscr{A})$ *be a measurable space, let* $\mathbf{f} \in \mathsf{L}^{(0)}((\mathrm{X}, \mathscr{A}); \mathbb{R}^{\mathrm{n}})$*, and let* $\boldsymbol{\phi}\colon \mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{n}}$ *be continuous. Then* $\boldsymbol{\phi} \circ \mathbf{f} \in \mathsf{L}^{(0)}((\mathrm{X}, \mathscr{A}); \mathbb{R}^{\mathrm{m}})$*.*

Proof This follows from Proposition 2.6.31, along with Example 2.6.3. ■

2.6.33 Corollary (Measurability of norms of functions) *Let* $(\mathsf{X}, \mathscr{A})$ *be a measurable space, let* $\mathbf{f} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \mathbb{R}^{\mathsf{n}})$*, and let* $\boldsymbol{\phi}\colon \mathbb{R}^{\mathsf{n}} \to \mathbb{R}^{\mathsf{n}}$ *be continuous. Then the function* $\mathsf{x} \mapsto \|\mathbf{f}(\mathsf{x})\|_{\mathbb{R}^{\mathsf{n}}}$ *is* $\mathscr{A}$*-measurable.*

Proof This follows from the previous corollary, along with continuity of the norm (). ■ what?

Next we consider the restrictions of measurable functions, recalling from Proposition 2.2.6 the definition of the restriction $\mathscr{A}_A$ of a measurable space $(X, \mathscr{A})$ to a measurable subset $A \in \mathscr{A}$.

2.6.34 Proposition (Measurability and restriction) *Let* $(\mathsf{X}, \mathscr{A})$ *be a measurable space, let* $\mathbf{f}\colon \mathsf{X} \to \mathbb{R}^{\mathsf{n}}$ *be* $\mathscr{A}$*-measurable, and let* $\mathsf{A} \in \mathscr{A}$*. Then* $\mathbf{f}|\mathsf{A}$ *is* $\mathscr{A}_{\mathsf{A}}$*-measurable.*

Moreover, if $\mathsf{B} = \mathsf{X} \setminus \mathsf{A}$ *and if we have* $\mathscr{A}_{\mathsf{A}}$*- and* $\mathscr{A}_{\mathsf{B}}$*-measurable functions* $\mathbf{f}_{\mathsf{A}}\colon \mathsf{A} \to \mathbb{R}^{\mathsf{n}}$ *and* $\mathbf{f}_{\mathsf{B}}\colon \mathsf{B} \to \mathbb{R}^{\mathsf{n}}$*, respectively, then the function* $\mathbf{f}\colon \mathsf{X} \to \mathbb{R}^{\mathsf{n}}$ *defined by*

$$\mathbf{f}(\mathsf{x}) = \begin{cases} \mathbf{f}_{\mathsf{A}}(\mathsf{x}), & \mathsf{x} \in \mathsf{A}, \\ \mathbf{f}_{\mathsf{B}}(\mathsf{x}), & \mathsf{x} \in \mathsf{B} \end{cases}$$

is $\mathscr{A}$*-measurable.*

Proof Let $B \in \mathscr{A}$ so that $A \cap B \in \mathscr{A}_A$. Then, by Proposition I-1.3.5,

$$\boldsymbol{f}^{-1}(A \cap B) = \boldsymbol{f}^{-1}(A) \cap \boldsymbol{f}^{-1}(B),$$

and from this we deduce that $\boldsymbol{f}^{-1}(A \cap B)$ is the intersection of measurable sets, and so measurable.

For the second assertion of the proposition, let $E \in \mathscr{A}$ and write $E = (A \cap E) \cup (B \cap E)$. Then, again by Proposition I-1.3.5,

$$\boldsymbol{f}^{-1}(E) = \boldsymbol{f}^{-1}(A \cap E) \cup \boldsymbol{f}^{-1}(B \cap E) = \boldsymbol{f}_A^{-1}(A \cap E) \cup \boldsymbol{f}_B^{-1}(B \cap E).$$

Since $\boldsymbol{f}_A$ and $\boldsymbol{f}_B$ are $\mathscr{A}_A$- and $\mathscr{A}_B$-measurable, $\boldsymbol{f}_A^{-1}(A \cap E) \in \mathscr{A}_A$ and $\boldsymbol{f}_B^{-1}(B \cap E) \in \mathscr{A}_B$. Since $\mathscr{A}_A, \mathscr{A}_B \subseteq \mathscr{A}$, $\boldsymbol{f}^{-1}(E)$ is the union of $\mathscr{A}$-measurable sets, and so is $\mathscr{A}$-measurable. ■

Finally, we consider measurability of limits of vector-valued functions. We consider a sequence $S = (\boldsymbol{f}_j)_{j \in \mathbb{Z}_{>0}}$ of $\mathbb{R}^n$-valued functions on a measurable space $(X, \mathscr{A})$. Let us denote

$$A_S = \left\{x \in \mathsf{X} \;\middle|\; \lim_{j \to \infty} \boldsymbol{f}_j(x) \text{ exists}\right\}$$

and define $\lim S\colon A_S \to \mathbb{R}^n$ by $\lim S(x) = \lim_{j \to \infty} \boldsymbol{f}_j(x)$. With this notation we have the following result.

2.6.35 Proposition (Pointwise limits of sequences of measurable functions) *Let* $(\mathsf{X}, \mathscr{A})$ *be a measurable space and let* $\mathsf{S} = (\mathbf{f}_{\mathsf{j}})_{\mathsf{j} \in \mathbb{Z}_{>0}}$ *be a sequence in* $\mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \mathbb{R}^{\mathsf{n}})$*. Then the set* A_{S} *and the function* $\lim \mathsf{S}$ *are* $\mathscr{A}$*-measurable.*

Proof Let $f_{1,j}, \dots, f_{n,j}$ be the components of $\boldsymbol{f}_j$, $j \in \mathbb{Z}_{>0}$, and for $k \in \{1, \dots, n\}$ define

$$A_{S,k} = \left\{x \in \mathsf{X} \;\middle|\; \lim_{j \to \infty} f_{k,j}(x) \text{ exists}\right\}.$$

Note that $A_S = \cap_{k=1}^n A_{S,k}$ so that A_S is measurable, being a finite intersection of measurable sets. From Propositions 2.6.15 and 2.6.18 it follows that the function

$$A_S \ni x \mapsto \lim_{j\to\infty} f_{k,j}(x) \in \mathbb{R}$$

is $\mathscr{A}$-measurable. The definition of measurability of vector-valued functions now gives the result. ■

For almost everywhere pointwise convergent sequences, this gives the following result.

2.6.36 Corollary (Measurability of almost everywhere convergent sequences) *Let $(\mathsf{X}, \mathscr{A}, \mu)$ be a complete measure space, let $(\mathbf{f}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \mathbb{R}^n)$, and let $\mathbf{f}\colon \mathsf{X} \to \mathbb{R}^n$ be such that*

$$\mu\Big(\mathsf{X} \setminus \Big\{x \in \mathsf{X} \;\Big|\; \mathbf{f}(x) = \lim_{j\to\infty} \mathbf{f}_j(x)\Big\}\Big) = 0.$$

Then $\mathbf{f}$ is $\mathscr{A}$-measurable.

Proof Let $f_{1,j}, \dots, f_{n,j}$ be the components of $\boldsymbol{f}_j$, $j \in \mathbb{Z}_{>0}$, let $f_1, \dots, f_n$ be the components of $\boldsymbol{f}$, and define

$$B_k = \Big\{x \in A \;\Big|\; f_k(x) = \lim_{j\to\infty} f_{k,j}(x)\Big\}, \qquad k \in \{1, \dots, n\},$$

and

$$B = \Big\{x \in A \;\Big|\; f(x) = \lim_{j\to\infty} f_j(x)\Big\}.$$

Note that $B = \cap_{k=1}^n B_k$. By hypothesis, $\mu(X \setminus B) = 0$, and we claim that $\mu(X \setminus B_k) = 0$ for $k \in \{1, \dots, n\}$. Indeed, suppose that $\mu(X \setminus B_k) > 0$ for some $k_0 \in \{1, \dots, n\}$. Then

$$\mu(X \setminus B) = \mu(X \setminus \cap_{j=1}^k B_k) = \mu(\cup_{k=1}^n X \setminus B_k) \geq \mu(X \setminus B_{k_0}) > 0,$$

contrary to our hypothesis. Since $\mu(X \setminus B_k) = 0$ for every $k \in \{1, \dots, n\}$, it follows from Corollary 2.6.19 that f_k is measurable, and so $\boldsymbol{f}$ is also measurable. ■

2.6.5 Simple functions and approximations of measurable functions

In this section we consider a specific class of measurable functions that will be fundamental to our construction of the integral in Section 2.7. There are various ways to characterise this class of functions, and the following result gives some of these.

2.6.37 Proposition (Characterisations of simple functions) *For a measurable space $(\mathsf{X}, \mathscr{A})$ and a function $\mathsf{f}\colon \mathsf{X} \to \overline{\mathbb{R}}$ the following statements are equivalent:*

(i) $\mathrm{image}(\mathsf{f}) = \{\mathsf{a}_1, \dots, \mathsf{a}_k\} \subseteq \overline{\mathbb{R}}$ *and the sets* $\mathsf{f}^{-1}(\mathsf{a}_j)$, $\mathsf{j} \in \{1, \dots, \mathsf{k}\}$, *are measurable;*

(ii) there exists $\mathsf{B}_1, \dots, \mathsf{B}_m \in \mathscr{A}$ *and* $\mathsf{b}_1, \dots, \mathsf{b}_m \in \overline{\mathbb{R}}$ *such that* $\mathsf{f} = \sum_{j=1}^m \mathsf{b}_j \chi_{\mathsf{B}_j}$*;*

(iii) there exist pairwise disjoint sets $C_1, \dots, C_r \in \mathscr{A}$ *and* $c_1, \dots, c_r \in \overline{\mathbb{R}}$ *such that* $f = \sum_{j=1}^{r} c_j \chi_{C_j}$.

Proof (i) $\Longrightarrow$ (ii) Given $f\colon X \to \overline{\mathbb{R}}$ satisfying condition (i), take $m = k$, $b_j = a_j$, and $B_j = f^{-1}(a_j)$, $j \in \{1, \dots, k\}$. If $x \in B_j$ then we clearly have $f(x) = b_j$ and so $f = \sum_{j=1}^{m} b_j \chi_{B_j}$, as desired.

(ii) $\Longrightarrow$ (iii) Let $f\colon X \to \overline{\mathbb{R}}$ satisfy condition (ii). If $x \in \cup_{j=1}^{m} B_j$ then there exists unique $j_1(x), \dots, j_{r(x)}(x) \in \{1, \dots, m\}$ such that $x \in C(x) \triangleq B_{j_1(x)} \cap \dots \cap B_{j_{r(x)}}(x)$, but $x \notin B_j$ for $j \notin \{j_1(x), \dots, j_{r(x)}(x)\}$. Moreover, $f(x) = b_{j_1(x)} + \dots + b_{j_{r(x)}(x)}$. Since there is a finite number of sets $B_1, \dots, B_m$ there are only finitely many possible intersections of these sets. Thus $\{C(x)\}_{x \in X} = \{C_1, \dots, C_r\}$ for disjoint sets $C_1, \dots, C_r$. Since each of the sets $C_1, \dots, C_r$ is a finite intersection of measurable sets, these sets are measurable. By construction of the sets $C(x)$, $x \in X$, the sets $C_1, \dots, C_m$ are pairwise disjoint. Moreover, the value of f on C_j is constant for $j \in \{1, \dots, r\}$. From these observations we immediately conclude that f satisfies property (iii).

(iii) $\Longrightarrow$ (i) First suppose that $\cup_{j=1}^{r} C_j = X$. Then, given $f\colon X \to \overline{\mathbb{R}}$ satisfying condition (iii), define $k = r$ and $a_j = c_j$, $j \in \{1, \dots, r\}$. Clearly $\operatorname{image}(f) = \{a_1, \dots, a_r\}$ and, since $f^{-1}(a_j)$ is a union of the measurable sets $C_1, \dots, C_r$ (it might be a union in case the numbers $c_1, \dots, c_r$ are not distinct), these sets are measurable. If $\cup_{j=1}^{r} C_j \subset X$ then define $k = r + 1$ and let $C_{r+1} = X \setminus \cup_{j=1}^{r} C_j$ and $c_{r+1} = 0$. The first part of the proof can now be repeated to give the desired conclusion in this case. ■

We now give a function having any of the preceding properties a name.

2.6.38 Definition (Simple function) If $(X, \mathscr{A})$ is a measurable space, a function $f\colon X \to \overline{\mathbb{R}}$ satisfying any one of the three equivalent properties of Proposition 2.6.37 is a ***simple function***. For any subset $I \subseteq \overline{\mathbb{R}}$ (typically we will be concerned with $I \in \{\mathbb{R}, \overline{\mathbb{R}}_{\geq 0}\}$) we denote

$$\mathsf{S}(X; I) = \{f\colon X \to I \mid f \text{ is simple}\},$$

with the understanding that the σ-algebra $\mathscr{A}$ is implicit. •

Simple functions can be thought of playing for the integral on measure spaces the rôle of step functions in the construction of the Riemann integral. For the Riemann integral, Riemann integrable functions are *defined* by their ability to be well approximated by step functions. For the integral defined on measure spaces, there exists a notion, definable only in terms of measurable sets, of a class of functions that are well approximated by simple functions. These are none other than the measurable functions that we have been talking about in this section. The following result illustrates this.

2.6.39 Proposition (Approximations of measurable functions by simple functions) *For a measurable space* $(X, \mathscr{A})$ *and for an* $\mathscr{A}$*-measurable function* $f\colon A \to \overline{\mathbb{R}}$*, the following statements hold:*

(i) *there exists a sequence* $(f_k)_{k\in\mathbb{Z}_{>0}}$ *of simple functions having the property that, for each* $x \in X$*, we have*

$$\lim_{k\to\infty} f_k(x) = f(x);$$

(ii) *if* f *is* $\overline{\mathbb{R}}_{\geq 0}$*-valued, the sequence* $(f_k)_{k\in\mathbb{Z}_{>0}}$ *of part (i) may be chosen so that the functions are* $\mathbb{R}_{\geq 0}$*-valued, and so that, for each* $x \in X$*, the sequence* $(f_k(x))_{k\in\mathbb{Z}_{>0}}$ *is increasing.*

Proof We prove part (ii) first, with (i) then following easily. Thus suppose that $f(x) \geq 0$ for each $x \in X$. Let $k \in \mathbb{Z}_{>0}$. For $j \in \{1, \dots, k2^k\}$, define

$$A_{k,j} = \{x \in X \mid 2^{-k}(j-1) \leq f(x) < 2^{-k}j\}.$$

As f is measurable, each of these sets is measurable (why?). We then define $f_k(x)$ by

$$f_k(x) = \begin{cases} 2^{-k}(j-1), & x \in A_{k,j}, \\ k, & x \in A \setminus (\cup_{j=1}^{k2^k} A_{k,j}). \end{cases}$$

If $f(x) < \infty$ then the sequence $(f_k(x))_{k\in\mathbb{Z}_{>0}}$ converges to $f(x)$ by construction. If $f(x) = \infty$ then $f_k(x) = k$ for all $k \in \mathbb{Z}_{>0}$, and again the sequence converges, i.e., diverges to ∞.

This proves the result when f is positive-valued. If f is not positive-valued, then one writes $f = f_+ - f_-$ where

$$f_+(x) = \max\{f(x), 0\}, \quad f_-(x) = \max\{-f(x), 0\},$$

cf. Corollary 2.6.17. In this case, the preceding argument can be applied to f_+ and f_- separately, giving (i). ■

One can also consider simple functions that are $\mathbb{C}$- or $\mathbb{R}^n$-valued. Let us first consider the vector-valued case.

2.6.40 Definition (Vector-valued simple function) For a measurable space $(X, \mathscr{A})$, a function $\boldsymbol{f}\colon X \to \mathbb{R}^n$ is a ***simple function*** if each of its components $f_j\colon X \to \mathbb{R}$, $j \in \{1, \dots, n\}$, is a simple function. •

The following characterisation of $\mathbb{R}^n$-valued simple functions is then useful.

2.6.41 Proposition (Characterisation of vector-valued simple functions) *For a measurable space* $(X, \mathscr{A})$ *and for* $\mathbf{f}\colon X \to \mathbb{R}^n$*, the following statements are equivalent:*

(i) $\mathbf{f}$ *is a simple function;*

(ii) $\mathrm{image}(\mathbf{f}) = \{\mathbf{a}_1, \dots, \mathbf{a}_k\} \subseteq \mathbb{R}^n$ *and the sets* $\mathbf{f}^{-1}(\mathbf{a}_j)$, $j \in \{1, \dots, k\}$*, are measurable;*

(iii) *there exists* $B_1, \dots, B_m \in \mathscr{A}$ *and* $\mathbf{b}_1, \dots, \mathbf{b}_m \in \mathbb{R}^n$ *such that* $\mathbf{f} = \sum_{j=1}^m \mathbf{b}_j \chi_{B_j}$*;*

(iv) *there exist pairwise disjoint sets* $C_1, \dots, C_r \in \mathscr{A}$ *and* $\mathbf{c}_1, \dots, \mathbf{c}_r \in \mathbb{R}^n$ *such that* $\mathbf{f} = \sum_{j=1}^r \mathbf{c}_j \chi_{C_j}$.

Proof It suffices to show the equivalence of any of the last three statements with the first. The arguments from Proposition 2.6.37 can then be applied to show the equivalence with the other two statements, the only difference being the replacement of $\overline{\mathbb{R}}$ with $\mathbb{R}^n$. We shall show that the first statement is equivalent to the fourth.

First suppose that $\boldsymbol{f}$ is a simple function and write

$$f_j = \sum_{k=1}^{r_j} c_{j,k}\chi_{C_{j,k}},$$

for $c_{j,k} \in \mathbb{R}$ and for pairwise disjoint sets $C_{j,k} \in \mathscr{A}$, $j \in \{1,\dots,n\}$, $k \in \{1,\dots,r_j\}$. Let $x \in X$ and denote

$$C(x) = \cap\{C_{j,k} \mid j \in \{1,\dots,n\},\ k \in \{1,\dots,r_j\},\ x \in C_{j,k}\}.$$

Since there are finitely many sets $C_{j,k}$, $j \in \{1,\dots,n\}$, $k \in \{1,\dots,r_j\}$, it follows that there are finitely many possible intersections of these sets. Therefore, there are pairwise disjoint measurable sets $C_1,\dots,C_r$ such that $\{C(x)\}_{x\in X} = \{C_1,\dots,C_r\}$. Moreover, if $x \in X$ and if $\boldsymbol{f}(x) \neq \boldsymbol{0}$, then $x \in C_l$ for some $l \in \{1,\dots,r\}$. Moreover, since $C_l = C_{j_1,k_1} \cap \dots \cap C_{j_m,k_m}$ for some distinct $j_1,\dots,j_m \in \{1,\dots,n\}$ and some $k_l \in \{1,\dots,r_{j_l}\}$, $l \in \{1 \dots, m\}$, we have

$$f_j(x) = \begin{cases} c_{j_l,k_l}, & j = j_l \text{ for some } l \in \{1,\dots,m\}, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore, taking $\boldsymbol{c}_l$ to be the vector whose jth component is given by the expression on the right above, we have

$$\boldsymbol{f} = \sum_{l=1}^{r} \boldsymbol{c}_l\chi_{C_l},$$

as desired.

Conversely, suppose that $\boldsymbol{f}$ satisfies the fourth condition with

$$\boldsymbol{f} = \sum_{l=1}^{r} \boldsymbol{c}_l\chi_{C_l}.$$

Then

$$f_j = \sum_{l=1}^{r} c_{l,j}\chi_{C_l},$$

where $c_{l,j}$, $j \in \{1,\dots,n\}$, is the jth component of $\boldsymbol{c}_l$, $l \in \{1,\dots,r\}$. This shows that $\boldsymbol{f}$ is a simple function. ■

The same constructions obviously apply to $\mathbb{C}$-valued functions, and we record the constructions here.

2.6.42 Definition ($\mathbb{C}$-valued simple function) For a measurable space $(X, \mathscr{A})$, a function $f\colon X \to \mathbb{C}$ is a ***simple function*** if $\mathrm{Re}(f), \mathrm{Im}(f)\colon X \to \mathbb{R}$ are simple functions. •

2.6.43 Corollary (Characterisation of $\mathbb{C}$-valued simple functions) *For a measurable space* $(X, \mathscr{A})$ *and for* $f\colon X \to \mathbb{C}$, *the following statements are equivalent:*

(i) f *is a simple function;*

(ii) $\operatorname{image}(f) = \{a_1, \dots, a_k\} \subseteq \mathbb{C}$ *and the sets* $f^{-1}(a_j)$, $j \in \{1, \dots, k\}$, *are measurable;*

(iii) *there exists* $B_1, \dots, B_m \in \mathscr{A}$ *and* $b_1, \dots, b_m \in \mathbb{C}$ *such that* $f = \sum_{j=1}^m b_j \chi_{B_j}$;

(iv) *there exist pairwise disjoint sets* $C_1, \dots, C_r \in \mathscr{A}$ *and* $c_1, \dots, c_r \in \mathbb{C}$ *such that* $f = \sum_{j=1}^r c_j \chi_{C_j}$.

One can also use $\mathbb{C}$- or vector-valued simple functions to approximate $\mathbb{C}$- or vector-valued measurable functions.

2.6.44 Proposition (Approximation of vector-valued measurable functions by simple functions) *If* $(X, \mathscr{A})$ *is a measure space and if* $\mathbf{f}\colon X \to \mathbb{R}^n$ *is measurable, then there exists a sequence* $(\mathbf{f}_k)_{k \in \mathbb{Z}_{>0}}$ *of* $\mathbb{R}^n$*-valued simple functions such that*

(i) $\lim_{k\to\infty} \mathbf{f}_k(x) = \mathbf{f}(x)$ *for each* $x \in X$ *and*

(ii) $\|\mathbf{f}_k(x)\|_{\mathbb{R}^n} \le \|\mathbf{f}(x)\|_{\mathbb{R}^n}$ *for each* $x \in X$.

Proof Let $f_1, \dots, f_n$ be the components of $\boldsymbol{f}$. For each $j \in \{1, \dots, n\}$, if we apply the construction of Proposition 2.6.39, we arrive at a sequence $(f_{j,k})_{k \in \mathbb{Z}_{>0}}$ of simple functions for which

1. $\lim_{k\to\infty} f_{j,k}(x) = f_j(x)$ for every $x \in X$ and
2. $|f_{j,k}(x)| \le |f_j(x)|$ for every $x \in X$

(the verification of the second property requires looking for a moment at the particular construction of Proposition 2.6.39. If we take

$$\boldsymbol{f}_k(x) = (f_{1,k}(x), \dots, f_{n,k}(x)), \qquad x \in X,\ k \in \mathbb{Z}_{>0}$$

then one sees easily that the sequence $(\boldsymbol{f}_k)_{k \in \mathbb{Z}_{>0}}$ has the desired properties. ■

Of course, this specialises to the $\mathbb{C}$ case.

2.6.45 Corollary (Approximation of $\mathbb{C}$-valued measurable functions by simple functions) *If* $(X, \mathscr{A})$ *is a measure space and if* $f\colon X \to \mathbb{C}$ *is measurable, then there exists a sequence* $(f_k)_{k \in \mathbb{Z}_{>0}}$ *of* $\mathbb{C}$*-valued simple functions such that*

(i) $\lim_{k\to\infty} f_k(x) = f(x)$ *for each* $x \in X$ *and*

(ii) $|f_k(x)| \le |f(x)|$ *for each* $x \in X$.

2.6.6 Topological characterisations of convergence for sequences of measurable functions[8]

In this section we characterise some of the modes of convergence for sequences of measurable functions in terms of topological constructions. We let $(X, \mathscr{A}, \mu)$ be a measure space. It will be useful to characterise measurable functions as

[8]The results in this section are not used in an essential way elsewhere in the text, except in Sections 2.7.5 and 2.9.11.

equivalence classes of functions that agree up to sets of measure zero. Thus we say that $f, g \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ are ***equivalent*** if

$$\mu(\{x \in X \mid f(x) \neq g(x)\}) = 0.$$

This is readily seen to define an equivalence relation in $\mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ and we denote by $\mathsf{L}^{0}((X, \mathscr{A}); \overline{\mathbb{R}})$ the set of equivalence classes, an equivalence class being denoted by $[f]$ for $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$. The following result shows that convergence pointwise almost everywhere is defined independently of equivalence classes.

2.6.46 Lemma (Almost everywhere pointwise convergence is independent of equivalence) *Let $(X, \mathscr{A}, \mu)$ be a measure space. For a sequence $([f_j])_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{L}^{0}((X, \mathscr{A}); \mathbb{R})$ and for $[f] \in \mathsf{L}^{0}((X, \mathscr{A}); \mathbb{R})$ the following statements are equivalent:*

(i) there exists a sequence $(g_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ and $g \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ such that

- *(a) $[g_j] = [f_j]$ for $j \in \mathbb{Z}_{>0}$,*
- *(b) $[g] = [f]$, and*
- *(c) $(g_j)_{j\in\mathbb{Z}_{>0}}$ converges pointwise almost everywhere to g.*

(ii) for every sequence $(g_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ and for every $g \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ satisfying

- *(a) $[g_j] = [f_j]$ for $j \in \mathbb{Z}_{>0}$ and*
- *(b) $[g] = [f]$,*

it holds that $(g_j)_{j\in\mathbb{Z}_{>0}}$ converges pointwise almost everywhere to g.

Proof It is clear that the second statement implies the first, so we only prove the converse. Thus we let $(g_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ and $g \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ be such that

1. $[g_j] = [f_j]$ for $j \in \mathbb{Z}_{>0}$,
2. $[g] = [f]$, and
3. $(g_j)_{j\in\mathbb{Z}_{>0}}$ converges pointwise almost everywhere to g.

Let $(h_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ and let $h \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ be such that

1. $[h_j] = [f_j]$ for $j \in \mathbb{Z}_{>0}$ and
2. $[h] = [f]$.

Define

$$A = \{x \in X \mid g(x) \neq f(x)\}, \quad B = \{x \in X \mid h(x) \neq f(x)\}$$

and, for $j \in \mathbb{Z}_{>0}$, define

$$A_j = \{x \in X \mid g_j(x) \neq f_j(x)\}, \quad B_j = \{x \in X \mid h_j(x) \neq f_j(x)\}$$

and note that

$$x \in X \setminus (A \cup B) = (X \setminus A) \cap (X \setminus B) \quad \Longrightarrow \quad h(x) = f(x) = g(x)$$

and

$$x \in X \setminus (A_j \cup B_j) = (X \setminus A_j) \cap (X \setminus B_j) \quad \Longrightarrow \quad h_j(x) = f_j(x) = g_j(x).$$

Thus,

$$x \in X \setminus \Big((\cup_{j\in\mathbb{Z}_{>0}} A_j \cup B_j) \cup (A \cup B)\Big) \quad \implies \quad \lim_{j\to\infty} h_j(x) = \lim_{j\to\infty} g_j(x) = g(x) = h(x).$$

Since $(\cup_{j\in\mathbb{Z}_{>0}} A_j \cup B_j) \cup (A \cup B)$ is a countable union of sets of measure zero, it has zero measure, and so $(h_j)_{j\in\mathbb{Z}_{>0}}$ converges pointwise almost everywhere to h. ■

With the preceding lemma, the following definition makes sense.

2.6.47 Definition (Almost everywhere convergence of sequences of equivalence classes of functions) Let $(X, \mathscr{A}, \mu)$ be a measure space, let $([f_j])_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{L}^0((X,\mathscr{A});\overline{\mathbb{R}})$, and let $[f] \in \mathsf{L}^0((X,\mathscr{A});\overline{\mathbb{R}})$. The sequence $([f_j])_{j\in\mathbb{Z}_{>0}}$ ***converges pointwise almost everywhere*** to $[f]$ if

$$\mu\Big(X \setminus \Big\{x \in X \;\Big|\; f(x) = \lim_{j\to\infty} f_j(x)\Big\}\Big) = 0. \qquad \bullet$$

We begin by indicating that the convergence defined by almost everywhere pointwise convergence cannot arise from a topology.

2.6.48 Proposition (Almost everywhere pointwise convergence is not always topological) *Let $(X, \mathscr{A}, \mu)$ be a measure space and let $\mathscr{T}_{\mathrm{a.e.}}$ be the set of topologies τ on $\mathsf{L}^0((X,\mathscr{A});\mathbb{R})$ such that the convergent sequences in τ are precisely the almost everywhere pointwise convergent sequences. If there exists a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{L}^{(0)}((X,\mathscr{A});\mathbb{R})$ and $f \in \mathsf{L}^{(0)}((X,\mathscr{A});\mathbb{R})$ such that $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges in measure to f but does not converge pointwise almost everywhere to f, then $\mathscr{T}_{\mathrm{a.e.}} = \emptyset$.*

Proof Let us denote by $z \in \mathsf{L}^{(0)}((X,\mathscr{A});\mathbb{R})$ the zero function. The hypotheses ensure that the sequence $(g_j \triangleq f_j - f)_{j\in\mathbb{Z}_{>0}}$ converges to z in measure, but does not converges pointwise almost everywhere to z. Suppose that $\mathscr{T}_{\mathrm{a.e.}} \neq \emptyset$ and let $\tau \in \mathscr{T}_{\mathrm{a.e.}}$. Since almost everywhere pointwise convergence agrees with convergence in τ, there exists a neighbourhood U of $[z]$ in $\mathsf{L}^0((X,\mathscr{A});\mathbb{R})$ such that the set

$$\{j \in \mathbb{Z}_{>0} \mid [f_j] \in U\}$$

is finite. By Proposition 2.6.26 there exists a subsequence $(f_{j_k})_{k\in\mathbb{Z}_{>0}}$ of $(f_j)_{j\in\mathbb{Z}_{>0}}$ that converges pointwise almost everywhere to z. Thus the sequence $([f_{j_k}])_{k\in\mathbb{Z}_{>0}}$ converges pointwise almost everywhere to $[z]$, and so converges to $[z]$ in τ. Thus, in particular, the set

$$\{k \in \mathbb{Z}_{>0} \mid [f_{j_k}] \in U\}$$

is infinite, which is a contradiction. ■

In particular, we have the following result which shows that in the most common situation where one wishes to study almost everywhere pointwise convergence, this sort of convergence is not topological.

2.6.49 Corollary (Almost everywhere pointwise convergence is not topological for the Lebesgue measure) *Let $\mathscr{T}_{\mathrm{a.e.}}$ be the set of topologies τ on $\mathsf{L}^0((\mathbb{R}^n, \mathscr{L}(\mathbb{R}^n)); \mathbb{R})$ such that the convergent sequences in τ are precisely the almost everywhere pointwise convergent sequences using the Lebesgue measure on $\mathbb{R}^n$. Then $\mathscr{T}_{\mathrm{a.e.}} = \emptyset$.*

Proof In Example 2.6.25 we have seen that there exists a sequence in $\mathsf{L}^{(0)}((\mathbb{R}, \mathscr{L}(\mathbb{R})); \mathbb{R})$ that converges in measure but does not converge pointwise almost everywhere. This example is easily adapted to $\mathsf{L}^{(0)}((\mathbb{R}^n, \mathscr{L}(\mathbb{R}^n)); \mathbb{R})$, and the result then follows from Proposition 2.6.48. ■

Now one can ask if there is a framework in which almost everywhere pointwise convergence can be studied. Indeed there is such a framework. The construction relies on notions concerning filters and nets from . where?

2.6.50 Definition (Limit structure) A ***limit structure*** on a set S is a subset $\mathscr{L} \subseteq \mathscr{F}(S) \times S$ with the following properties:

(i) if $x \in S$ then $(\mathcal{F}_x, x) \in \mathscr{L}$;
(ii) if $(\mathcal{F}, x) \in \mathscr{L}$ and if $\mathcal{F} \subseteq \mathcal{G} \in \mathscr{F}(S)$ then $(\mathcal{G}, x) \in \mathscr{L}$;
(iii) if $(\mathcal{F}, x), (\mathcal{G}, x) \in \mathscr{L}$ then $(\mathcal{F} \cap \mathcal{G}, x) \in \mathscr{L}$.

If $(\Lambda, \preceq)$ is a directed set, a Λ-net $\phi\colon \Lambda \to S$ is ***$\mathscr{L}$-convergent*** to $x \in S$ if $(\mathcal{F}_\phi, x) \in \mathscr{L}$. Let us denote by $\mathscr{S}(\mathscr{L})$ the set of $\mathscr{L}$-convergent $\mathbb{Z}_{>0}$-nets, i.e., the set of $\mathscr{L}$-convergent sequences. •

The intuition behind the notion of a limit structure is as follows. Condition (i) says that the trivial filter converging to x should be included in the limit structure, condition (ii) says that if a filter converges to x, then every coarser filter also converges to x, and condition (iii) says that "mixing" filters converging to x should give a filter converging to x. Starting from the definition of a limit structure, one can reproduce many of the concepts from topology, e.g., openness, closedness, compactness, continuity.

We are interested in the special case of limit structures on a vector space V. We suppose that V is defined over a field F. For $\mathcal{F}, \mathcal{G} \in \mathscr{F}(\mathsf{V})$ and for $a \in \mathsf{F}$ we denote

$$\mathcal{F} + \mathcal{G} = \{A + B \mid A \in \mathcal{F},\ B \in \mathcal{G}\}, \quad a\mathcal{F} = \{aA \mid A \in \mathcal{F}\},$$

where, as usual,

$$A + B = \{u + v \mid u \in A,\ v \in B\}, \quad aA = \{au \mid u \in A\}.$$

We say that a limit structure $\mathscr{L}$ on a vector space V is ***linear*** if $(\mathcal{F}_1, v_1), (\mathcal{F}_2, v_2) \in \mathscr{L}$ implies that $(\mathcal{F}_1 + \mathcal{F}_2, v_1 + v_2) \in \mathscr{L}$ and if $a \in \mathsf{F}$ and $(\mathcal{F}, v) \in \mathscr{L}$ then $(a\mathcal{F}, av) \in \mathscr{L}$.

For $[f] \in \mathsf{L}^0((X, \mathscr{A}); \mathbb{R})$ define

$$\mathscr{F}_{[f]} = \{\mathcal{F} \in \mathscr{F}(\mathsf{L}^0((X, \mathscr{A}); \mathbb{R})) \mid \mathcal{F}_\phi \subseteq \mathcal{F} \text{ for some } \mathbb{Z}_{>0}\text{-net } \phi \text{ such that } (\phi(j))_{j \in \mathbb{Z}_{>0}} \text{ is almost everywhere pointwise convergent to } [f]\}.$$

We may now define a limit structure on $\mathsf{L}^0((X, \mathscr{A}); \mathbb{R})$ as follows.

2.6.51 Theorem (Almost everywhere pointwise convergence is defined by a limit structure) *The subset of* $\mathscr{F}(\mathsf{L}^0((\mathsf{X},\mathscr{A});\mathbb{R})) \times \mathsf{L}^0((\mathsf{X},\mathscr{A});\mathbb{R})$ *defined by*

$$\mathscr{L}_\mu = \{(\mathcal{F}, [\mathrm{f}]) \mid \mathcal{F} \in \mathscr{F}_{[\mathrm{f}]}\}$$

is a linear limit structure on $\mathsf{L}^0((\mathsf{X},\mathscr{A});\mathbb{R})$. *Moreover, a sequence* $([\mathrm{f}_j])_{j\in\mathbb{Z}_{>0}}$ *is* $\mathscr{L}_\mu$*-convergent to* $[\mathrm{f}]$ *if and only if the sequence is almost everywhere pointwise convergent to* $[\mathrm{f}]$.

Proof Let $[f] \in \mathsf{L}^0((X,\mathscr{A});\mathbb{R})$. Consider the trivial $\mathbb{Z}_{>0}$-net $\phi_{[f]}\colon \mathbb{Z}_{>0} \to \mathsf{L}^0((X,\mathscr{A});\mathbb{R})$ defined by $\phi_{[f]}(j) = [f]$. Since $\mathcal{F}_\phi = \mathcal{F}_{[f]}$ and since $(\mathcal{F}_\phi, [f]) \in \mathscr{L}_\mu$, the condition (i) for a limit structure is satisfied.

Let $(\mathcal{F}, [f]) \in \mathscr{L}_\mu$ and suppose that $\mathcal{F} \subseteq \mathcal{G}$. Then $\mathcal{F} \in \mathscr{F}_{[f]}$ and so $\mathcal{F} \supseteq \mathcal{F}_\phi$ for some $\mathbb{Z}_{>0}$-net ϕ that converges pointwise almost everywhere to $[f]$. Therefore, we immediately have $\mathcal{F}_\phi \subseteq \mathcal{G}$ and so $(\mathcal{G}, [f]) \in \mathscr{L}_\mu$. This verifies condition (ii) in the definition of a limit structure.

Finally, let $(\mathcal{F},[f]), (\mathcal{G},[f]) \in \mathscr{L}_\mu$ and let ϕ and ψ be $\mathbb{Z}_{>0}$-nets that converge pointwise almost everywhere to $[f]$ and satisfy $\mathcal{F}_\phi \subseteq \mathcal{F}$ and $\mathcal{F}_\psi \subseteq \mathcal{G}$. Define a $\mathbb{Z}_{>0}$-net $\phi \wedge \psi$ by

$$\phi \wedge \psi(j) = \begin{cases} \phi(\frac{1}{2}(j+1)), & j \text{ odd},\\ \psi(\frac{1}{2}j), & j \text{ even}.\end{cases}$$

We first claim that $\phi \wedge \psi$ converges pointwise almost everywhere to $[f]$. Let

$$A = \Big\{x \in X \;\Big|\; \lim_{j\to\infty} \phi(j)(x) \neq f(x)\Big\}, \quad B = \Big\{x \in X \;\Big|\; \lim_{j\to\infty} \psi(j)(x) \neq f(x)\Big\}.$$

If $x \in X \setminus (A \cup B)$ then

$$\lim_{j\to\infty} \phi(j)(x) = \lim_{j\to\infty} \psi(j)(x) = f(x).$$

Thus, for $x \in X \setminus (A \cup B)$ and $\epsilon \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{Z}_{>0}$ such that

$$|f(x) - \phi(j)(x)|, |f(x) - \psi(j)(x)| < \epsilon, \qquad j \geq N.$$

Therefore, for $j \geq 2N$ and for $x \in X \setminus (A \cup B)$ we have $|f(x) - \phi \wedge \psi(j)(x)| < \epsilon$ and so

$$\lim_{j\to\infty} \phi \wedge \psi(j)(x) = f(x), \qquad x \in X \setminus (A \cup B).$$

Since $\mu(A \cup B) = 0$ it indeed follows that $\phi \wedge \psi$ converges pointwise almost everywhere to $[f]$.

We next claim that $\mathcal{F}_{\phi\wedge\psi} \subseteq \mathcal{F} \cap \mathcal{G}$. Indeed, let $S \in \mathcal{F}_{\phi\wedge\psi}$. Then there exists $N \in \mathbb{Z}_{>0}$ such that $T_{\phi\wedge\psi}(N) \subseteq S$. Therefore, there exists $N_\phi, N_\psi \in \mathbb{Z}_{>0}$ such that $T_\phi(N_\phi) \subseteq S$ and $T_\psi(N_\psi) \subseteq S$. That is, $S \in \mathcal{F}_\phi \cap \mathcal{F}_\psi \subseteq \mathcal{F} \cap \mathcal{G}$. This shows that $(\mathcal{F} \cap \mathcal{G}, [f]) \in \mathscr{L}_\mu$ and so shows that condition (iii) in the definition of a limit structure holds.

Thus we have shown that $\mathscr{L}_\mu$ is a limit structure. Let us show that it is a linear limit structure. Let $(\mathcal{F}_1, [f_1]), (\mathcal{F}_2, v_2) \in \mathscr{L}_\mu$. Thus there exists $\mathbb{Z}$-nets ϕ_1 and ϕ_2 in $\mathsf{L}^0((X,\mathscr{A});\mathbb{R})$ converging pointwise almost everywhere to $[f_1]$ and $[f_2]$, respectively, and such that $\mathcal{F}_{\phi_1} \subseteq \mathcal{F}_1$ and $\mathcal{F}_{\phi_2} \subseteq \mathcal{F}_2$. Let us denote by $(f_{1,j})_{j\in\mathbb{Z}_{>0}}$ and $(f_{2,j})_{j\in\mathbb{Z}_{>0}}$

sequences in $\mathsf{L}^{(0)}((X,\mathscr{A});\mathbb{R})$ such that $[f_{1,j}]=\phi_1(j)$ and $[f_{2,j}]=\phi_2(j)$ for $j\in\mathbb{Z}_{>0}$. Then, as in the proof of Lemma 2.6.46, there exists a subset $A\subseteq X$ of zero measure such that

$$\lim_{j\to\infty} f_{j,1}(x)=f_1(x),\quad \lim_{j\to\infty} f_{2,j}(x)=f_2(x),\qquad x\in X\setminus A.$$

Thus, for $x\in X\setminus A$,

$$\lim_{j\to\infty}(f_{1,j}+f_{2,j})(x)=(f_1+f_2)(x).$$

This shows that the $\mathbb{Z}_{>0}$-net $\phi_1+\phi_2$ converges pointwise almost everywhere to $[f_1+f_2]$. Since $\mathscr{F}_{\phi_1+\phi_2}\subseteq\mathscr{F}_1+\mathscr{F}_2$, it follows that $(\mathscr{F}_1+\mathscr{F}_2,[f_1+f_2])\in\mathscr{L}_\mu$. An entirely similarly styled argument gives $(a\mathscr{F},av)\in\mathscr{L}_\mu$ for $(\mathscr{F},v)\in\mathscr{L}_\mu$.

We now need to show that $\mathscr{S}(\mathscr{L}_\mu)$ consists exactly of the almost everywhere pointwise convergent sequences. The very definition of $\mathscr{L}_\mu$ ensures that if a $\mathbb{Z}_{>0}$-net ϕ is almost everywhere pointwise convergent then $\phi\in\mathscr{S}(\mathscr{L}_\mu)$. We prove the converse, and so let ϕ be $\mathscr{L}_\mu$-convergent to $[f]$. Therefore, by definition of $\mathscr{L}_\mu$, there exists a $\mathbb{Z}_{>0}$-net ψ converging pointwise almost everywhere to $[f]$ such that $\mathscr{F}_\psi\subseteq\mathscr{F}_\phi$.

1 **Lemma** *There exists of a subsequence ψ' of ψ such that $\mathscr{F}_{\psi'}=\mathscr{F}_\phi$.*

Proof Let $n\in\mathbb{Z}_{>0}$ and note that $T_\psi(n)\in\mathscr{F}_\psi\subseteq\mathscr{F}_\phi$. Thus there exists $k\in\mathbb{Z}_{>0}$ such that $T_\phi(k)\subseteq T_\psi(n)$. Then define

$$k_n=\min\{k\in\mathbb{Z}_{>0}\mid T_\phi(k)\subseteq T_\psi(n)\},$$

the minimum being well-defined since

$$k>k'\quad\implies\quad T_\phi(k)\subseteq T_\phi(k').$$

This uniquely defines, therefore, a sequence $(k_n)_{n\in\mathbb{Z}_{>0}}$. Moreover, if $n_1>n_2$ then $T_\psi(n_2)\subseteq T_\psi(n_1)$ which implies that $T_\phi(k_{n_2})\subseteq T_\psi(n_1)$. Therefore, $k_{n_2}\geq k_{n_1}$, showing that the sequence $(k_n)_{n\in\mathbb{Z}_{>0}}$ is nondecreasing.

Now define $\theta\colon\mathbb{Z}_{>0}\to\mathbb{Z}_{>0}$ as follows. If $j<k_n$ for every $n\in\mathbb{Z}_{>0}$ then define $\theta(j)$ in an arbitrary manner. If $j\geq k_1$ then note that $\phi(j)\in T_\phi(k_1)\subseteq T_\psi(1)$. Thus there exists (possibly many) $m\in\mathbb{Z}_{>0}$ such that $\phi(j)=\psi(m)$. If $j\geq k_n$ for $n\in\mathbb{Z}_{>0}$ then there exists (possibly many) $m\geq n$ such that $\phi(j)=\psi(m)$. Thus for any $j\in\mathbb{Z}_{>0}$ we can define $\theta(j)\in\mathbb{Z}_{>0}$ such that $\phi(j)=\psi(\theta(j))$ if $j\geq k_1$ and such that $\theta(j)\geq n$ if $j\geq k_n$.

Note that any function $\theta\colon\mathbb{Z}_{>0}\to\mathbb{Z}_{>0}$ as constructed above is unbounded. Therefore, there exists a strictly increasing function $\rho\colon\mathbb{Z}_{>0}\to\mathbb{Z}_{>0}$ such that $\mathrm{image}(\rho)=\mathrm{image}(\theta)$. We claim that $\mathscr{F}_\rho=\mathscr{F}_\theta$. First let $n\in\mathbb{Z}_{>0}$ and let $j\geq k_{\rho(n)}$. Then $\theta(j)\geq\rho(n)$. Since $\mathrm{image}(\rho)=\mathrm{image}(\theta)$ there exists $m\in\mathbb{Z}_{>0}$ such that $\rho(m)=\theta(j)\geq\rho(n)$. Since ρ is strictly increasing, $m\geq n$. Thus $\theta(j)\in T_\rho(n)$ and so $T_\theta(k_{\rho(n)})\subseteq T_\rho(n)$. This implies that $\mathscr{F}_\rho\subseteq\mathscr{F}_\theta$.

Conversely, let $n\in\mathbb{Z}_{>0}$ and let $r_n\in\mathbb{Z}_{>0}$ be such that

$$\rho(r_n)>\max\{\theta(1),\dots,\theta(n)\};$$

this is possible since ρ is unbounded. If $j\geq r_n$ then

$$\rho(j)\geq\rho(r_n)>\max\{\theta(1),\dots,\theta(n)\}.$$

Since image(ρ) = image(θ) we have $\rho(j) = \theta(m)$ for some $m \in \mathbb{Z}_{>0}$. We must have $m > n$ and so $\rho(j) \in T_\theta(n)$. Thus $T_\rho(r_n) \subseteq T_\theta(n)$ and so $\mathcal{F}_\theta \subseteq \mathcal{F}_\rho$.

To arrive at the conclusions of the lemma we first note that, by definition of θ, $\mathcal{F}_\phi = \mathcal{F}_{\psi\circ\theta}$. We now define $\psi' = \psi \circ \rho$ and note that

$$\mathcal{F}_\phi = \mathcal{F}_{\psi\circ\theta} = \psi(\mathcal{F}_\theta) = \psi(\mathcal{F}_\rho) = \mathcal{F}_{\psi\circ\rho},$$

as desired. ▼

Since a subsequence of an almost everywhere pointwise convergent sequence is almost everywhere pointwise convergent to the same limit, it follows that ψ', and so ϕ, converges almost everywhere pointwise to $[f]$. ■

what?

Note that we have already seen in Sections 1.9.1 and 1.9.2 that pointwise and uniform convergence is prescribed by a topology. We shall see in that convergence in measure is topological.

Exercises

2.6.1 Let $(X, \mathscr{A}, \mu)$ be a measure space that is not complete. Show that Proposition 2.6.10 fails in this case.

2.6.2 Let $(X, \mathscr{A}, \mu)$ be a measure space that is not complete. Show that Corollary 2.6.19 fails in this case.

2.6.3 Let $(X, \mathscr{A})$ be a measurable space and let $A, B \in \mathscr{A}$ be such that $X = A \mathbin{\mathring{\cup}} B$. Let $f_A \colon A \to \overline{\mathbb{R}}$ be $\mathscr{A}_A$-measurable and let $f_B \colon B \to \overline{\mathbb{R}}$ be $\mathscr{A}_B$-measurable. Show that $f \colon X \to \overline{\mathbb{R}}$ defined by

$$f(x) = \begin{cases} f_A(x), & x \in A, \\ f_B(x), & x \in B \end{cases}$$

is $\mathscr{A}$-measurable.

2.6.4 Give an example of a measure space $(X, \mathscr{A}, \mu)$, a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$, and a function $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ such that $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges pointwise almost everywhere to f, but does not converges pointwise to f.

Section 2.7

Integration on measure spaces

Up to now, we have studied measurable and measure spaces in some detail. These subjects certainly have some value in their own right, particularly in the domain of probability theory which we discuss in . In particular, the properties of the Lebesgue measure on $\mathbb{R}$ and $\mathbb{R}^n$ considered in Sections 2.4 and 2.5 are substantially useful. Following our discussion of measure, we introduced a particular class of functions on measurable spaces called measurable functions. While we showed in Sections 2.9.1 and 2.10.1 that for the Lebesgue measure that these functions are not too far from easily understood functions such as step or continuous functions, the importance of measurable functions is perhaps not so easily understood. What we see in this section is that these functions form the basis for a powerful and general theory of integration. For the Lebesgue measure, this construction of the integral generalises the Riemann integral, and repairs some of the defects of the latter as seen in Section 2.1. where?

The treatment of the integral is as easily carried out in the general setting of a general measure space as it is for the specific case of the Lebesgue integral in particular. Thus we do much of the work in this general setting. In Sections 2.9 and 2.10 we consider the Lebesgue integral, but only its particular properties that rely on the structure of Lebesgue measure. Thus a reader wanting only to learn about the Lebesgue integral will have to learn it here. A reader only believing they are interested in Lebesgue integration will have to be satisfied by mentally making the replacement of "$(X, \mathscr{A}, \mu)$" with "$(\mathbb{R}, \mathscr{L}(\mathbb{R}), \lambda)$" or "$(\mathbb{R}^n, \mathscr{L}(\mathbb{R}^n), \lambda_n)$."

Do I need to read this section? Clearly if you are reading this chapter, then you must read this section. •

2.7.1 Definition of the integral

We consider a measure space $(X, \mathscr{A}, \mu)$. The objective is to define the integral of a measurable function $f: X \to \overline{\mathbb{R}}$. We do this in three stages.

Integration of nonnegative simple functions

Let $f \in \mathsf{S}(X; \overline{\mathbb{R}}_{\geq 0})$ be written as $f = \sum_{j=1}^{k} a_j \chi_{A_j}$ for a partition $(A_1, \dots, A_k)$ of X into measurable sets. Let us first make an observation concerning the fact that the numbers $a_1, \dots, a_k$ and the sets $A_1, \dots, A_k$ are not uniquely prescribed by f.

2.7.1 Proposition (Independence of integral of simple functions on partition) *For a measure space* $(\mathsf{X}, \mathscr{A}, \mu)$ *suppose that* $\mathrm{f} \in \mathsf{S}(\mathsf{X}; \overline{\mathbb{R}}_{\geq 0})$ *satisfies*

$$\mathrm{f} = \sum_{\mathrm{j}=1}^{\mathrm{k}} \mathrm{a_j} \chi_{\mathrm{A_j}} = \sum_{\mathrm{l}=1}^{\mathrm{m}} \mathrm{b_l} \chi_{\mathrm{B_l}}$$

for $\mathrm{a}_1, \dots, \mathrm{a_k}, \mathrm{b}_1, \dots, \mathrm{b_m} \in \overline{\mathbb{R}}_{\geq 0}$ *and* $\mathrm{A}_1, \dots, \mathrm{A_k} \in \mathscr{A}$ *disjoint and* $\mathrm{B}_1, \dots, \mathrm{B_m} \in \mathscr{A}$ *disjoint. Then*

$$\sum_{\mathrm{j}=1}^{\mathrm{k}} \mathrm{a_j} \mu(\mathrm{A_j}) = \sum_{\mathrm{l}=1}^{\mathrm{m}} \mathrm{b_l} \mu(\mathrm{B_l}).$$

Proof Without loss of generality we suppose that none of $a_1, \dots, a_k$ and $b_1, \dots, b_m$ are zero. It therefore follows that $\cup_{j=1}^k A_j = \cup_{l=1}^m B_l$. Note that if $A_j \cap B_m \neq \emptyset$ for some $j \in \{1, \dots, k\}$ and $l \in \{1, \dots, m\}$, it follows that $a_j = b_l$. Therefore, we have

$$\sum_{j=1}^{k} a_j \mu(A_j) = \sum_{j=1}^{k} \sum_{l=1}^{m} a_j \mu(A_j \cap B_l) = \sum_{l=1}^{m} \sum_{j=1}^{k} b_l \mu(B_l \cap A_j) = \sum_{l=1}^{m} b_l \mu(B_l),$$

as desired. ■

Given the preceding result, the following definition makes sense.

2.7.2 Definition (Integral of nonnegative simple function) For a measure space $(X, \mathscr{A}, \mu)$ and for $f \in \mathsf{S}(X; \overline{\mathbb{R}}_{\geq 0})$ given by $f = \sum_{j=1}^k a_j \chi_{A_j}$ for a partition $(A_1, \dots, A_k)$ of X into measurable sets, the ***integral*** of f is

$$\int_X f \, \mathrm{d}\mu = \sum_{j=1}^{k} a_j \mu(A_j).$$ •

Note that the notion of integral for a simple function is a natural adaptation of the notion of integral for a step function in our development of the Riemann integral in Sections I-3.4 and II-1.6.

Let us give some examples of simple functions and their integrals.

2.7.3 Examples (Positive simple functions and their integrals)

1. Let $P = (I_1, \dots, I_k)$ be a partition of $[a, b] \subseteq \mathbb{R}$ with endpoints $\mathrm{EP}(P) = (x_0, x_1, \dots, x_k)$ and let $f \colon [a, b] \to \mathbb{R}$ be a step function taking value c_j on the interval I_j, $j \in \{1, \dots, k\}$. Clearly then, f is also a simple function since intervals are measurable. Moreover,

$$\int_{[a,b]} f \, \mathrm{d}\lambda = \int_a^b f(x) \, \mathrm{d}x = \sum_{j=1}^{k} c_j (x_j - x_{j-1}),$$

since the Lebesgue measure of an interval is its length.

2. Let us consider the measure space $(\mathbb{R}, \mathscr{L}(\mathbb{R}), \lambda)$ and take $A = \mathbb{Q}$. By Exercise I-2.5.10 it follows that $\lambda(A) = 0$. Therefore, the simple function χ_A has measure zero.
3. Let X be a set, let $\mathscr{A} = \mathbf{2}^X$, and let μ_Σ be the counting measure on X; see Example 2.3.9–3. Let $A_1, \dots, A_k \subseteq X$ be nonempty disjoint subsets, let $a_1, \dots, a_k \in \overline{\mathbb{R}}_{\geq 0}$, and define $f = \sum_{j=1}^k a_j \chi_{A_j}$. If $\mathrm{card}(A_j) = \infty$ for any $j \in \{1, \dots, k\}$ for which $a_j \neq 0$ or if $a_j = \infty$ for any $j \in \{1, \dots, k\}$, then $\int_X f \, \mathrm{d}\mu_\Sigma = \infty$. Otherwise,

$$\int_X f \, \mathrm{d}\mu_\Sigma = \sum_{j=1}^k a_j \, \mathrm{card}(A_j). \qquad \bullet$$

Integration of nonnegative measurable functions

Using the definition of the integral for simple functions, it is possible to immediately deduce a definition of the integral for nonnegative-valued functions. This is done as follows.

2.7.4 Definition (Integral of a nonnegative measurable function) For a measure space $(X, \mathscr{A}, \mu)$ and for $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$, the ***integral*** of f is

$$\int_X f \, \mathrm{d}\mu = \sup \left\{ \int_X g \, \mathrm{d}\mu \;\middle|\; g \in \mathsf{S}(X; \overline{\mathbb{R}}_{\geq 0}) \text{ satisfies } 0 \leq g(x) \leq f(x) \text{ for } x \in X \right\}. \qquad \bullet$$

The following result gives a useful characterisation of the integral of nonnegative-valued functions. It also gives an idea of why measurable functions are the “right” class of functions to integrate, since they are well-approximated by simple functions.

2.7.5 Proposition (Sequential characterisation of the integral for nonnegative functions) *Let $(X, \mathscr{A}, \mu)$ be a measure space, let $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$, and let $(f_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence of increasing positive simple functions converging to f as in Proposition 2.6.39. Then*

$$\int_X f \, \mathrm{d}\mu = \lim_{j \to \infty} \int_X f_j \, \mathrm{d}\mu.$$

Proof First we prove the result in the case that f is a simple function.

1 Lemma *Let $(X, \mathscr{A}, \mu)$ be a measure space, let $f \in \mathsf{S}(X; \overline{\mathbb{R}}_{\geq 0})$, and let $(f_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence of increasing positive simple functions converging to f as in Proposition 2.6.39. Then*

$$\int_X f \, \mathrm{d}\mu = \lim_{j \to \infty} \int_X f_j \, \mathrm{d}\mu.$$

Proof By Exercise 2.7.1 the sequence $(\int_X f_j \, \mathrm{d}\mu)_{j \in \mathbb{Z}_{>0}}$ is increasing and bounded above by $\int_X f \, \mathrm{d}\mu$. Thus the sequence $(\int_X f_j \, \mathrm{d}\mu)_{j \in \mathbb{Z}_{>0}}$ converges in $\overline{\mathbb{R}}_{\geq 0}$, by Theorem I-2.3.8 if the limit is finite, tautologically otherwise. Thus we have

$$\lim_{j\to\infty}\int_X f_j\,\mathrm{d}\mu \leq \int_X f\,\mathrm{d}\mu.$$

Next let $\epsilon \in (0,1)$. Let us write $f = \sum_{l=1}^m a_l\chi_{A_l}$ for $a_1,\dots,a_m \in \overline{\mathbb{R}}_{\geq 0}$ and disjoint $A_1,\dots,A_m \in \mathscr{A}$. For $l \in \{1,\dots,m\}$ and $j \in \mathbb{Z}_{>0}$ denote

$$A_{j,l} = \{x \in A_l \mid f_j(x) \geq (1-\epsilon)a_l\},$$

noting that $A_{j,l} \in \mathscr{A}$ since f_j is measurable. Since the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ is monotonically increasing, the sequence $(A_{j,l})_{j\in\mathbb{Z}_{>0}}$ satisfies

$$A_{j,l} \subseteq A_{j+1,l}, \qquad \cup_{j\in\mathbb{Z}_{>0}} A_{j,l} = A_l.$$

Let us define simple functions

$$g_j = \sum_{l=1}^m (1-\epsilon)a_l\chi_{A_{j,l}}, \qquad j \in \mathbb{Z}_{>0}.$$

By Proposition 2.3.3 we have

$$\lim_{j\to\infty}\int_X g_j\,\mathrm{d}\mu = \lim_{j\to\infty}\sum_{l=1}^m (1-\epsilon)a_l\mu(A_{j,l}) = \sum_{l=1}^m (1-\epsilon)a_l\mu(A_l) = (1-\epsilon)\int_X f\,\mathrm{d}\mu.$$

Since $g_j(x) \leq f_j(x)$ for every $j \in \mathbb{Z}_{>0}$, by Exercise 2.7.1 we have

$$\begin{aligned} & \int_X g_j\,\mathrm{d}\mu \leq \int_X f_j\,\mathrm{d}\mu \\ \implies\quad & \lim_{j\to\infty}\int_X g_j\,\mathrm{d}\mu \leq \lim_{j\to\infty}\int_X f_j\,\mathrm{d}\mu \\ \implies\quad & (1-\epsilon)\int_X f\,\mathrm{d}\mu \leq \lim_{j\to\infty}\int_X f_j\,\mathrm{d}\mu \leq \int_X f\,\mathrm{d}\mu. \end{aligned}$$

Since ϵ is arbitrary, this implies that

$$\lim_{j\to\infty}\int_X f_j\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\mu,$$

as desired. ▼

In the case that f is a general nonnegative-valued measurable function, we note that

$$\int_X f_j\,\mathrm{d}\mu \leq \int_X f_{j+1}\,\mathrm{d}\mu, \qquad j \in \mathbb{Z}_{>0},$$

and

$$\int_X f_j\,\mathrm{d}\mu \leq \int_X f\,\mathrm{d}\mu, \qquad j \in \mathbb{Z}_{>0}.$$

Thus the sequence $(\int_X f_j \,\mathrm{d}\mu)_{j\in\mathbb{Z}_{>0}}$ converges in $\overline{\mathbb{R}}_{\geq 0}$ to a limit (by Theorem I-2.3.8 if the limit is finite, tautologically otherwise) and this limit satisfies

$$\lim_{j\to\infty} \int_X f_j \,\mathrm{d}\mu \leq \int_X f \,\mathrm{d}\mu.$$

Next let $\epsilon \in \mathbb{R}_{>0}$ and let $g \in \mathsf{S}(X; \overline{\mathbb{R}}_{\geq 0})$ be such that

$$\int_X g \,\mathrm{d}\mu \geq (1-\epsilon) \int_X f \,\mathrm{d}\mu.$$

Define $g_j(x) = \min\{g(x), f_j(x)\}$, and note that g_j is a nonnegative simple function, and that the sequence $(g_j(x))_{j\in\mathbb{Z}_{>0}}$ converges to $g(x)$ for each $x \in X$. By the lemma above we thus have

$$\lim_{j\to\infty} \int_X g_j \,\mathrm{d}\mu = \int_X g \,\mathrm{d}\mu.$$

By Exercise 2.7.1 we have

$$\int_X g_j \,\mathrm{d}\mu \leq \int_X f_j \,\mathrm{d}\mu \quad\Longrightarrow\quad \int_X g \,\mathrm{d}\mu \leq \lim_{j\to\infty} \int_X f_j \,\mathrm{d}\mu$$

which gives

$$(1-\epsilon) \int_X f \,\mathrm{d}\mu \leq \int_X g \,\mathrm{d}\mu \leq \lim_{j\to\infty} \int_X f_j \,\mathrm{d}\mu \leq \int_X f \,\mathrm{d}\mu,$$

which gives

$$\lim_{j\to\infty} \int_X f_j \,\mathrm{d}\mu = \int_X f \,\mathrm{d}\mu$$

since ϵ is arbitrary. ■

The following corollary to the preceding result ensures consistency of Definition 2.7.4 with Definition 2.7.2.

2.7.6 Corollary (Consistency of integral definitions) *If $(\mathsf{X}, \mathscr{A}, \mu)$ is a measure space and if $\mathsf{f} \in \mathsf{S}(\mathsf{X}; \overline{\mathbb{R}}_{\geq 0})$ then the integral of f as in Definition 2.7.4 agrees with the integral of f as in Definition 2.7.2.*

Proof Consider the constant sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ defined by $f_j = f$, $j \in \mathbb{Z}_{>0}$. By Proposition 2.7.5 it follows that the integral of f from Definition 2.7.4 satisfies

$$\int_X f \,\mathrm{d}\mu = \lim_{j\to\infty} \int_X f_j \,\mathrm{d}\mu,$$

where the integrals on the left are as in Definition 2.7.2. However, each of these integrals is exactly the integral of f itself as in Definition 2.7.2. ■

Let us give a somewhat simple application of the preceding result that uses the counting measure. This example is interesting in and of itself as it begins the casting of the notion of summation using general index sets from Section I-2.4.7 in the framework of integration on measure spaces; this programme is completed in Example 2.7.10 below. For other examples of integration we shall wait until Sections 2.9 and 2.10.

2.7.7 Example (Sums as integrals) Let X be a set, take $\mathscr{A} = \mathbf{2}^X$, and let μ_Σ be the counting measure; see Example 2.3.9–3. Note that all functions $f\colon X \to \overline{\mathbb{R}}$ are measurable. Let $f \in \mathsf{L}^{(0)}((X, \mathscr{A}), \overline{\mathbb{R}}_{\geq 0})$ be a positive nonnegative-valued function. Let us attempt to understand the integral of f. We denote

$$\operatorname{supp}(f) = \{x \in X \mid f(x) \neq 0\}$$

and then consider three cases.

1. $\operatorname{supp}(f)$ *is finite:* Here f is a simple function and we immediately have

$$\int_X f \, d\mu_\Sigma = \sum_{x \in \operatorname{supp}(f)} f(x),$$

   using the definition of the integral of a simple function and the definition of the counting measure.

2. $\operatorname{supp}(f)$ *is countably infinite:* In this case we write $\operatorname{supp}(f) = \{x_j\}_{j \in \mathbb{Z}_{>0}}$ for distinct $x_j \in X$, $j \in \mathbb{Z}_{>0}$. Let us then define a sequence $(f_k)_{k \in \mathbb{Z}_{>0}}$ of $\overline{\mathbb{R}}_{\geq 0}$-valued functions on X by

$$f_k(x) = \begin{cases} f(x), & x \in \{x_1, \dots, x_k\}, \\ 0, & \text{otherwise}. \end{cases}$$

   Then the sequence $(f_k)_{k \in \mathbb{Z}_{>0}}$ is monotonically increasing and satisfies $\lim_{k \to \infty} f_k(x) = f(x)$ for every $x \in X$. Note that the functions f_k, $k \in \mathbb{Z}_{>0}$, are simple and that

$$\int_X f_k \, d\mu_\Sigma = \sum_{j=1}^{k} f(x_j),$$

   using the definition of the integral of a simple function and the definition of the counting measure. Thus, by Proposition 2.7.5 we have

$$\int_X f \, d\mu_\Sigma = \lim_{k \to \infty} \int_X f_k(x) \, d\mu_\Sigma = \sum_{j=1}^{\infty} f(x_j).$$

   In other words,

$$\int_X f \, d\mu_\Sigma = \sum_{x \in X} f(x),$$

   where the sum is interpreted as in Section I-2.4.7, and where we allow the sum to be infinite.

3. $\operatorname{supp}(f)$ *is uncountable:* For $k \in \mathbb{Z}_{>0}$ define

$$A_k = \left\{x \in X \;\middle|\; f(x) \geq \tfrac{1}{k}\right\}.$$

We claim that one of the sets A_k must be infinite for some $k \in \mathbb{Z}_{>0}$. Indeed, if all of the sets A_k, $k \in \mathbb{Z}_{>0}$, is finite then, since $\operatorname{supp}(f) = \cup_{k \in \mathbb{Z}_{>0}} A_k$, it follows that $\operatorname{supp}(f)$ is countable by Proposition I-1.7.16. Thus it must be the case that A_k is infinite for some $k \in \mathbb{Z}_{>0}$. In case A_k is uncountable, let A'_k be a countable subset of A_k. Now define $f_k \colon X \to \overline{\mathbb{R}}_{\geq 0}$ by

$$f_k(x) = \begin{cases} f(x), & x \in A'_k, \\ 0, & \text{otherwise}. \end{cases}$$

Note that $f_k(x) \leq f(x)$ for every $x \in X$. Then, using Exercise 2.7.1 and the fact that we know how to integrate f_k from the preceding case, we have

$$\int_X f \, \mathrm{d}\mu_\Sigma \geq \int_{A_k} f \, \mathrm{d}\mu_\Sigma = \sum_{x \in A_k} f_k(x) \geq \sum_{x \in A_k} \frac{1}{k} = \infty.$$

Thus the integral of f is infinite.

Thus, in summary, we have

$$\int_X f \, \mathrm{d}\mu_\Sigma = \sum_{x \in X} f(x),$$

using the definition of series using arbitrary index sets in Section I-2.4.7, and with the convention that the integral is allowed to be infinite, and indeed will be infinite if $\operatorname{supp}(f)$ is uncountable. •

Integration of general measurable functions

It is now relatively easy to define the integral for general measurable functions on a measure space $(X, \mathscr{A}, \mu)$. To do so, if $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ we define $f_+, f_- \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ by

$$f_+(x) = \max\{f(x), 0\}, \qquad f_-(x) = \max\{-f(x), 0\},$$

noting that these functions are indeed measurable by Corollary 2.6.17. We may now directly give the definition of the integral.

2.7.8 Definition (Integral of measurable function) For a measure space $(X, \mathscr{A}, \mu)$ and for $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$, we have the following definitions.

(i) If at least one of $\int_X f_+ \, \mathrm{d}\mu$ or $\int_X f_- \, \mathrm{d}\mu$ are finite then the integral of f with respect to μ ***exists*** and is given by

$$\int_X f \, \mathrm{d}\mu = \int_X f_+ \, \mathrm{d}\mu - \int_X f_- \, \mathrm{d}\mu,$$

this being the ***integral*** of f with respect to μ.

(ii) If both $\int_X f_+ \, \mathrm{d}\mu$ and $\int_X f_- \, \mathrm{d}\mu$ are infinite then the integral of f with respect to μ ***does not exist***.

(iii) If $\int_X f_+ \, \mathrm{d}\mu < \infty$ and $\int_X f_- \, \mathrm{d}\mu < \infty$ then f is ***integrable*** with respect to μ.

For a subset $I \subseteq \overline{\mathbb{R}}$ we denote the set of I-valued functions integrable with respect to μ by $\mathsf{L}^{(1)}((X, \mathscr{A}, \mu); I)$, or simply by $\mathsf{L}^{(1)}(X; I)$ if $\mathscr{A}$ and μ are understood. •

2.7.9 Notation ($\mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$) The notation $\mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ seems a little odd at this point. For example, what does the superscript "1" mean? And why are there parentheses around the "1." This will be presented in context in Section 3.8.8, so the reader should perhaps not worry at this point what is the precise meaning of the "1." We might mention, however, that the "L" refers to "Lebesgue," as this notation was first used in the context of the Lebesgue integral, and this will be the setting where the notation will be mainly used by us in these volumes. •

Again, we delay until Sections 2.9 and 2.10 the presentation of examples related to the Lebesgue measure. However, we can at this point complete our example of how the integral includes the usual notion of series.

2.7.10 Example (Sums as integrals (cont'd)) As in Example 2.7.7 we consider a set X, we let $\mathscr{A} = \mathbf{2}^X$, and we let μ_Σ be the counting measure defined in Example 2.3.9–3. We let $f\colon X \to \overline{\mathbb{R}}$, noting again that all functions are measurable. We then note that, as in Example 2.7.7, we have

$$\int_X f_+ \, \mathrm{d}\mu_\Sigma = \sum_{x \in X} f_+(x), \qquad \int_X f_- \, \mathrm{d}\mu_\Sigma = \sum_{x \in X} f_-(x), \tag{2.15}$$

using the notion of sums with arbitrary index sets from Section I-2.4.7, and allowing that these quantities may be infinite. Note that the general summation construction of Section I-2.4.7, along with the definition of the integral, then immediately gives

$$\int_X f \, \mathrm{d}\mu_\Sigma = \sum_{x \in X} f(x)$$

if either of the sums in (2.15) is finite, and otherwise the integral is undefined.

Using Proposition I-2.4.32 we see that in the case that $X = \mathbb{Z}_{>0}$, a function is integrable if and only if the sum $\sum_{j=1}^{\infty} f(j)$ is absolutely convergent. In this case, the value of the integral is exactly the sum of the series. Thus we see that the construction of the integral we give generalises the notion of an *absolutely* convergent series. Note that it does not generalise the notion of a convergent series. It can be made to do so by using special constructions. We do this for the Lebesgue integral in Sections 2.9.2 and 2.10.2. •

Let us close this section by giving a few more or less obvious properties of the integral.

2.7.11 Proposition (Integrals of functions agreeing almost everywhere) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space and let* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}})$ *have the property that* $\mathsf{f}(\mathsf{x}) = \mathsf{g}(\mathsf{x})$ *for almost every* $\mathsf{x} \in \mathsf{X}$*. Then the integral of* f *exists if and only if the integral of* g *exists, and if either integral exists then we have*

$$\int_{\mathsf{X}} \mathsf{f}\, \mathrm{d}\mu = \int_{\mathsf{X}} \mathsf{g}\, \mathrm{d}\mu.$$

Proof By breaking both f and g into their positive and negative parts, we can without loss of generality suppose that both functions take values in $\overline{\mathbb{R}}_{\geq 0}$. Let Z be the set where f and g are not equal and let h take the value ∞ on Z and zero elsewhere. Since $f \leq g + h$ we have

$$\int_X f\, \mathrm{d}\mu \leq \int_X g\, \mathrm{d}\mu + \int_X h\, \mathrm{d}\mu,$$

by Propositions 2.7.16 and 2.7.19. The argument can be reversed to give

$$\int_X g\, \mathrm{d}\mu \leq \int_X f\, \mathrm{d}\mu + \int_X h\, \mathrm{d}\mu,$$

and the result follows since $\int_X h\, \mathrm{d}\mu = 0$. ■

The following simple result comes up on occasion in our presentation, so we state it explicitly. Since the result is "obvious," we shall often use it without mention.

2.7.12 Proposition (Integrable functions are almost everywhere finite) *If* $(\mathsf{X}, \mathscr{A}, \mu)$ *is a measure space and if* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *then*

$$\mu\Big(\{\mathsf{x} \in \mathsf{X} \mid \mathsf{f}(\mathsf{x}) \notin \mathbb{R}\}\Big) = 0.$$

Proof Since f is integrable if both its positive and negative parts, f_+ and f_-, are integrable, we may as well assume that f takes values in $\overline{\mathbb{R}}_{\geq 0}$. Suppose that $f(x) = \infty$ for $x \in A$ with $\mu(A) > 0$. For $N \in \mathbb{Z}_{>0}$ consider the simple function

$$g_N(x) = \begin{cases} N, & x \in A, \\ 0, & \text{otherwise.} \end{cases}$$

We have $g_N(x) \leq f(x)$ for all $x \in A$ and $\int_X g_N\, \mathrm{d}\mu = N\mu(A) > 0$. By the definition of the integral we have $\int_X f\, \mathrm{d}\mu \geq N\mu(A)$, so showing that the integral of f is not finite, since this holds for all $N \in \mathbb{Z}_{>0}$. ■

2.7.13 Remark (Integrable functions may as well be $\mathbb{R}$-valued) Combining Propositions 2.7.11 and 2.7.12 we see that if $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ then, for the purposes of integration, we may as well suppose that f is $\mathbb{R}$-valued. Indeed, if we define

$$g(x) = \begin{cases} f(x), & f(x) \in \mathbb{R}, \\ 0, & f(x) \in \{-\infty, \infty\}, \end{cases}$$

then $\int_X g\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\mu$. For this reason, when we discuss spaces of integrable functions in Section 3.8, we will assume all functions are finite-valued. It is really only useful to allow functions to take infinite values when doing constructions with pointwise limits. •

The following result is another "obvious" result that we will use without mention throughout the text.

2.7.14 Proposition (Positive functions with zero integral) *If* $(\mathsf{X}, \mathscr{A}, \mu)$ *is a measure space and if* $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ *satisfies* $\int_{\mathsf{X}} \mathsf{f}\,\mathrm{d}\mu = 0$ *then*

$$\mu\Big(\{\mathsf{x} \in \mathsf{X} \mid \mathsf{f}(\mathsf{x}) \neq 0\}\Big) = 0.$$

Proof Suppose that $A \subseteq X$ has positive Lebesgue measure and that $f(x) > 0$ for all $x \in A$. Since $f \geq f\chi_A$, by Proposition 2.7.19 it follows that

$$\int_X f\,\mathrm{d}\mu \geq \int_X f\chi_A\,\mathrm{d}\mu > 0,$$

which gives the result. ■

As a final result in this section we record the relationship between functions that are measurable on the completion of a measure space and those that are measurable on the incomplete measure space.

2.7.15 Proposition (Integrable functions on the completion) *Let* $(\mathsf{X}, \mathscr{A}_\mu, \overline{\mu})$ *be the completion of the measure space* $(\mathsf{X}, \mathscr{A}, \mu)$ *and let* $\mathsf{f}\colon \mathsf{X} \to \overline{\mathbb{R}}$ *be* $\mathscr{A}_\mu$*-measurable. Then there exists a function* $\mathsf{g}\colon \mathsf{X} \to \overline{\mathbb{R}}$ *that is* $\mathscr{A}$*-measurable and with the property that*

$$\mu(\{\mathsf{x} \in \mathsf{X} \mid \mathsf{g}(\mathsf{x}) \neq \mathsf{f}(\mathsf{x})\}) = 0.$$

Moreover, the integral of f *with respect to* $\overline{\mu}$ *exists if and only if the integral of* g *with respect to* μ *exists, and in this case,*

$$\int_{\mathsf{X}} \mathsf{f}\,\mathrm{d}\overline{\mu} = \int_{\mathsf{X}} \mathsf{g}\,\mathrm{d}\mu.$$

Proof First suppose that f takes values in $\overline{\mathbb{R}}_{\geq 0}$. By Proposition 2.6.39 let $(g_j)_{j\in\mathbb{Z}_{>0}}$ be a monotonically increasing sequence of simple functions for which $\lim_{j\to\infty} g_j(x) = f(x)$ for all $x \in X$. This means that we may write f as an infinite sum of characteristic functions:

$$f(x) = \sum_{j=1}^{\infty} c_j \chi_{A_j}(x),$$

where $c_j \in \mathbb{R}_{\geq 0}$ and $A_j \in \mathscr{A}_\mu$, $j \in \mathbb{Z}_{>0}$. For $j \in \mathbb{Z}_{>0}$ let $L_j, U_j \in \mathscr{A}$ have the property that $L_j \subseteq A_j \subseteq U_j$ and $\mu(U_j \setminus L_j) = 0$. Taking

$$g(x) = \sum_{j=1}^{\infty} c_j \chi_{U_j}(x)$$

for $x \in X$ gives the first part of the result in this case since f and g differ on the set $(\cup_{j\in\mathbb{Z}_{>0}} U_j \setminus A_j) \subseteq (\cup_{j\in\mathbb{Z}_{>0}} U_j \setminus L_j)$, and this latter set has measure zero by Exercise 2.3.4.

Now suppose that f is now allowed to take arbitrary values in $\overline{\mathbb{R}}$. Write $f = f_+ - f_-$, where

$$f_+(x) = \max\{f(x), 0\}, \qquad f_-(x) = \max\{-f(x), 0\}.$$

These functions are $\mathscr{A}_\mu$-measurable by Corollary 2.6.17. Therefore, there exist $\mathscr{A}$-measurable functions g_+ and g_- such that f_+ differs from g_+ and f_- differs from g_- on a set of measure zero. Therefore, f differs from $g = g_+ - g_-$ on a set of measure zero. The result follows since g is $\mathscr{A}$-measurable by Proposition 2.6.11.

Now let us prove the last assertion of the proposition. Write $f = g + h$ for f being $\mathscr{A}_\mu$-measurable, for g being $\mathscr{A}$-measurable, and for

$$\mu(\{x \in X \mid h(x) \neq 0\}) = 0.$$

Let $Z \in \mathscr{A}$ be a set such that $h(x) = 0$ for $x \in X \setminus Z$ and such that $\mu(Z) = 0$. Then

$$\int_X f \, \mathrm{d}\overline{\mu} = \int_{X\setminus Z} g \, \mathrm{d}\overline{\mu} + \int_Z (g + h) \, \mathrm{d}\overline{\mu} = \int_{X\setminus Z} g \, \mathrm{d}\overline{\mu} = \int_X g \, \mathrm{d}\overline{\mu},$$

using Proposition 2.7.11. Now note that since g if integrable with respect to μ, its integral with respect to $\overline{\mu}$ can be constructed using the definition of the integral without reference to the distinction between μ and $\overline{\mu}$. That is to say,

$$\int_X g \, \mathrm{d}\overline{\mu} = \int_X g \, \mathrm{d}\mu,$$

and from this the result follows. ■

2.7.2 The integral and operations on functions

In this section we provide the more or less expected result regarding the interaction of the integral with the standard operations one may perform on functions. It is useful to record two different versions of results, one for arbitrary positive measurable functions and one for integrable functions.

We begin with the relationships between the integral and the standard algebraic operations on functions. We recall from Proposition 2.6.11 that $\mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ is a subset of the set $\overline{\mathbb{R}}^X$ of all $\overline{\mathbb{R}}$-valued functions on X, and this subset is closed under addition and multiplication on $\overline{\mathbb{R}}$. With this in mind we have the following results.

2.7.16 Proposition (Algebraic operations on positive measurable functions) *For a measure space* $(\mathsf{X}, \mathscr{A}, \mu)$*, for* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$*, and for* $\alpha \in \overline{\mathbb{R}}_{\geq 0}$*, the following statements hold:*

(i) $\displaystyle\int_\mathsf{X} (\mathsf{f} + \mathsf{g}) \, \mathrm{d}\mu = \int_\mathsf{X} \mathsf{f} \, \mathrm{d}\mu + \int_\mathsf{X} \mathsf{g} \, \mathrm{d}\mu;$

(ii) $\displaystyle\int_\mathsf{X} \alpha\mathsf{f} \, \mathrm{d}\mu = \alpha \int_\mathsf{X} \mathsf{f} \, \mathrm{d}\mu.$

Proof We let $(f_j)_{j\in\mathbb{Z}_{>0}}$ and $(g_j)_{j\in\mathbb{Z}_{>0}}$ be sequences of simple functions converging to f and g, respectively, as in Proposition 2.6.39.

(i) Note that if either $\lim_{j\to\infty} f_j(x)$ or $\lim_{j\to\infty} g_j(x)$ is infinite, then

$$\lim_{j\to\infty}(f_j+g_j)(x) = \lim_{j\to\infty} f_j(x) + \lim_{j\to\infty} g_j(x) = f(x)+g(x) = \infty.$$

If both $\lim_{j\to\infty} f_j(x)$ and $\lim_{j\to\infty} g_j(x)$ are finite then we have

$$\lim_{j\to\infty}(f_j+g_j)(x) = \lim_{j\to\infty} f_j(x) + \lim_{j\to\infty} g_j(x) = f(x)+g(x)$$

by Proposition I-2.3.23. Thus $(f_j+g_j)_{j\in\mathbb{Z}_{>0}}$ is a monotonically increasing sequence of simple functions converging to $f+g$. Thus this part of the result will follow from Proposition 2.7.5 if we can establish it for simple functions. Thus we assume that f and g are simple functions and denote

$$f = \sum_{j=1}^{k} a_j\chi_{A_j}, \quad g = \sum_{l=1}^{m} b_l\chi_{B_l}.$$

for $a_1,\dots,a_k,b_1,\dots,b_l \in \overline{\mathbb{R}}$ and $A_1,\dots,A_k$ and $B_1,\dots,B_m$ are disjoint. We assume without loss of generality that $\cup_{j=1}^{k} A_j = \cup_{l=1}^{m} B_l$. Then

$$\begin{aligned}\int_A (f+g)\,\mathrm{d}\mu &= \sum_{j=1}^{k}\sum_{l=1}^{m}(a_j+b_l)\mu(A_j\cap B_l)\\ &= \sum_{j=1}^{k}\sum_{l=1}^{m} a_j\mu(A_j\cap B_l) + \sum_{j=1}^{k}\sum_{l=1}^{m} b_l\mu(A_j\cap B_l)\\ &= \sum_{j=1}^{k} a_j\mu(A_j) + \sum_{l=1}^{m} b_l\mu(B_l)\\ &= \int_A f\,\mathrm{d}\mu + \int_A g\,\mathrm{d}\mu,\end{aligned}$$

so giving (i).

(ii) If either α or $\lim_{j\to\infty} f_j(x)$ is infinite then obviously we have

$$\lim_{j\to\infty} \alpha f_j(x) = \alpha f(x) = \infty.$$

If both α and $\lim_{j\to\infty} f_j(x)$ are finite then we have

$$\lim_{j\to\infty} \alpha f_j(x) = \alpha f(x)$$

by Proposition I-2.3.23. Thus $(\alpha f_j)_{j\in\mathbb{Z}_{>0}}$ is a monotonically increasing sequence of positive simple functions that converges to αf. Part (ii) then follows from Proposition 2.7.5. ■

2.7.17 Proposition (Algebraic operations on integrable functions) *For a measure space* $(\mathsf{X}, \mathscr{A}, \mu)$*, for* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$*, and for* $\alpha \in \mathbb{R}$*, the following statements hold:*

(i) $\mathsf{f} + \mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *and*

$$\int_{\mathsf{X}} (\mathsf{f} + \mathsf{g})\, \mathrm{d}\mu = \int_{\mathsf{X}} \mathsf{f}\, \mathrm{d}\mu + \int_{\mathsf{X}} \mathsf{g}\, \mathrm{d}\mu;$$

(ii) $\alpha \mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *and*

$$\int_{\mathsf{X}} \alpha \mathsf{f}\, \mathrm{d}\mu = \alpha \int_{\mathsf{X}} \mathsf{f}\, \mathrm{d}\mu.$$

Proof The proposition follows from Proposition 2.7.16 by breaking f and g into their positive and negative parts, and applying the lemma to both resulting integrals. ■

One might wonder about the relationships between integrals and other algebraic operations on functions, like multiplication and division. Generally speaking, these operations fail to preserve integrability.

2.7.18 Examples (Multiplication, division, and the integral)

1. We take $X = \mathbb{Z}_{>0}$ with the σ-algebra $\mathscr{A} = \mathbf{2}^{\mathbb{Z}_{>0}}$ and the counting measure μ_Σ. In this case, integrable functions are those functions $f\colon \mathbb{Z}_{>0} \to \mathbb{R}$ satisfying $\sum_{j=1}^{\infty} |f(j)| < \infty$; this follows from Example 2.7.10, or more straightforwardly from Exercise 2.7.3. Let us define $f\colon \mathbb{Z}_{>0} \to \mathbb{R}$ by $f(j) = \frac{1}{j^2}$. By Example I-2.4.2–4 it follows that $f \in \mathsf{L}^{(1)}((\mathbb{Z}_{>0}, \mathbf{2}^{\mathbb{Z}_{>0}}, \mu_\Sigma); \mathbb{R})$. However, since $f^2(j) = \frac{1}{j}$, it follows from Example I-2.4.2–2 that $f^2 \notin \mathsf{L}^{(1)}((\mathbb{Z}_{>0}, \mathbf{2}^{\mathbb{Z}_{>0}}, \mu_\Sigma); \mathbb{R})$. Thus products of integrable functions need not be integrable functions.
2. We take $X = \mathbb{Z}_{>0}$, $\mathscr{A} = \mathbf{2}^{\mathbb{Z}_{>0}}$, and $\mu = \mu_\Sigma$ as in the previous example. We note that if we define $f, g\colon \mathbb{Z}_{>0} \to \mathbb{R}$ by $f(j) = \frac{1}{j^2}$ and $g(j) = \frac{1}{j^3}$; as above, $f, g \in \mathsf{L}^{(1)}((\mathbb{Z}_{>0}, \mathbf{2}^{\mathbb{Z}_{>0}}, \mu_\Sigma); \mathbb{R})$. However, clearly $\frac{f}{g}(j) = \frac{1}{j}$ and so $\frac{f}{g} \notin \mathsf{L}^{(1)}((\mathbb{Z}_{>0}, \mathbf{2}^{\mathbb{Z}_{>0}}, \mu_\Sigma); \mathbb{R})$. Thus the quotient of two integrable functions is not necessarily integrable, even when the denominator function is nowhere zero. •

For functions whose values are related by the total order on $\overline{\mathbb{R}}$ we have the following result applies.

2.7.19 Proposition (The integral and total order on $\overline{\mathbb{R}}$) *If* $(\mathsf{X}, \mathscr{A}, \mu)$ *is a measure space and if* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *(resp.* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$*) satisfy* $\mathsf{f}(\mathsf{x}) \leq \mathsf{g}(\mathsf{x})$ *for almost all* $\mathsf{x} \in \mathsf{X}$*, then*

$$\int_{\mathsf{X}} \mathsf{f}\, \mathrm{d}\mu \leq \int_{\mathsf{X}} \mathsf{g}\, \mathrm{d}\mu.$$

Proof Without loss of generality we may suppose that $f(x) \leq g(x)$ for all $x \in X$. Indeed, if this inequality holds except on a set Z which has zero measure, then we have

$$\int_X f\,\mathrm{d}\mu = \int_{X\setminus Z} f\,\mathrm{d}\mu + \int_Z f\,\mathrm{d}\mu = \int_{X\setminus Z} f\,\mathrm{d}\mu,$$

and so we can simply replace X with $X \setminus Z$.

Now we may use part (i) from Proposition 2.7.16 or Proposition 2.7.17 to write

$$\int_X g\,\mathrm{d}\mu = \int_X (f + (g - f))\,\mathrm{d}\mu = \int_X f\,\mathrm{d}\mu + \int_X (g - f)\,\mathrm{d}\mu \geq \int_A f\,\mathrm{d}\mu,$$

as desired. ■

This result has the following corollary which we often apply.

2.7.20 Corollary (Functions bounded by integrable functions are integrable) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space and let* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}})$ *satisfy* $|\mathsf{f}(\mathsf{x})| \leq |\mathsf{g}(\mathsf{x})|$ *for almost every* $\mathsf{x} \in \mathsf{X}$*. If* $\mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *then* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$.

Proof Write $f = f_+ - f_-$ and $g = g_+ - g_-$ for $f_+, g_+, f_-, g_- \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$. Then we obviously have

$$f_+(x) \leq g_+(x), \quad f_- \leq g_-(x)$$

for almost every $x \in X$. Thus, by Proposition 2.7.19 we have

$$\int_X f_+\,\mathrm{d}\mu \leq \int_X g_+\,\mathrm{d}\mu, \quad \int_X f_-\,\mathrm{d}\mu \leq \int_X g_-\,\mathrm{d}\mu.$$

Therefore,

$$\int_X |f|\,\mathrm{d}\mu = \int_X f_+\,\mathrm{d}\mu + \int_X f_-\,\mathrm{d}\mu \leq \int_X g_+\,\mathrm{d}\mu + \int_X g_-\,\mathrm{d}\mu = \int_X |g|\,\mathrm{d}\mu,$$

as desired. ■

The following result follows pretty much from the definitions surrounding the Lebesgue integral.

2.7.21 Proposition (The integral and absolute value) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space and let* $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}})$*. Then* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *if and only if* $|\mathsf{f}| \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$*, and if* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *then*

$$\left|\int_{\mathsf{X}} \mathsf{f}\,\mathrm{d}\mu\right| \leq \int_{\mathsf{X}} |\mathsf{f}|\,\mathrm{d}\mu.$$

Proof The first assertion is Exercise 2.7.4. For the second assertion, write $f = f_+ - f_-$ as the sum of its positive and negative parts. Then

$$\left|\int_X f\,\mathrm{d}\mu\right| \leq \left|\int_X f_+\,\mathrm{d}\mu\right| + \left|\int_X f_-\,\mathrm{d}\mu\right| = \int_X |f|\,\mathrm{d}\mu,$$

using the fact that for a positive function the integral is positive. ■

It is at times useful to break an integral into two parts by breaking the domain of integration into two parts.

2.7.22 Proposition (Breaking the integral in two) *Let* $(X, \mathscr{A}, \mu)$ *be a measure space, let* $A, B \in \mathscr{A}$ *be sets such that* $X = A \overset{\circ}{\cup} B$, *and let* $f \in L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$. *Then* $f \in L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *if and only if* $f|A \in L^{(1)}((A, \mathscr{A}_A, \mu|\mathscr{A}_A); \overline{\mathbb{R}})$ *and* $f|B \in L^{(1)}((B, \mathscr{A}_b, \mu|\mathscr{A}_B); \overline{\mathbb{R}})$. *Furthermore, if* $f \in L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *then we have*

$$\int_X f\, d\mu = \int_A (f|A)\, d\mu_A + \int_B (f|B)\, d\mu_B.$$

Proof Let us define $f_A, f_B \colon X \to \overline{\mathbb{R}}$ by $f_A = f\chi_A$ and $f_B = f\chi_B$. By Proposition 2.6.15 the functions f_A and f_B are measurable. We claim that, provided that f_A and $f|A$ are integrable,

$$\int_X f_A\, d\mu = \int_A (f|A)\, d\mu_A. \tag{2.16}$$

To see this, first suppose that f is $\overline{\mathbb{R}}_{\geq 0}$-valued and let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of simple functions converging to f as in Proposition 2.6.39. Then the sequence $(f_{A,j})_{j\in\mathbb{Z}_{>0}}$ defined by $f_{A,j} = f_j\chi_A$ is a sequence of simple functions converging to f_A as in Proposition 2.6.39. Moreover,

$$\int_X f_{A,j}\, d\mu = \int_A (f_j|A)\, d\mu_A$$

by Exercise 2.7.2. Therefore, by Proposition 2.7.5 we have

$$\int_X f_A\, d\mu = \lim_{j\to\infty} \int_X f_{A,j}\, d\mu = \lim_{j\to\infty} \int_A (f_j|A)\, d\mu_A = \int_A (f|A)\, d\mu_A,$$

giving (2.16) when f is $\overline{\mathbb{R}}_{\geq 0}$-valued. For $\overline{\mathbb{R}}$-valued f the same conclusion follows by breaking f into its positive and negative parts. Similarly, of course, we have

$$\int_X f_B\, d\mu = \int_B (f|B)\, d\mu_B,$$

and so Proposition 2.7.17 gives the final assertion of the result provided that f, f_A, and f_B are integrable.

Now, if f_A and f_B are integrable, by Proposition 2.7.17 it follows that f is integrable. Conversely, if either of f_A or f_B are not integrable, then neither can f be integrable (why?). ■

A more general version of the preceding result is useful, but is only valid for complete measure spaces.

2.7.23 Corollary (Breaking the integral almost in two) *Let* $(X, \mathscr{A}, \mu)$ *be a complete measure space, let* $A, B \in \mathscr{A}$ *be such that* $\mu(A \cap B) = 0$ *and such that* $X = A \cup B$, *and let* $f \in L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$. *Then* $f \in L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *if and only if* $f|A \in L^{(1)}((A, \mathscr{A}_A, \mu|\mathscr{A}_A); \overline{\mathbb{R}})$ *and* $f|B \in L^{(1)}((B, \mathscr{A}_B, \mu|\mathscr{A}_B); \overline{\mathbb{R}})$. *Furthermore, if* $f \in L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *then we have*

$$\int_X f\, d\mu = \int_A (f|A)\, d\mu_A + \int_B (f|B)\, d\mu_B.$$

Proof Let $Z = A \cap B$, let $A' = A - Z$ and $B' = B - Z$ and write $X = A' \mathring{\cup} B' \mathring{\cup} Z$. Note that Z, A', and B' are measurable since X is complete. Applying Proposition 2.7.22 (or more properly, its obvious extension to finitely many disjoint components) gives

$$\int_X f\,\mathrm{d}\mu = \int_{A'} (f|A')\,\mathrm{d}\mu_{A'} + \int_{B'} (f|B')\mathrm{d}\mu_{B'} + \int_Z (f|Z)\,\mathrm{d}\mu.$$

The last integral is zero by Proposition 2.7.11 and, by the same result,

$$\int_{A'} (f|A')\,\mathrm{d}\mu_{A'} = \int_A (f|A)\,\mathrm{d}\mu_A$$

and

$$\int_{B'} (f|B')\,\mathrm{d}\mu_{B'} = \int_B (f|B)\,\mathrm{d}\mu_B,$$

giving the result. ■

2.7.3 Limit theorems

In Section 2.1 we suggested that one of the reasons why the Riemann integral was not satisfactory was that it did not have useful properties with respect to swapping of limits and integration. In this section we prove some powerful theorems for the integral on measure spaces which give very general conditions under which limits and integrals will swap. When these are applied to the Lebesgue integral in Sections 2.9 and 2.10, this will show that we have produced a theory of integration that generalises the Riemann integral, and which has at least some more desirable properties.

Our first theorem has very weak hypotheses, but only applies to nonnegative functions.

2.7.24 Theorem (Monotone Convergence Theorem I) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space and let* $(\mathsf{f}_j)_{j\in\mathbb{Z}_{>0}}$ *be a sequence in* $\mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ *such that, for almost every* $\mathsf{x} \in \mathsf{X}$, $\mathsf{f}_j(\mathsf{x}) \leq \mathsf{f}_{j+1}(\mathsf{x})$ *for every* $j \in \mathbb{Z}_{>0}$. *If* $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}})$ *has the property that* $\mathsf{f}(\mathsf{x}) = \lim_{j\to\infty} \mathsf{f}_j(\mathsf{x})$ *for almost every* $\mathsf{x} \in \mathsf{X}$, *then*

$$\int_{\mathsf{X}} \mathsf{f}\,\mathrm{d}\mu = \lim_{j\to\infty} \int_{\mathsf{X}} \mathsf{f}_j\,\mathrm{d}\mu.$$

Proof First let us show that we may assume without loss of generality that the relations $f_j(x) \leq f_{j+1}(x)$, $j \in \mathbb{Z}_{>0}$, and $f(x) = \lim_{j\to\infty} f_j(x)$ hold for all $x \in X$. Let Z be the set on which these relation do not hold, noting that Z has measure zero being a union of two sets of measure zero. Let $Y = X \setminus Z$. The sequence of functions $(f_j\chi_Y)_{j\in\mathbb{Z}_{>0}}$ and the function $f\chi_Y$ then satisfy the relations for all $x \in X$. If the theorem holds in this case, then the result will follow from Proposition 2.7.11. For the remainder of the proof we therefore assume that $f_j(x) \leq f_{j+1}(x)$, $j \in \mathbb{Z}_{>0}$, and $f(x) = \lim_{j\to\infty} f_j(x)$ for all $x \in X$.

By Proposition 2.7.19 we have

$$\int_X f_j\,\mathrm{d}\mu \leq \int_X f_{j+1}\,\mathrm{d}\mu, \qquad j \in \mathbb{Z}_{>0},$$

and

$$\int_X f_j \, d\mu \leq \int_X f \, d\mu, \qquad j \in \mathbb{Z}_{>0}.$$

Thus the sequence $(\int_X f_j \, d\mu)_{j\in\mathbb{Z}_{>0}}$ converges in $\overline{\mathbb{R}}_{\geq 0}$ to a limit and this limit satisfies

$$\lim_{j\to\infty} \int_X f_j \, d\mu \leq \int_X f \, d\mu.$$

We wish to establish the opposite inequality. For each $j \in \mathbb{Z}_{>0}$ let $(g_{j,k})_{k\in\mathbb{Z}_{>0}}$ be a sequence of simple functions whose limit is f_j, as in Proposition 2.6.39. Now define $h_k(x) = \max\{g_{1,k}(x), \dots, g_{k,k}(x)\}$, and note that $(h_k)_{k\in\mathbb{Z}_{>0}}$ is a monotonically increasing sequence of simple functions converging to f, and that $h_k(x) \leq f_k(x)$ for all $x \in X$. By our above arguments for the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$, we have

$$\lim_{k\to\infty} \int_X h_k \, d\mu \leq \lim_{j\to\infty} \int_X f_j \, d\mu \leq \int_X f \, d\mu.$$

The theorem now follows by Proposition 2.7.5. ■

In the next assertion, the condition that the functions be nonnegative is relaxed, but one must add an integrability condition for one of the functions in the sequence.

2.7.25 Theorem (Monotone Convergence Theorem II) *Let* $(X, \mathscr{A}, \mu)$ *be a measure space and let* $(f_j)_{j\in\mathbb{Z}_{>0}}$ *be a sequence in* $L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ *such that, for almost every* $x \in X$, $f_j(x) \leq f_{j+1}(x)$ *for every* $j \in \mathbb{Z}_{>0}$ *and such that* $f_1 \in L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$*. If* $f \in L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ *has the property that* $f(x) = \lim_{j\to\infty} f_j(x)$ *for almost every* $x \in X$*, then*

$$\int_X f \, d\mu = \lim_{j\to\infty} \int_X f_j \, d\mu.$$

Proof As in the proof of Theorem 2.7.24 we can assume that $f_j(x) \leq f_{j+1}(x)$, $j \in \mathbb{Z}_{>0}$, and $f(x) = \lim_{j\to\infty} f_j(x)$ for every $x \in X$. Note that the sequence $(f_j - f_1)$ is then in $L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ and satisfies $\lim_{j\to\infty}(f_j(x) - f_1(x)) = f(x) - f_1(x)$ for every $x \in X$. Note that if $f \in L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ then we have

$$\int_X f \, d\mu - \int_X f_1 \, d\mu = \int_X (f - f_1) \, d\mu$$

by Proposition 2.7.17. If $f \notin L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ then the previous relation still holds with value ∞ on both sides (why?). Therefore, by Theorem 2.7.24, we have

$$\int_X f \, d\mu - \int_X f_1 \, d\mu = \int_X (f - f_1) \, d\mu = \lim_{j\to\infty} \int_X (f_j - f_1) \, d\mu = \lim_{j\to\infty} \int_X f_j \, d\mu - \int_X f_1 \, d\mu,$$

which gives the result since $f_1 \in L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$. ■

We also have the following immediate corollary to the Monotone Convergence Theorem.

2.7.26 Corollary (Beppo Levi's[9] Theorem) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space and let* $(\mathsf{f}_j)_{j\in\mathbb{Z}_{>0}}$ *be a sequence in* $\mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}}_{\geq 0})$*. If* $\mathsf{f}\colon \mathsf{X}\to\overline{\mathbb{R}}_{\geq 0}$ *is defined by*

$$\mathsf{f}(x) = \sum_{j=1}^{\infty} \mathsf{f}_j(x),$$

then f *is measurable and we have*

$$\int_{\mathsf{X}} \mathsf{f}\,\mathrm{d}\mu = \sum_{j=1}^{\infty} \int_{\mathsf{X}} \mathsf{f}_j\,\mathrm{d}\mu.$$

Proof Define $g_k(x) = \sum_{j=1}^k f_j(x)$, noting that $g_k \in \mathsf{L}^{(0)}((X,\mathscr{A});\overline{\mathbb{R}}_{\geq 0})$ by Proposition 2.6.11. Moreover, for every $x\in X$ we have $g_k(x)\leq g_{k+1}(x)$. Thus Theorem 2.7.24 and Proposition 2.7.16 imply that

$$\int_X f\,\mathrm{d}\mu = \lim_{k\to\infty}\int_X g_k\,\mathrm{d}\mu = \lim_{k\to\infty}\sum_{j=1}^k\int_X f_j\,\mathrm{d}\mu = \sum_{j=1}^{\infty}\int_X f_j\,\mathrm{d}\mu,$$

as desired. ■

The following result is also useful, but with weaker hypotheses and conclusions than the Monotone Convergence Theorem.

2.7.27 Theorem (Fatou's[10] Lemma) *If* $(\mathsf{X}, \mathscr{A}, \mu)$ *is a measure space and if* $(\mathsf{f}_j)_{j\in\mathbb{Z}_{>0}}$ *is a sequence in* $\mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}}_{\geq 0})$*, then*

$$\int_{\mathsf{X}} \liminf_{j\to\infty} \mathsf{f}_j\,\mathrm{d}\mu \leq \liminf_{j\to\infty}\int_{\mathsf{X}} \mathsf{f}_j\,\mathrm{d}\mu.$$

Proof For $k\in\mathbb{Z}_{>0}$ define $g_k(x) = \inf_{j\geq k} f_j(x)$, noting that g_k so defined is measurable by Proposition 2.6.18. We then note that the sequence $(g_k)_{k\in\mathbb{Z}_{>0}}$ is increasing and that

$$\liminf_{j\to\infty} f_j(x) = \lim_{k\to\infty} g_k(x)$$

for $x\in X$. From Theorem 2.7.24 we then have

$$\int_X \liminf_{j\to\infty} f_j\,\mathrm{d}\mu = \lim_{k\to\infty}\int_X g_k\,\mathrm{d}\mu \leq \liminf_{j\to\infty}\int_X f_j\,\mathrm{d}\mu,$$

since $g_j(x)\leq f_j(x)$ for $j\in\mathbb{Z}_{>0}$ and $x\in X$. ■

The most frequently useful of the limit theorems is the following. It is a result that is used with great regularity in integration theory. For example, many of the fundamental results we state in Sections 3.8 and IV-1.3 and in Chapters IV-5 and IV-6 rely at their core on this important theorem.

[9]Beppo Levi (1875–1961) was an Italian mathematician who made mathematical contributions to algebra and analysis. As a Jew, he left Italy after the rise of Mussolini for Argentina, where he spent much of his professional life.

[10]Pierre Joseph Louis Fatou (1878–1929) was a French mathematician who made substantial contributions to analysis, particularly complex analysis.

2.7.28 Theorem (Dominated Convergence Theorem I) *Let* $(\mathsf{X},\mathscr{A},\mu)$ *be a measure space and let* $(\mathsf{f}_j)_{j\in\mathbb{Z}_{>0}}$ *be a sequence in* $\mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}})$ *having the following properties:*

(i) the limit $\mathsf{f}(x)=\lim_{j\to\infty}\mathsf{f}_j(x)$ *exists for almost every* $x\in\mathsf{X}$*;*

(ii) there exists $g\in\mathsf{L}^{(1)}((\mathsf{X},\mathscr{A},\mu);\overline{\mathbb{R}}_{\ge0})$ *such that, for almost every* $x\in\mathsf{X}$, $|\mathsf{f}_j(x)|\le g(x)$ *for every* $j\in\mathbb{Z}_{>0}$.

Then the functions f *and* f_j, $j\in\mathbb{Z}_{>0}$, *are integrable and*

$$\int_{\mathsf{X}}\mathsf{f}\,\mathrm{d}\mu=\lim_{j\to\infty}\int_{\mathsf{X}}\mathsf{f}_j\,\mathrm{d}\mu.$$

Proof The integrability of f and f_j, $j\in\mathbb{Z}_{>0}$, follows from Corollary 2.7.20. As with our proof of Theorem 2.7.24, we can without loss of generality suppose that (i) and (ii) hold for all $x\in X$. Furthermore, since g is integrable, we may as well suppose that $g(x)\in\mathbb{R}$ for every x, again by Proposition 2.7.11. The sequence $(g+f_j)_{j\in\mathbb{Z}_{>0}}$ is then a sequence of nonnegative functions for which

$$\lim_{j\to\infty}(g+f_j)(x)=(g+f)(x),\qquad x\in X.$$

By Fatou's Lemma this gives

$$\begin{aligned}&\int_X(g+f)\,\mathrm{d}\mu\le\liminf_{j\to\infty}\int_X(g+f_j)\,\mathrm{d}\mu\\ \implies\quad&\int_X f\,\mathrm{d}\mu\le\liminf_{j\to\infty}\int_X f_j\,\mathrm{d}\mu.\end{aligned}$$

Similarly we can show that

$$\begin{aligned}&\int_X(g-f)\,\mathrm{d}\mu\le\liminf_{j\to\infty}\int_X(g-f_j)\,\mathrm{d}\mu\\ \implies\quad&\int_X f\,\mathrm{d}\mu\le\limsup_{j\to\infty}\int_X f_j\,\mathrm{d}\mu.\end{aligned}$$

This gives the result. ■

The Dominated Convergence Theorem also has the following weaker form for more general sequences.

2.7.29 Theorem (Dominated Convergence Theorem II) *Let* $(\mathsf{X},\mathscr{A},\mu)$ *be a measure space and let* $(\mathsf{f}_j)_{j\in\mathbb{Z}_{>0}}$ *be a sequence in* $\mathsf{L}^{(0)}((\mathsf{X},\mathscr{A});\overline{\mathbb{R}})$ *for which there exists* $g\in\mathsf{L}^{(1)}((\mathsf{X},\mathscr{A},\mu);\overline{\mathbb{R}}_{\ge0})$ *such that, for almost every* $x\in\mathsf{X}$, $|\mathsf{f}_j(x)|\le g(x)$ *for every* $j\in\mathbb{Z}_{>0}$. *Then the functions* f_j, $j\in\mathbb{Z}_{>0}$, *are integrable and*

(i) $\displaystyle\int_{\mathsf{X}}\liminf_{j\to\infty}\mathsf{f}_j\,\mathrm{d}\mu\le\liminf_{j\to\infty}\int_{\mathsf{X}}\mathsf{f}_j\,\mathrm{d}\mu$ *and*

(ii) $\displaystyle\int_{\mathsf{X}}\limsup_{j\to\infty}\mathsf{f}_j\,\mathrm{d}\mu\ge\limsup_{j\to\infty}\int_{\mathsf{X}}\mathsf{f}_j\,\mathrm{d}\mu.$

Proof The proofs for both conclusions are similar, so we only prove (i). The integrability f_j, $j \in \mathbb{Z}_{>0}$, follows from Corollary 2.7.20. The measurability of $x \mapsto \liminf_{j\to\infty} f_j(x)$ follows from Proposition 2.6.18. As in the proof of Theorem 2.7.24 we may as well assume that $|f_j(x)| \leq |g(x)|$ for all $x \in X$ and $j \in \mathbb{Z}_{>0}$. In this case, the sequence $(g + f_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence in $\mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ and so, by Fatou's Lemma and Proposition 2.7.16, we have

$$\begin{aligned}\int_X g \, d\mu + \int_X \liminf_{j\in\infty} f_j \, d\mu &= \int_X \liminf_{j\to\infty}(g + f_j) \, d\mu \\ &\leq \liminf_{j\to\infty} \int_X (g + f_j) \, d\mu \\ &= \int_X g \, d\mu + \liminf_{j\to\infty} \int_X f_j \, d\mu,\end{aligned}$$

which gives the result since $g \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$. ■

Let us illustrate how one might use the preceding results.

2.7.30 Examples (Illustration of limit theorems) In both of the examples, we consider the measure space $(X, \mathscr{A}, \mu)$ with $X = \mathbb{Z}_{>0}$, $\mathscr{A} = \mathbf{2}^{\mathbb{Z}_{>0}}$, and $\mu = \mu_\Sigma$, the counting measure. Thus, as we have seen in Example 2.7.10, integrable functions are absolutely convergent series.

1. The Monotone Convergence Theorem is often helpful for showing that a certain integral diverges. Let us illustrate this as follows. We wish to ascertain whether the limit

$$\lim_{\alpha\downarrow 1} \sum_{k=1}^{\infty} \frac{1}{k^\alpha} \tag{2.17}$$

exists. Let us define $f_\alpha \in \mathsf{L}^{(0)}((\mathbb{Z}_{>0}, \mathbf{2}^{\mathbb{Z}_{>0}}); \overline{\mathbb{R}}_{\geq 0})$ by $f_\alpha(k) = \frac{1}{k^\alpha}$ for $\alpha \in [1, 2]$. Let $(\alpha_j)_{j\in\mathbb{Z}_{>0}}$ be a strictly monotonically decreasing sequence such that $\alpha_1 = 2$ and $\lim_{j\to\infty} \alpha_j = 1$. We then have

$$\lim_{\alpha\downarrow 1} \sum_{k=1}^{\infty} \frac{1}{k^\alpha} = \lim_{\alpha\downarrow 1} \int_{\mathbb{Z}_{>0}} f_\alpha \, d\mu_\Sigma = \lim_{j\to\infty} \int_{\mathbb{Z}_{>0}} f_{\alpha_j} \, d\mu_\Sigma.$$

Note that $f_{\alpha_j}(k) < f_{\alpha_{j+1}}(k)$ for every $k \in \mathbb{Z}_{>0}$ and $j \in \mathbb{Z}_{>0}$. Therefore, the sequence $(f_{\alpha_j})_{j\in\mathbb{Z}_{>0}}$ satisfies the hypotheses of the Monotone Convergence Theorem. Therefore, we have

$$\lim_{j\to\infty} \int_{\mathbb{Z}_{>0}} f_{\alpha_j} \, d\mu_\Sigma = \int_{\mathbb{Z}_{>0}} \lim_{j\to\infty} f_{\alpha_j} \, d\mu_\Sigma = \sum_{k=1}^{\infty} \frac{1}{k} = \infty.$$

Thus the limit (2.17) does not exist, at least not in $\mathbb{R}$.

2. Let us use the Dominated Convergence Theorem to determine the value of the following limit:

$$\lim_{\alpha\downarrow 1}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k^{2\alpha}}.$$

We proceed much as above, defining $f_\alpha \in \mathsf{L}^{(0)}((\mathbb{Z}_{>0}, \mathbf{2}^{\mathbb{Z}_{>0}}); \overline{\mathbb{R}})$ by $f_\alpha(k) = \frac{(-1)^{k+1}}{k^{2\alpha}}$. We let $(\alpha_j)_{j\in\mathbb{Z}_{>0}}$ be a strictly monotonically decreasing sequence such that $\alpha_1 = 2$ and $\lim_{j\to\infty}\alpha_j = 1$. It then holds that

$$\lim_{\alpha\downarrow 1}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k^{2\alpha}} = \lim_{\alpha\downarrow 1}\int_{\mathbb{Z}_{>0}} f_{\alpha_j}\,\mathrm{d}\mu_\Sigma = \lim_{j\to\infty}\int_{\mathbb{Z}_{>0}} f_{\alpha_j}\,\mathrm{d}\mu_\Sigma.$$

We then have

$$|f_{\alpha_j}(k)| = \frac{1}{k^{2\alpha_j}} < \frac{1}{k^2}$$

for every $j \in \mathbb{Z}_{>0}$ and $k \in \mathbb{Z}_{>0}$. Define $g \in \mathsf{L}^{(0)}((\mathbb{Z}_{>0}, \mathbf{2}^{\mathbb{Z}_{>0}}); \overline{\mathbb{R}})$ by $g(k) = \frac{1}{k^2}$ and note that

$$\int_X g\,\mathrm{d}\mu_\Sigma = \sum_{k=1}^{\infty}\frac{1}{k^2} < \infty$$

by Example I-2.4.2–4. Therefore, the hypotheses of the Dominated Convergence Theorem apply, and we have

$$\lim_{j\to\infty}\int_{\mathbb{Z}_{>0}} f_{\alpha_j}\,\mathrm{d}\mu_\Sigma = \int_{\mathbb{Z}_{>0}}\lim_{j\to\infty} f_{\alpha_j}\,\mathrm{d}\mu_\Sigma = \sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k^2} = \frac{\pi^2}{12},$$

where we look up the last sum. •

2.7.4 Integration with respect to probability measures

In Section 2.3.5 we introduced the notion of a probability space. In this section we investigate integration on probability spaces, giving a few results peculiar and useful for such measure spaces.

The following general result concerning how integrals behave under composition by certain classes of functions. Recall from Sections I-3.1.6 and I-3.2.6 the notion of a convex function. We shall use properties of convex functions we proved in those sections.

2.7.31 Theorem (Jensen's[11] inequality) *Let $(\mathsf{X}, \mathscr{A}, \mu)$ be a finite measure space, let $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R})$, and let $\phi\colon \mathbb{R}\to\mathbb{R}$ be convex. Then*

$$\phi\Big(\int_X \mathsf{f}\,\mathrm{d}\mu\Big) \le \int_X (\phi\circ \mathsf{f})\,\mathrm{d}\mu.$$

[11] Johan Ludwig William Valdemar Jensen (1859–1925) was a Danish telephone company employee who did some mathematics in his spare time.

Proof From Proposition I-3.2.30(i) we have

$$\phi'(y_0+)(y-y_0)+\phi(y_0)\le\phi(y)$$

for every $y\in\mathbb{R}$. Let $x\in X$ and let us take

$$y_0=\int_X f\,\mathrm{d}\mu,\quad y=f(x),$$

so that the above inequality reads

$$\phi(y_0)\le\phi\circ f(x)-\phi'(y_0+)(f(x)-y_0).$$

By Proposition 2.7.19 we have

$$\int_X\phi(y_0)\,\mathrm{d}\mu\le\int_X\phi\circ f\,\mathrm{d}\mu-\int_X\phi'(y_0+)(f-y_0)\,\mathrm{d}\mu.$$

Since μ is a probability measure (i.e., $\int_X\mathrm{d}\mu=1$) and since the integral is linear, we have

$$\int_X\phi(y_0)\,\mathrm{d}\mu=\phi(y_0)$$

and

$$\int_X\phi'(y_0+)(f-y_0)\,\mathrm{d}\mu=\phi'(y_0+)\int_X f\,\mathrm{d}\mu-\phi'(y_0+)y_0=0.$$

This immediately gives the result. ■

The following version of Jensen's inequality is often useful. Here we make use of the Lebesgue integral on $\mathbb{R}$ discussed in detail in Section 2.9.

2.7.32 Corollary (Jensen's inequality for integration on intervals) *Let* $[a,b]\subseteq\mathbb{R}$ *be a compact interval, let* $f\in\mathsf{L}^{(1)}(([a,b],\mathscr{L}([a,b]),\lambda_{[a,b]});\mathbb{R})$*, and let* $\phi\colon\mathbb{R}\to\mathbb{R}$ *be convex. Then*

$$\phi\Big(\int_{[a,b]}f\,\mathrm{d}\lambda_{[a,b]}\Big)\le\frac{1}{b-a}\int_{[a,b]}\phi\circ((b-a)f)\,\mathrm{d}\lambda_{[a,b]}.$$

Proof We shall use Riemann integral notation in the proof, cf. Notation 2.9.13. By the change of variable theorem, Theorem 2.9.38,

$$\int_a^b f(x)\,\mathrm{d}x=\int_0^1(b-a)f(a+(b-a)s)\,\mathrm{d}s.$$

By Jensen's inequality above,

$$\phi\Big(\int_a^b f(x)\,\mathrm{d}x\Big)\le\int_0^1\phi(b-a)f(a+(b-a)s))\,\mathrm{d}s=\frac{1}{b-a}\int_a^b\phi((b-a)f(x))\,\mathrm{d}x,$$

which is the result. ■

Now we give a few characterisations of how a function deviates from its mean. For this, the following simple definition is useful.

2.7.33 Definition (Mean of a function) Let $(X, \mathscr{A}, \mu)$ be a measure space and let $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$. The ***mean*** of f is

$$\mathrm{mean}(f) = \int_X f \, \mathrm{d}\mu. \qquad \bullet$$

With this notion, we have the following results.

2.7.34 Theorem (Markov's[12] inequality) *Let* $(\mathrm{X}, \mathscr{A}, \mu)$ *be a measure space and let* $\mathrm{f} \in \mathsf{L}^{(1)}((\mathrm{X}, \mathscr{A}, \mu); \overline{\mathbb{R}}_{\geq 0})$. *Then, for any* $\mathrm{a} \in \mathbb{R}_{>0}$ *it holds that*

$$\mu(\{\mathrm{x} \in \mathrm{X} \mid \mathrm{f}(\mathrm{x}) \geq \mathrm{a}\}) \leq \frac{1}{\mathrm{a}}\mathrm{mean}(\mathrm{f}).$$

Proof Let us abbreviate

$$M_a = \{x \in \mathrm{X} \mid f(x) \geq a\}.$$

Then, for every $x \in X$.

$$a \leq a\chi_{M_a}(x) \leq f(x)\chi_{M_a}(x) \leq f(x).$$

Therefore, by Proposition 2.7.19,

$$\int_X (a\chi_{M_a}) \, \mathrm{d}\mu \leq \int_{M_a} f \, \mathrm{d}\mu_{M_a} \leq \int_X f \, \mathrm{d}\mu.$$

Dividing by a gives $\mu(M_a) \leq \frac{1}{a}\mathrm{mean}(f)$, as desired. ■

Very often Markov's inequality gives rather course estimates, and moreover only applies to nonnegative-valued functions. In this respect, the following results are sometimes useful.

2.7.35 Theorem (General Chebychev[13] inequality) *Let* $(\mathrm{X}, \mathscr{A}, \mu)$ *be a measure space, let* $\mathrm{f} \in \mathsf{L}^{(1)}((\mathrm{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$, *and let* $\phi \colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ *be such that* $\phi(\mathrm{y}_1) \leq \mathrm{g}(\mathrm{y}_2)$ *for all* $\mathrm{y}_1, \mathrm{y}_2 \in \mathrm{image}(\mathrm{f})$ *with* $\mathrm{y}_1 < \mathrm{y}_2$. *Then, for any* $\mathrm{a} \in \mathbb{R}$ *for which* $\phi(\mathrm{a}) \in \mathbb{R}$, *it holds that*

$$\mu(\{\mathrm{x} \in \mathrm{X} \mid \mathrm{f}(\mathrm{x}) \geq \mathrm{a}\}) \leq \frac{1}{\phi(\mathrm{a})}\mathrm{mean}(\phi \circ \mathrm{f}).$$

Proof Let

$$M_a = \{x \in \mathrm{X} \mid f(x) \geq a\}.$$

Then, for any $x \in X$,

$$\phi(a)\chi_{M_a}(x) \leq \phi \circ f(x)\chi_{M_a}(x) \leq \phi \circ f(x),$$

[12]Andrei Andreyevich Markov (1856–1922) did mathematical research in analysis, and was one of the pioneers in the early development of what we now know as probability theory. He also involved himself in the political turmoil in which Russia was involved during his lifetime.

[13]Pafnuty Lvovich Chebyshev (1821–1894) was a Russian mathematician, making contributions to the areas of analysis, number theory, and approximation theory, and was one of the early researchers in the area of modern probability theory.

noting that $\phi \circ f(x) \geq \phi(a)$ for $x \in M_a$ since ϕ is monotonically increasing. Using Proposition 2.7.19, just as in the proof of Markov's inequality, we have

$$\phi(a)\mu(M_a) \leq \text{mean}(\phi \circ f),$$

as desired. ■

The usual form of Chebychev's inequality is the following.

2.7.36 Corollary (Usual form of Chebychev's inequality) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space and let* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R}_{\geq 0})$*. Then, for any* $\mathsf{a} \in \mathbb{R}_{>0}$ *it holds that*

$$\mu(\{\mathsf{x} \in \mathsf{X} \mid |\mathsf{f}(\mathsf{x})| \geq \mathsf{a}\}) \leq \frac{1}{\mathsf{a}^2} \int_{\mathsf{X}} \mathsf{f}^2 \, \mathrm{d}\mu.$$

Proof Applying the general form of Chebychev's inequality with

$$\phi(y) = \begin{cases} y^2, & y \in \overline{\mathbb{R}}_{\geq 0}, \\ 0, & \text{otherwise} \end{cases}$$

and replacing f with $|f|$ gives the result. ■

Our final result of this form is the following result which follows from our general for the Chebychev inequality. For $c \in \mathbb{R}$ let us denote $\exp_c \colon \overline{\mathbb{R}} \to \overline{\mathbb{R}}_{\geq 0}$ by

$$\exp_c(y) = \begin{cases} \mathrm{e}^{cy}, & y \in \mathbb{R}_{\geq 0}, \\ \lim_{y \to -\infty} \mathrm{e}^{cy}, & y = -\infty, \\ \lim_{y \to \infty} \mathrm{e}^{cy}, & y = \infty, \end{cases}$$

allowing that one of the limits will be ∞. With this notation we have the following result.

2.7.37 Corollary (Chernoff's[14] inequality) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space and let* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$*. Then, for any* $\mathsf{a}, \mathsf{c} \in \mathbb{R}_{>0}$*, it holds that*

$$\mu(\{\mathsf{x} \in \mathsf{X} \mid \mathsf{f}(\mathsf{x}) \geq \mathsf{a}\}) \leq \mathrm{e}^{-\mathsf{ca}} \int_{\mathsf{X}} \exp_{\mathsf{c}} \circ \mathsf{f} \, \mathrm{d}\mu.$$

Proof Applying the general form of Chebychev's inequality with $\phi = \exp_c$ gives the result. ■

Note that it might very well be the case that the right-hand side of either of the inequalities in the preceding two corollaries will be infinite. In this case the inequalities hold vacuously, and so do not give useful information.

[14] Herman Chernoff, born in New York in 1923, is an American statistician.

2.7.5 Topological characterisations of limit theorems[15]

It turns out that there is a very simple way to restate usual version of the Dominated Convergence Theorem using the notion of a limit structure for almost everywhere pointwise convergence from Theorem 2.6.51. For this purpose, it is advantageous to have at hand two versions of the Dominated Convergence Theorem. One is that stated as Theorem 2.7.28, and the other, an "everywhere" rather than an "almost everywhere" version, being the following.

2.7.38 Theorem ("Everywhere" Dominated Convergence Theorem) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space and let* $(\mathsf{f}_j)_{j\in\mathbb{Z}_{>0}}$ *be a sequence in* $\mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \mathbb{R})$ *having the following properties:*

(i) the limit $\mathsf{f}(x) = \lim_{j\to\infty} \mathsf{f}_j(x)$ *exists for every* $x \in \mathsf{X}$*;*

(ii) there exists $g \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}}_{\geq 0})$ *such that, for every* $x \in \mathsf{X}$*,* $|\mathsf{f}_j(x)| \leq g(x)$ *for every* $j \in \mathbb{Z}_{>0}$*.*

Then the functions f *and* f_j*,* $j \in \mathbb{Z}_{>0}$*, are integrable and*

$$\int_{\mathsf{X}} \mathsf{f}\, d\mu = \lim_{j\to\infty} \int_{\mathsf{X}} \mathsf{f}_j\, d\mu.$$

Proof This follows immediately from Theorem 2.7.28. ■

Our objective is to restate the "everywhere" and "almost everywhere" versions of the Dominated Convergence Theorem in topological terms. First let us consider the "everywhere" version of the Dominated Convergence Theorem, Theorem 2.7.38. In this case we use the topology C_p of pointwise convergence on $\mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ described in Section 1.9.1. Note that Proposition 2.6.18 implies that $\mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ is a sequentially closed subspace of $\mathbb{R}^X$ using this topology. Let us say that a subset $A \subseteq \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ is ***$\mathbf{C_p}$-sequentially closed*** if every C_p convergent sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in A converges to a function in A. A subset $B \subseteq \mathbb{R}^X$ is ***$\mathbf{C_p}$-bounded*** if, for every sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in B and every sequence $(a_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}$ converging to 0, the sequence $(a_j f_j)_{j\in\mathbb{Z}_{>0}}$ converges to the zero function in the C_p-topology. This notion of boundedness may look strange at present. We shall examine the general context from which this definition is derived in . where?

The following result characterises C_p-bounded sets.

2.7.39 Proposition (Characterisation of $\mathbf{C_p}$-bounded functions) *Let* X *be a set. A subset* $\mathsf{B} \subseteq \mathbb{R}^{\mathsf{X}}$ *is* C_p*-bounded if and only if there exists a nonnegative-valued* $g \in \mathbb{R}^{\mathsf{X}}$ *such that*

$$\mathsf{B} \subseteq \{\mathsf{f} \in \mathbb{R}^{\mathsf{X}} \mid |\mathsf{f}(x)| \leq g(x) \textit{ for every } x \in \mathsf{X}\}.$$

Proof Suppose that there exists a nonnegative-valued $g \in \mathbb{R}^X$ such that $|f(x)| \leq g(x)$ for every $x \in X$ if $f \in B$. Let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in B and let $(a_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence

[15] The results in this section are not used in an essential way elsewhere in the text, except in Section 2.9.11.

in $\mathbb{R}$ converging to 0. If $x \in X$ then

$$\lim_{j\to\infty} |a_j f_j(x)| \le \lim_{j\to\infty} |a_j| g(x) = 0,$$

which gives C_p-convergence of the sequence $(a_j f_j)_{j\in\mathbb{Z}_{>0}}$ to zero.

Next suppose that there exists no nonnegative-valued function $g \in \mathbb{R}^X$ such that $|f(x)| \le g(x)$ for every $x \in X$ if $f \in B$. This means that there exists $x_0 \in X$ such that, for every $M \in \mathbb{R}_{>0}$, there exists $f \in B$ such that $|f(x_0)| > M$. Let $(a_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}$ converging to 0 and such that $a_j \ne 0$ for every $j \in \mathbb{Z}_{>0}$. Then let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in B such that $|f_j(x_0)| > |a_j^{-1}|$ for every $j \in \mathbb{Z}_{>0}$. Then $|a_j f_j(x_0)| > 1$ for every $j \in \mathbb{Z}_{>0}$, implying that the sequence $(a_j f_j)_{j\in\mathbb{Z}_{>0}}$ cannot C_p-converge to zero. Thus B is not C_p-bounded. ■

With the preceding development, we can now state the "everywhere" Dominated Convergence Theorem in terms of the C_p-topology.

2.7.40 Theorem (Topological "everywhere" Dominated Convergence Theorem) *If* $(\mathsf{X}, \mathscr{A}, \mu)$ *is a measure space then* C_p*-bounded subsets of* $\mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R})$ *are* C_p*-sequentially closed.*

Proof This follows immediately from Theorem 2.7.38 and the definitions of the terms involved. ■

Now we turn to the "almost everywhere" Dominated Convergence Theorem. Here matters are possibly (and often) complicated by the fact that almost everywhere pointwise convergence is not topological, as shown in Proposition 2.6.48. However, we can effectively replace the rôle of the C_p-topology above with the $\mathscr{L}_\mu$-limit structure. To this end, let us say that a subset $A \subseteq \mathsf{L}^0((X, \mathscr{A}); \mathbb{R})$ is ***$\mathscr{L}_\mu$-sequentially closed*** if every $\mathscr{L}_\mu$ convergent sequence $([f_j])_{j\in\mathbb{Z}_{>0}}$ in A converges to an equivalence class of functions in A. A subset $B \subseteq \mathsf{L}^0((X, \mathscr{A}); \mathbb{R})$ is ***$\mathscr{L}_\mu$-bounded*** if, for every sequence $([f_j])_{j\in\mathbb{Z}_{>0}}$ in B and every sequence $(a_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}$ converging to 0, the sequence $([a_j f_j])_{j\in\mathbb{Z}_{>0}}$ converges to the zero equivalence class in the $\mathscr{L}_\mu$-topology.

The following result characterises $\mathscr{L}_\mu$-bounded sets.

2.7.41 Proposition *A subset* $\mathsf{B} \subseteq \mathsf{L}^0((\mathsf{X}, \mathscr{A}); \mathbb{R})$ *is* $\mathscr{L}_\mu$*-bounded if and only if there exists a nonnegative-valued* $\mathsf{g} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \mathbb{R})$ *such that*

$$\mathsf{B} \subseteq \{[\mathsf{f}] \in \mathsf{L}^0((\mathsf{X}, \mathscr{A}); \mathbb{R}) \mid |\mathsf{f}(\mathsf{x})| \le \mathsf{g}(\mathsf{x}) \textit{ for almost every } \mathsf{x} \in \mathsf{X}\}.$$

Proof We first observe that the condition that $|f(x)| \le g(x)$ for almost every $x \in X$ is independent of the choice of representative f from the equivalence class $[f]$.

Suppose that there exists a nonnegative-valued $g \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R})$ such that, if $[f] \in B$, then $|f(x)| \le g(x)$ for almost every $x \in X$. Let $([f_j])_{j\in\mathbb{Z}_{>0}}$ be a sequence in B and let $(a_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}$ converging to zero. For $j \in \mathbb{Z}_{>0}$ define

$$A_j = \{x \in X \mid |f_j(x)| \le g(x)\}.$$

Note that if $x \in X \setminus (\cup_{j\in\mathbb{Z}_{>0}} A_j)$ then

$$\lim_{j\to\infty} |a_j f_j(x)| \le \lim_{j\to\infty} |a_j| g(x) = 0.$$

Since $\mu(\cup_{j\in\mathbb{Z}_{>0}} A_j) = 0$ this implies that the sequence $(a_j[f_j])_{j\in\mathbb{Z}_{>0}}$ is $\mathscr{L}_\mu$-convergent to zero. One may show that this argument is independent of the choice of representatives f_j from the equivalence classes $[f_j]$, $j \in \mathbb{Z}_{>0}$.

Conversely, suppose that there exists no nonnegative-valued function $g \in \mathsf{L}^{(0)}((X,\mathscr{A});\mathbb{R})$ such that, for every $[f] \in B$, $|f(x)| \le g(x)$ for almost every $x \in X$. This means that there exists a set $E \subseteq X$ of positive measure such that, for any $M \in \mathbb{R}_{>0}$, there exists $[f] \in B$ such that $|f(x)| > M$ for almost every $x \in E$. Let $(a_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}$ converging to 0 and such that $a_j \neq 0$ for every $j \in \mathbb{Z}_{>0}$. Then let $([f_j])_{j\in\mathbb{Z}_{>0}}$ be a sequence in B such that $|f_j(x)| > |a_j^{-1}|$ for almost every $x \in E$ and for every $j \in \mathbb{Z}_{>0}$. Define

$$A_j = \{x \in E \mid \ |f_j(x)| > |a_j^{-1}|\}.$$

If $x \in E \setminus (\cup_{j\in\mathbb{Z}_{>0}} A_j)$ then $|a_j f_j(x)| > 1$ for every $j \in \mathbb{Z}_{>0}$. Since $\mu(E \setminus (\cup_{j\in\mathbb{Z}_{>0}} A_j)) > 0$ it follows that $(a_j[f_j])_{j\in\mathbb{Z}_{>0}}$ cannot $\mathscr{L}_\mu$-converge to zero, and so B is not $\mathscr{L}_\mu$-bounded. ■

We can then state the following characterisation of the "almost everywhere" Dominated Convergence Theorem. We denote by $\mathsf{L}^1((X,\mathscr{A},\mu);\mathbb{R})$ the image of $\mathsf{L}^{(1)}((X,\mathscr{A},\mu);\mathbb{R})$ under the projection from $\mathsf{L}^{(0)}((X,\mathscr{A});\mathbb{R})$ to $\mathsf{L}^0((X,\mathscr{A});\mathbb{R})$. Thus elements of $\mathsf{L}^1((X,\mathscr{A},\mu);\mathbb{R})$ are equivalence classes of integrable $\mathbb{R}$-valued functions under the equivalence relation of almost everywhere equality. The space $\mathsf{L}^1((X,\mathscr{A},\mu);\mathbb{R})$ will be studied in detail as part of Section 3.8.8.

2.7.42 Theorem (Limit structure "almost everywhere" Dominated Convergence Theorem) *If* $(\mathrm{X},\mathscr{A},\mu)$ *is a measure space then* $\mathscr{L}_\mu$*-bounded subsets of* $\mathsf{L}^1((\mathrm{X},\mathscr{A},\mu);\mathbb{R})$ *are* $\mathscr{L}_\mu$*-sequentially closed.*

Proof This follows immediately from Theorem 2.7.28 and the definitions of the terms involved. ■

2.7.6 Image measure and integration by image measure

In this section we provide the definition of a measure induced by a map. We shall not use this construction frequently, but it does arise, for example, in parts of our discussion of convolution in Chapter IV-4. anywhere else?

The construction is as follows.

2.7.43 Proposition (Characterisation of image measure) *Let* $(\mathrm{X},\mathscr{A},\mu)$ *be a measure space, let* $(\mathrm{Y},\mathscr{B})$ *be a measurable space, and let* $\phi\colon \mathrm{X}\to\mathrm{Y}$ *be a* $(\mathscr{A},\mathscr{B})$*-measurable map. If we define* $\mu\phi^{-1}\colon \mathscr{B}\to\overline{\mathbb{R}}_{\ge 0}$ *by* $\mu\phi^{-1}(\mathrm{B}) = \mu(\phi^{-1}(\mathrm{B}))$*, then* $(\mathrm{Y},\mathscr{B},\mu\phi^{-1})$ *is a measure space.*

Proof Since $\phi^{-1}(\varnothing) = \varnothing$ we have $\mu\phi^{-1}(\varnothing) = 0$. Now let $(B_j)_{j\in\mathbb{Z}_{>0}}$ be a pairwise disjoint family of subsets from $\mathscr{B}$. We claim that $(\phi^{-1}(B_j))_{j\in\mathbb{Z}_{>0}}$ is pairwise disjoint. This follows since $\phi^{-1}(B_j)\cap\phi^{-1}(B_k) = \phi^{-1}(B_j\cap B_k)$ by Proposition I-1.3.5. It, therefore, follows that

$$\sum_{j=1}^\infty \mu\phi^{-1}(B_j) = \sum_{j=1}^\infty \mu(\phi^{-1}(B_j)) = \mu\Big(\bigcup_{j\in\mathbb{Z}_{>0}} \phi^{-1}(B_j)\Big) = \mu\phi^{-1}\Big(\bigcup_{j\in\mathbb{Z}_{>0}} B_j\Big),$$

again with an application of Proposition I-1.3.5. ■

The measure $\mu\phi^{-1}$ has a name.

2.7.44 Definition (Image measure) For $(X, \mathscr{A}, \mu)$, $(Y, \mathscr{B})$, and ϕ as in Proposition 2.7.43, the measure $\mu\phi^{-1}$ is the ***image measure*** of μ by ϕ. •

One can characterise the functions integrable by the image measure.

2.7.45 Proposition (Integration by the image measure) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space, let* $\phi\colon \mathsf{X} \to \mathsf{Y}$ *be a* $(\mathsf{Y}, \mathscr{B})$ *be a* $(\mathscr{A}, \mathscr{B})$*-measurable map, and let* $\mu\phi^{-1}$ *be the image measure of* μ *by* ϕ*. Then* $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{Y}, \mathscr{B}); \overline{\mathbb{R}})$ *is integrable with respect to* $\mu\phi^{-1}$ *if and only if* $\mathsf{f} \circ \phi$ *is integrable with respect to* μ*. Moreover, if* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{Y}, \mathscr{B}, \mu\phi^{-1}); \overline{\mathbb{R}})$ *then we have*

$$\int_{\mathsf{Y}} \mathsf{f}\, \mathrm{d}(\mu\phi^{-1}) = \int_{\mathsf{X}} (\mathsf{f} \circ \phi)\, \mathrm{d}\mu.$$

Proof Suppose that f is $\mu\phi^{-1}$-integrable. By Proposition 2.6.6 this means that f is $(\mathscr{B}, \mathscr{B}(\overline{\mathbb{R}}))$-measurable. Since ϕ is $(\mathscr{A}, \mathscr{B})$-measurable, it follows easily that $f \circ \phi$ is $(\mathscr{A}, \mathscr{B}(\overline{\mathbb{R}}))$-measurable, and so measurable.

Now let $B \in \mathscr{B}$ and note that $\chi_B \circ \phi = \chi_{\phi^{-1}(B)}$, as can be directly verified. Therefore,

$$\int_Y \chi_B\, \mathrm{d}(\mu\phi^{-1}) = \mu\phi^{-1}(B) = \mu(\phi^{-1}(B)) = \int_X \chi_{\phi^{-1}(B)}\, \mathrm{d}\mu = \int_X \chi_B \circ \phi\, \mathrm{d}\mu.$$

By linearity of the integral, Proposition 2.7.17, this implies that if $f \in \mathsf{L}^{(0)}((Y, \mathscr{B}); \overline{\mathbb{R}})$ is a simple function we have

$$\int_Y f\, \mathrm{d}(\mu\phi^{-1}) = \int_X (f \circ \phi)\, \mathrm{d}\mu. \tag{2.18}$$

If $f \in \mathsf{L}^{(1)}((Y, \mathscr{B}, \mu\phi^{-1}); \overline{\mathbb{R}}_{\geq 0})$ then by Proposition 2.6.39 there exists a sequence of monotonically increasing simple functions $(g_j)_{j \in \mathbb{Z}_{>0}}$ such that $f(y) = \lim_{j\to\infty} g_j(y)$ for each $y \in Y$. The sequence $(g_j \circ \phi)_{j \in \mathbb{Z}_{>0}}$ is then itself a sequence of monotonically increasing functions such that $f \circ \phi(x) = \lim_{j\to\infty} g_j \circ \phi(x)$. By the Monotone Convergence Theorem, (2.18) then holds for $f \in \mathsf{L}^{(1)}((Y, \mathscr{B}, \mu\phi^{-1}); \overline{\mathbb{R}}_{\geq 0})$. For general integrable functions, breaking the function f into its positive and negative parts and using linearity of the integral gives (2.18) in this case. This shows that if $f \in \mathsf{L}^{(1)}((Y, \mathscr{B}, \mu\phi^{-1}); \overline{\mathbb{R}})$ then $f \circ \phi \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ and the functions have equal integrals.

The argument above also clearly shows that if $f \circ \phi \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ then $f \in \mathsf{L}^{(1)}((Y, \mathscr{B}, \mu\phi^{-1}); \overline{\mathbb{R}})$, as desired. ■

2.7.7 The integral for $\mathbb{C}$- and vector-valued functions

Thus far, we have always assumed that functions take values in $\overline{\mathbb{R}}$ or subsets of $\overline{\mathbb{R}}$. Some of the time, however, we wish to integrate functions that are vector-valued, or particularly $\mathbb{C}$-valued. The extension to these sorts of functions is easily made, and in this section we write the (hopefully) expected results. The reader will wish to recall our discussion in Section 2.6.4 of measurable vector-valued functions.

We begin with the definitions.

2.7.46 Definition (Integrable vector-valued function) For a measure space $(X, \mathscr{A}, \mu)$, a function $\boldsymbol{f}\colon X \to \mathbb{R}^n$ is ***integrable*** if its components $f_1, \dots, f_n$ are integrable in the sense of Definition 2.7.8. The ***integral*** of an integrable function $\boldsymbol{f}\colon X \to \mathbb{R}^n$ is

$$\int_X \boldsymbol{f}\,\mathrm{d}\mu = \Big(\int_X f_1\,\mathrm{d}\mu, \dots, \int_X f_n\,\mathrm{d}\mu\Big).$$

We denote the set of integrable $\mathbb{R}^n$-valued maps by $\mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$. •

The following result gives a useful characterisation of the integrability of $\mathbb{R}^n$-valued functions.

2.7.47 Proposition (Characterisation of vector-valued integrable functions) *For a measure space* $(\mathrm{X}, \mathscr{A}, \mu)$ *and* $\mathbf{f} \in \mathsf{L}^{(0)}((\mathrm{X}, \mathscr{A}); \mathbb{R}^{\mathrm{n}})$*, the following statements are equivalent:*

(i) $\mathbf{f} \in \mathsf{L}^{(1)}((\mathrm{X}, \mathscr{A}, \mu); \mathbb{R}^{\mathrm{n}})$*;*

(ii) the $\mathbb{R}$*-valued function* $\mathrm{x} \mapsto \|\mathbf{f}\|_{\mathbb{R}^{\mathrm{n}}}(\mathrm{x})$ *is integrable.*

Moreover, if either of the above equivalent conditions holds, then

$$\Big\|\int_{\mathrm{X}} \mathbf{f}\,\mathrm{d}\mu\Big\|_{\mathbb{R}^{\mathrm{n}}} \le \int_{\mathrm{X}} \|\mathbf{f}\|_{\mathbb{R}^{\mathrm{n}}}\,\mathrm{d}\mu.$$

Proof (i) $\Longrightarrow$ (ii) By Proposition 2.6.11 and Corollary 2.6.33 it follows that $x \mapsto \|\boldsymbol{f}(x)\|_{\mathbb{R}^n}$ is measurable. From Lemma II-1.2.67 we have

$$\|\boldsymbol{f}(x)\|_{\mathbb{R}^n} \le |f_1(x)| + \dots + |f_n(x)|$$

for every $x \in X$. Therefore, by Propositions 2.7.17 and 2.7.19,

$$\int_X \|\boldsymbol{f}\|_{\mathbb{R}^n}\,\mathrm{d}\mu \le \int_X |f_1|\,\mathrm{d}\mu + \dots + \int_X |f_n|\,\mathrm{d}\mu < \infty,$$

giving the result.

(ii) $\Longrightarrow$ (i) From Lemma II-1.2.67 we have

$$|f_1(x)| + \dots + |f_n(x)| \le \sqrt{n}\|\boldsymbol{f}(x)\|_{\mathbb{R}^n}$$

for every $x \in X$. Therefore, by Proposition 2.7.19, for each $j \in \{1, \dots, n\}$ we have

$$\int_X |f_j|\,\mathrm{d}\mu \le \int_X \|\boldsymbol{f}\|_{\mathbb{R}^n}\,\mathrm{d}\mu < \infty,$$

as desired.

Now we prove the final assertion of the proposition. The inequality obviously holds if $\int_X \boldsymbol{f}\,\mathrm{d}\mu = \boldsymbol{0}$, so we may suppose that $\int_X \boldsymbol{f}\,\mathrm{d}\mu \ne \boldsymbol{0}$. Let $\boldsymbol{u} \in \mathbb{R}^n$ be such that $\|\boldsymbol{u}\|_{\mathbb{R}^n} = 1$ and

$$\int_X \boldsymbol{f}\,\mathrm{d}\mu = \boldsymbol{u}\Big\|\int_X \boldsymbol{f}\,\mathrm{d}\mu\Big\|_{\mathbb{R}^n}.$$

Therefore, using linearity of the integral and the fact that $\langle \boldsymbol{u}, \boldsymbol{u}\rangle_{\mathbb{R}^n} = 1$,

$$\int_X \langle \boldsymbol{u}, \boldsymbol{f}\rangle_{\mathbb{R}^n}\,\mathrm{d}\mu = \Big\langle \boldsymbol{u}, \int_X \boldsymbol{f}\,\mathrm{d}\mu\Big\rangle_{\mathbb{R}^n} = \Big\|\int_X \boldsymbol{f}\,\mathrm{d}\mu\Big\|_{\mathbb{R}^n}.$$

Since $|u_j| \le 1$ for each $j \in \{1,\dots,n\}$ we can use the Cauchy–Bunyakovsky–Schwarz inequality and Lemma II-1.2.67 to get

$$\langle \boldsymbol{u}, \boldsymbol{f}(x)\rangle_{\mathbb{R}^n} \le |\langle \boldsymbol{u}, \boldsymbol{f}(x)\rangle_{\mathbb{R}^n}| \le \|\boldsymbol{u}\|_{\mathbb{R}^n}\|\boldsymbol{f}(x)\|_{\mathbb{R}^n} = \|\boldsymbol{f}(x)\|_{\mathbb{R}^n}.$$

Therefore, by Proposition 2.7.19,

$$\Big\|\int_X \boldsymbol{f}\,\mathrm{d}\mu\Big\|_{\mathbb{R}^n} \le \int_X \|\boldsymbol{f}(x)\|_{\mathbb{R}^n}\,\mathrm{d}\mu,$$

as desired. ■

This result has the following immediate and useful corollary which gives an easy means of checking the integrability of a vector-valued function.

2.7.48 Corollary (Integrability of vector-valued functions) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space and let* $\mathbf{f} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \mathbb{R}^{\mathsf{n}})$ *and* $\mathsf{g} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}), \mathbb{R}_{\ge 0})$ *satisfy* $\|\mathbf{f}(\mathsf{x})\|_{\mathbb{R}^{\mathsf{n}}} \le \mathsf{g}(\mathsf{x})$ *for almost every* $\mathsf{x} \in \mathsf{X}$*. Then* $\mathbf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R}^{\mathsf{n}})$ *if* $\mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R}_{\ge 0})$ *and, in this case,*

$$\Big\|\int_{\mathsf{X}} \mathbf{f}\,\mathrm{d}\mu\Big\|_{\mathbb{R}^{\mathsf{n}}} \le \int_{\mathsf{X}} \mathsf{g}\,\mathrm{d}\mu.$$

Of course, the preceding definition and characterisation of integrable vector-valued functions applies immediately to $\mathbb{C}$-valued functions, using the fact that $\mathbb{C}$ and $\mathbb{R}^2$ are isomorphic as $\mathbb{R}$-vector spaces.

2.7.49 Definition (Integrable $\mathbb{C}$-valued function) For a measure space $(X, \mathscr{A}, \mu)$, a function $f\colon X \to \mathbb{C}$ is ***integrable*** if the $\mathbb{R}$-valued functions

$$\mathrm{Re}(f)\colon x \mapsto \mathrm{Re}(f(x)), \quad \mathrm{Im}(f)\colon x \mapsto \mathrm{Im}(f(x))$$

are integrable in the sense of Definition 2.7.8. The ***integral*** of an integrable function $f\colon X \to \mathbb{C}$ is

$$\int_X f\,\mathrm{d}\mu = \Big(\int_X \mathrm{Re}(f)\,\mathrm{d}\mu, \int_X \mathrm{Im}(f)\,\mathrm{d}\mu\Big).$$

We denote the set of integrable $\mathbb{C}$-valued maps by $\mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{C})$. •

Following immediately from Proposition 2.7.47 is the following result.

2.7.50 Corollary (Characterisation of $\mathbb{C}$-valued integrable functions) *For a measure space* $(X, \mathscr{A}, \mu)$ *and* $f \in L^{(0)}((X, \mathscr{A}); \mathbb{C})$*, the following statements are equivalent:*

(i) $f \in L^{(1)}((X, \mathscr{A}, \mu); \mathbb{C})$*;*

(ii) the $\mathbb{R}$*-valued function* $x \mapsto |f|(x)$ *is integrable.*

Moreover, if either of the above equivalent statements holds then

$$\left|\int_X f\,d\mu\right| \le \int_X |f|\,d\mu.$$

Most of the properties of the integral generalise to vector- or $\mathbb{C}$-valued integrals. For completeness we record the results explicitly for $\mathbb{R}^n$-valued functions, noting that these results apply immediately to $\mathbb{C}$-valued functions.

The following result is fundamental and often used without explicit mention.

2.7.51 Proposition (Integrals of functions agreeing almost everywhere) *Let* $(X, \mathscr{A}, \mu)$ *be a measure space and let* $\mathbf{f}, \mathbf{g} \in L^{(0)}((X, \mathscr{A}); \mathbb{R}^n)$ *have the property that* $\mathbf{f}(x) = \mathbf{g}(x)$ *for almost every* $x \in X$*. Then the integral of* $\mathbf{f}$ *exists if and only if the integral of* $\mathbf{g}$ *exists, and if either integral exists then we have*

$$\int_X \mathbf{f}\,d\mu = \int_X \mathbf{g}\,d\mu.$$

Proof This follows immediately from Proposition 2.7.11, along with the definition of the integral for vector-valued functions. ■

Next let us see that the vector-valued integral is linear.

2.7.52 Proposition (Algebraic operations on integrable functions) *For a measure space* $(X, \mathscr{A}, \mu)$*, for* $\mathbf{f}, \mathbf{g} \in L^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$*, and for* $\alpha \in \mathbb{R}$*, the following statements hold:*

(i) $\mathbf{f} + \mathbf{g} \in L^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$ *and*

$$\int_X (\mathbf{f} + \mathbf{g})\,d\mu = \int_X \mathbf{f}\,d\mu + \int_X \mathbf{g}\,d\mu;$$

(ii) $\alpha\mathbf{f} \in L^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$ *and*

$$\int_X \alpha\mathbf{f}\,d\mu = \alpha \int_X \mathbf{f}\,d\mu.$$

Proof This follows directly from Proposition 2.7.17 and the definition of the vector-valued integral. ■

It is also useful to know that the integral of $\mathbb{C}$-valued functions is $\mathbb{C}$-linear.

2.7.53 Corollary (Linearity of the $\mathbb{C}$ integral) *For a measure space $(X, \mathscr{A}, \mu)$, for $f, g \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{C})$, and for $\alpha \in \mathbb{C}$, the following statements hold:*

(i) $f + g \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{C})$ *and*

$$\int_X (f + g)\,d\mu = \int_X f\,d\mu + \int_X g\,d\mu;$$

(ii) $\alpha f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{C})$ *and*

$$\int_X \alpha f\,d\mu = \alpha \int_X f\,d\mu.$$

Proof The first assertion is a special case of the first assertion of Proposition 2.7.52. The second assertion also follows from Proposition 2.7.52 since

$$\operatorname{Re}(\alpha f) = \operatorname{Re}(\alpha)\operatorname{Re}(f) - \operatorname{Im}(\alpha)\operatorname{Im}(f), \qquad \operatorname{Im}(\alpha f) = \operatorname{Re}(\alpha)\operatorname{Im}(f) + \operatorname{Im}(\alpha)\operatorname{Re}(f). \quad \blacksquare$$

For integrating vector-valued functions over disjoint subsets, we have the following result.

2.7.54 Proposition (Breaking the integral in two) *Let $(X, \mathscr{A}, \mu)$, let $A, B \in \mathscr{A}$ be sets such that $X = A \mathbin{\mathring{\cup}} B$, and let $\mathbf{f} \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R}^n)$. Then $\mathbf{f} \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$ if and only if $\mathbf{f}|A \in \mathsf{L}^{(1)}((A, \mathscr{A}_A, \mu|\mathscr{A}_A); \mathbb{R}^n)$ and $\mathbf{f}|B \in \mathsf{L}^{(1)}((B, \mathscr{A}_B, \mu|\mathscr{A}_B); \mathbb{R}^n)$. Furthermore, if $\mathbf{f} \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$ then we have*

$$\int_X \mathbf{f}\,d\mu = \int_A (\mathbf{f}|A)\,d\mu_A + \int_B (\mathbf{f}|B)\,d\mu_B.$$

Proof Thus follows from Proposition 2.7.22, along with the definition of the integral for vector-valued functions. ■

As in the scalar case, this result has the following corollary.

2.7.55 Corollary (Breaking the integral almost in two) *Let $(X, \mathscr{A}, \mu)$ be a complete measure space, let $A, B \in \mathscr{A}$ be such that $\mu(A \cap B) = 0$ and such that $X = A \cup B$, and let $\mathbf{f} \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R}^n)$. Then $\mathbf{f} \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$ if and only if $\mathbf{f}|A \in \mathsf{L}^{(1)}((A, \mathscr{A}_A, \mu|\mathscr{A}_A); \mathbb{R}^n)$ and $\mathbf{f}|B \in \mathsf{L}^{(1)}((B, \mathscr{A}_B, \mu|\mathscr{A}_B); \mathbb{R}^n)$. Furthermore, if $\mathbf{f} \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$ then we have*

$$\int_X \mathbf{f}\,d\mu = \int_A (\mathbf{f}|A)\,d\mu_A + \int_B (\mathbf{f}|B)\,d\mu_B.$$

Proof This follows from Proposition 2.7.23, along with the definition of the vector-valued integral. ■

Finally, we can also state a version of the Dominated Convergence Theorem for vector-valued integrals.

2.7.56 Theorem (Vector-valued Dominated Convergence Theorem) *Let* $(\mathsf{X}, \mathscr{A}, \mu)$ *be a measure space and let* $(\mathbf{f}_j)_{j\in\mathbb{Z}_{>0}}$ *be a sequence in* $\mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \mathbb{R}^n)$ *having the following properties:*

(i) *the limit* $\mathbf{f}(x) = \lim_{j\to\infty} \mathbf{f}_j(x)$ *exists for almost every* $x \in \mathsf{X}$*;*

(ii) *there exists* $g \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R}_{\geq 0})$ *such that, for almost every* $x \in \mathsf{X}$, $\|\mathbf{f}_j(x)\|_{\mathbb{R}^n} \leq g(x)$ *for every* $j \in \mathbb{Z}_{>0}$*.*

Then the functions $\mathbf{f}$ *and* $\mathbf{f}_j$, $j \in \mathbb{Z}_{>0}$, *are integrable and*

$$\int_X \mathbf{f}\, d\mu = \lim_{j\to\infty} \int_X \mathbf{f}_j\, d\mu.$$

Proof For $k \in \{1, \dots, n\}$ denote by f_k the kth component of f and by $f_{j,k}$ the kth component of f_j, $j \in \mathbb{Z}_{>0}$. Then, for almost every $x \in X$, we have

$$\begin{aligned} |f_k(x)| &\leq \|f(x)\|_{\mathbb{R}^n} \leq g(x), \qquad k \in \{1, \dots, n\}, \\ |f_{j,k}(x)| &\leq \|f_j(x)\|_{\mathbb{R}^n} \leq g(x), \qquad k \in \{1, \dots, n\},\ j \in \mathbb{Z}_{>0}. \end{aligned}$$

This gives integrability of f and f_j, $j \in \mathbb{Z}_{>0}$, by definition of the vector-valued integral. The final equality of the theorem now follows from the scalar Dominated Convergence Theorem, Theorem 2.7.28. ■

2.7.8 Integration with respect to signed, complex, and vector measures

In this section to this point we have talked solely about positive measure spaces. Let us now see how signed, complex, and vector measure spaces arise in the integration story.

We begin by indicating how one can define integrals with respect to signed, complex, and vector measures. Here we use the Jordan decomposition of such measures in an essential way. Let us consider first the case where $(X, \mathscr{A}, \mu)$ is a signed measure space.

2.7.57 Definition (Integration with respect to a signed measure) For a signed measure space $(X, \mathscr{A}, \mu)$ let $\mu = \mu_+ - \mu_-$ be the Jordan decomposition of μ. For $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$, we have the following definitions.

(i) If neither of the conditions

(a) $\int_X f\, d\mu_+ = \infty$ and $\int_X f\, d\mu_- = \infty$ and

(b) $\int_X f\, d\mu_+ = -\infty$ and $\int_X f\, d\mu_- = -\infty$

holds, then the integral of f with respect to μ ***exists*** and is given by

$$\int_X f\, d\mu = \int_X f\, d\mu_+ - \int_X f\, d\mu_-,$$

this being the ***integral*** of f with respect to μ.

(ii) If either of the two conditions from part (i) hold then the integral of f with respect to μ ***does not exist***.

(iii) If $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu_+); \overline{\mathbb{R}})$ and $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu_-); \overline{\mathbb{R}})$ then f is ***integrable*** with respect to μ.

For a subset $I \subseteq \overline{\mathbb{R}}$ we denote the set of I-valued functions integrable with respect to μ by $\mathsf{L}^{(1)}((X, \mathscr{A}, \mu); I)$, or simply by $\mathsf{L}^{(1)}(X; I)$ if $\mathscr{A}$ and μ are understood. •

Using this definition of integrability and integral for signed measures, it is straightforward to define the corresponding notions for complex and vector measures. The essential idea is that a complex measure μ can be written as

$$\mu = \mathrm{Re}(\mu) + \mathrm{i}\,\mathrm{Im}(\mu)$$

for finite signed measures $\mathrm{Re}(\mu)$ and $\mathrm{Im}(\mu)$. For a vector measure $\boldsymbol{\mu}$ taking values in $\mathbb{R}^n$, we can write

$$\boldsymbol{\mu} = \mu_1 \boldsymbol{e}_1 + \cdots + \mu_n \boldsymbol{e}_n$$

for finite signed measures $\mu_1, \dots, \mu_j$ and where $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is the standard basis for $\mathbb{R}^n$.

2.7.58 Definition (Integration with respect to complex and vector measures) For a measurable space $(X, \mathscr{A})$ and for a complex measure μ on $\mathscr{A}$ and a vector measure $\boldsymbol{\mu}$ taking values in $\mathbb{R}^n$, write them as above. For $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$, we have the following definitions.

(i) the integral of f with respect to μ ***exists*** if the integrals of f with respect to $\mathrm{Re}(\mu)$ and $\mathrm{Im}(\mu)$ exist, and is given by

$$\int_X f\,\mathrm{d}\mu = \Big(\int_X f\,\mathrm{d}(\mathrm{Re}(\mu))\Big) + \mathrm{i}\Big(\int_X f\,\mathrm{d}(\mathrm{Im}(\mu))\Big),$$

this being the ***integral*** of f with respect to μ.

(ii) the integral of f with respect to $\boldsymbol{\mu}$ ***exists*** if the integrals of f with respect to $\mu_1, \dots, \mu_n$ exist, and is given by

$$\int_X f\,\mathrm{d}\boldsymbol{\mu} = \Big(\int_X f\,\mathrm{d}\mu_1, \dots, \int_X f\,\mathrm{d}\mu_n\Big),$$

this being the ***integral*** of f with respect to $\boldsymbol{\mu}$.

(iii) If the integral of f does not exist with respect to at least one of $\mathrm{Re}(\mu)$ and $\mathrm{Im}(\mu)$, then the integral of f ***does not exist***.

(iv) If the integral of f does not exist with respect to at least one of $\mathrm{Re}(\mu)$ and $\mathrm{Im}(\mu)$, then the integral of f ***does not exist***.

(v) If $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mathrm{Re}(\mu)); \overline{\mathbb{R}})$ and $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mathrm{Im}(\mu)); \overline{\mathbb{R}})$ then f is ***integrable*** with respect to μ.

(vi) If $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu_j); \overline{\mathbb{R}})$, $j \in \{1, \dots, n\}$, f is ***integrable*** with respect to $\boldsymbol{\mu}$.

For a subset $I \subseteq \overline{\mathbb{R}}$ we denote the set of I-valued functions integrable with respect to μ (resp. $\boldsymbol{\mu}$) by $\mathsf{L}^{(1)}((X, \mathscr{A}, \mu); I)$ (resp. $\mathsf{L}^{(1)}((X, \mathscr{A}, \boldsymbol{\mu}); I)$), or simply by $\mathsf{L}^{(1)}(X; I)$ if $\mathscr{A}$ and μ (resp. $\boldsymbol{\mu}$) are understood. •

Since, by virtue of the Jordan decomposition, integration with respect to signed, complex, and vector measures boils down to integration with respect to positive measures as usual, one anticipates that many of the properties of the integral with respect to positive measures will carry over to signed, complex, and vector measures. Let us record some of these.

First we relate the integral of a function with the integral with respect to a measure to the integral with respect to the variation of the measure.

2.7.59 Proposition (Characterisation of integrals with respect to signed, complex, and vector measures) *For a measurable space* $(\mathrm{X}, \mathscr{A})$ *and for* $\mathrm{f} \in \mathsf{L}^{(0)}((\mathrm{X}, \mathscr{A}); \overline{\mathbb{R}})$*, the following statements hold:*

(i) if μ *is a signed or complex measure on* $\mathscr{A}$*, then* $\mathrm{f} \in \mathsf{L}^{(1)}((\mathrm{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *if and only if* $\mathrm{f} \in \mathsf{L}^{(1)}((\mathrm{X}, \mathscr{A}, |\mu|); \overline{\mathbb{R}})$*, and if either of these equivalent statements holds, then*

$$\left|\int_{\mathrm{X}} \mathrm{f}\, \mathrm{d}\mu\right| \le \int_{\mathrm{X}} |\mathrm{f}|\, \mathrm{d}|\mu|;$$

(ii) if $\boldsymbol{\mu}$ *is a vector measure on* $\mathscr{A}$ *taking values in* $\mathbb{R}^{\mathrm{n}}$*, then* $\mathrm{f} \in \mathsf{L}^{(1)}((\mathrm{X}, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$ *if and only if* $\mathrm{f} \in \mathsf{L}^{(1)}((\mathrm{X}, \mathscr{A}, \|\mu\|_{\mathbb{R}^{\mathrm{n}}}); \overline{\mathbb{R}})$*, and if either of these equivalent statements holds, then*

$$\left\|\int_{\mathrm{X}} \mathrm{f}\, \mathrm{d}\boldsymbol{\mu}\right\|_{\mathbb{R}^{\mathrm{n}}} \le \int_{\mathrm{X}} |\mathrm{f}|\, \mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^{\mathrm{n}}}.$$

Proof Let us first consider the case where μ is a signed measure on $\mathscr{A}$ with Jordan decomposition $\mu = \mu_+ - \mu_-$. If $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ then, by definition $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu_+); \overline{\mathbb{R}})$ and $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu_-); \overline{\mathbb{R}})$. Therefore,

$$\int_X |f|\, \mathrm{d}|\mu| = \int_X |f|\, \mathrm{d}\mu_+ + \int_X |f|\, \mathrm{d}\mu_- < \infty,$$

and so $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, |\mu|); \overline{\mathbb{R}})$. Conversely, suppose that $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, |\mu|); \overline{\mathbb{R}})$. Then

$$\int_X |f|\, \mathrm{d}|\mu| = \int_X |f|\, \mathrm{d}\mu_+ + \int_X |f|\, \mathrm{d}\mu_- < \infty.$$

Thus $f \in \mathscr{L}^{(1)}((X, \mathscr{A}, \mu_+); \overline{\mathbb{R}})$ and $\in \mathscr{L}^{(1)}((X, \mathscr{A}, \mu_-); \overline{\mathbb{R}})$. Therefore,

$$\int_X f\, \mathrm{d}\mu = \int_X f\, \mathrm{d}\mu_+ - \int_X f\, \mathrm{d}\mu_-$$

is well-defined, and so $f \in \mathscr{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$.

For the final assertion of this part of the theorem, we compute

$$\begin{aligned}\Big|\int_X f\,\mathrm{d}\mu\Big| = \Big|\int_X f\,\mathrm{d}\mu_+ - \int_X f\,\mathrm{d}\mu_-\Big| &\le \Big|\int_X f\,\mathrm{d}\mu_+\Big| + \Big|\int_X f\,\mathrm{d}\mu_-\Big| \\ &\le \int_X |f|\,\mathrm{d}\mu_+ + \int_X |f|\,\mathrm{d}\mu_- = \int_X |f|\,\mathrm{d}|\mu|,\end{aligned}$$

as claimed.

Now we consider the case of a vector measure $\boldsymbol{\mu}$, the case of a complex measure following from this as a special case. Suppose first that $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$ so that, by definition, $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu_j); \overline{\mathbb{R}})$ for each $j \in \{1, \dots, n\}$. Let us first suppose that f is a nonnegative-valued simple function. Thus

$$f = \sum_{j=1}^{k} c_j \chi_{A_j}$$

for $c_j \in \overline{\mathbb{R}}_{\ge 0}$, $j \in \{1, \dots, k\}$, and for pairwise disjoint measurable sets A_j, $j \in \{1, \dots, k\}$. Then

$$\int_X f\,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n} = \sum_{j=1}^{k} c_j \|\boldsymbol{\mu}\|_{\mathbb{R}^n}(A_j) \le \sum_{j=1}^{k} c_j \sum_{l=1}^{n} |\mu_l|(A_j),$$

the last inequality holding by (2.8). Noting that

$$\int_X f\,\mathrm{d}|\mu_l| = \sum_{j=1}^{k} c_j |\mu_l|(A_j),$$

we deduce that

$$\int_X f\,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n} \le \sum_{l=1}^{n} \int_X f \mathrm{d}|\mu_l|,$$

giving $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \|\boldsymbol{\mu}\|_{\mathbb{R}^n}); \overline{\mathbb{R}})$ in the case when f is a nonnegative simple function. For a general nonnegative function $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$ we let $(f_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence of nonnegative simple functions such that $f_j(x) \le f_{j+1}(x)$ for $x \in X$ and $j \in \mathbb{Z}_{>0}$ and such that $\lim_{j \to \infty} f_j(x) = f(x)$; see Proposition 2.6.39. Then

$$\int_X f_j\,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n} \le \sum_{l=1}^{n} \int_X f_j \mathrm{d}|\mu_l| \le \sum_{l=1}^{n} \int_X f \mathrm{d}|\mu_l|,$$

the last inequality by Proposition 2.7.19. Thus, by the Monotone Convergence Theorem,

$$\int_X f\,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n} = \lim_{j \to \infty} \int_X f_j\,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n} \le \sum_{l=1}^{n} \int_X f \mathrm{d}|\mu_l|,$$

giving $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \|\boldsymbol{\mu}\|_{\mathbb{R}^n}); \overline{\mathbb{R}})$ for a nonnegative $\boldsymbol{\mu}$-integrable function f. For a general $\boldsymbol{\mu}$-integrable function f we then have

$$\int_X |f|\,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n} = \lim_{j \to \infty} \int_X f_j\,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n} \le \sum_{l=1}^{n} \int_X |f| \mathrm{d}|\mu_l|,$$

giving $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \|\boldsymbol{\mu}\|_{\mathbb{R}^n}); \overline{\mathbb{R}})$.

Now we suppose that $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \|\boldsymbol{\mu}\|_{\mathbb{R}^n}); \overline{\mathbb{R}})$. As above, we first suppose that f is a nonnegative-valued simple function:

$$f = \sum_{j=1}^{k} c_j \chi_{A_j}.$$

For $l \in \{1, \dots, n\}$ we have

$$\begin{aligned}\int_X f \,\mathrm{d}|\mu_l| &\le \sum_{l=1}^{n} \int_X f \,\mathrm{d}|\mu_l| \le \sum_{l=1}^{n} \sum_{j=1}^{k} c_j |\mu_l|(A_j) \\ &\le \sqrt{n} \sum_{j=1}^{k} c_j \|\boldsymbol{\mu}\|_{\mathbb{R}^n}(A_j) = \sqrt{n} \int_X f \,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n},\end{aligned}$$

using Exercise 2.3.6 and Proposition 2.3.55. Thus, for nonnegative simple functions $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \|\boldsymbol{\mu}\|_{\mathbb{R}^n}); \overline{\mathbb{R}})$ we have $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu_l); \overline{\mathbb{R}})$, $l \in \{1, \dots, n\}$, and so $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$. Now one can prove that

$$\int_X f \,\mathrm{d}|\mu_l| \le \sum_{l=1}^{n} \int_X f \,\mathrm{d}|\mu_l| \le \sqrt{n} \int_X f \mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n}$$

for general nonnegative functions $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \|\boldsymbol{\mu}\|_{\mathbb{R}^n}); \overline{\mathbb{R}})$ using an argument involving a sequence of simple functions a $(f_j)_{j \in \mathbb{Z}_{>0}}$ approximating f, just as in the preceding paragraph. Also just as in the preceding paragraph, it follows that, for a general $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \|\boldsymbol{\mu}\|_{\mathbb{R}^n}); \overline{\mathbb{R}})$,

$$\int_X |f| \,\mathrm{d}|\mu_l| \le \sum_{l=1}^{n} \int_X |f| \,\mathrm{d}|\mu_l| \le \sqrt{n} \int_X |f| \,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n}, \tag{2.19}$$

and so $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$.

Moreover, by Proposition II-1.1.11, by the fact that the proposition holds for signed measures, and by (2.19), we have

$$\Big\| \int_X f \,\mathrm{d}\boldsymbol{\mu} \Big\|_{\mathbb{R}^n} \le \sum_{l=1}^{n} \Big| \int_X f \,\mathrm{d}\mu_l \Big| \le \sum_{l=1}^{n} \int_X |f| \,\mathrm{d}|\mu_l| \le \sqrt{n} \int_X |f| \,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n},$$

which gives the final assertion of the proposition. ■

First we can show that the integral depends, in the appropriate sense, on the value of a function up to a set of measure zero.

2.7.60 Proposition (Integrals of functions agreeing almost everywhere) *For a measurable space* $(\mathsf{X}, \mathscr{A})$ *and for* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}})$*, the following statements hold:*

(i) if μ *is a signed or complex measure on* $\mathscr{A}$ *and if*

$$|\mu|(\{\mathsf{x} \in \mathsf{X} \mid \mathsf{f}(\mathsf{x}) \neq \mathsf{g}(\mathsf{x})\}) = 0,$$

then $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *if and only if* $\mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *and, if either of these conditions holds,*

$$\int_{\mathsf{X}} \mathsf{f}\, \mathrm{d}\mu = \int_{\mathsf{X}} \mathsf{g}\, \mathrm{d}\mu;$$

(ii) if $\boldsymbol{\mu}$ *is a vector measure on* $\mathscr{A}$ *taking values in* $\mathbb{R}^{\mathsf{n}}$ *and if*

$$\|\boldsymbol{\mu}\|_{\mathbb{R}^{\mathsf{n}}}(\{\mathsf{x} \in \mathsf{X} \mid \mathsf{f}(\mathsf{x}) \neq \mathsf{g}(\mathsf{x})\}) = 0,$$

then $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$ *if and only if* $\mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$ *and, if either of these conditions holds,*

$$\int_{\mathsf{X}} \mathsf{f}\, \mathrm{d}\boldsymbol{\mu} = \int_{\mathsf{X}} \mathsf{g}\, \mathrm{d}\boldsymbol{\mu}.$$

Proof Let us first consider the case of a signed measure μ. First suppose that $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$. By Proposition 2.7.59 it follows that $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, |\mu|); \overline{\mathbb{R}})$. Since g differs from f on a set whose $|\mu|$-measure is zero, it follows from Proposition 2.7.11 that $g \in \mathsf{L}^{(1)}((X, \mathscr{A}, |\mu|); \overline{\mathbb{R}})$ and so $g \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$, again by Proposition 2.7.59. Of course, the argument is reversible, showing that if $g \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ then $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$. If Z is the set of points where f and g differ, then

$$\left|\int_Z (f - g)\, \mathrm{d}\mu\right| \leq \int_Z |f - g|\, \mathrm{d}|\mu| = 0,$$

the first inequality by Proposition 2.7.59. Therefore, using the Proposition 2.7.62 below, we have

$$\int_X (f - g)\, \mathrm{d}\mu = \int_{X \setminus Z} (f - g)\, \mathrm{d}\mu + \int_Z (f - g)\, \mathrm{d}\mu = \int_{X \setminus Z} (f - g)\, \mathrm{d}\mu = 0.$$

By Proposition 2.7.61 we then have

$$\int_X f\, \mathrm{d}\mu = \int_X g\, \mathrm{d}\mu,$$

giving the first part of the result.

To conclude, we prove the proposition for vector measures, the complex case being a consequence of this. Suppose that Z denotes the set of points where f and g differ. Then

$$|\mu_l|(Z) = \int_X \chi_Z\, \mathrm{d}\mu_l \leq \sqrt{n} \int_X \chi_Z\, \mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n} = 0, \tag{2.20}$$

where we have used (2.19). Then the first part of the proof gives $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu_l); \overline{\mathbb{R}})$ if and only if $g \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu_l); \overline{\mathbb{R}})$ for each $l \in \{1, \dots, n\}$. The definition of the integral with respect to $\boldsymbol{\mu}$, along with the conclusions from the first part of the result, gives

$$\int_X f \, \mathrm{d}\boldsymbol{\mu} = \int_X g \, \mathrm{d}\boldsymbol{\mu},$$

as desired. ■

The following result concerning algebraic operations can be deduced immediately by applying the corresponding result for positive measures to the Jordan decomposition of the measures involved.

2.7.61 Proposition (Algebraic operations for the integral with respect to signed, complex, and signed measures) *For a measurable space* $(\mathsf{X}, \mathscr{A})$*, the following statements hold:*

(i) if μ *is a signed or complex measure and if* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$*, then* $\mathsf{f} + \mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *and*

$$\int_{\mathsf{X}} (\mathsf{f} + \mathsf{g}) \, \mathrm{d}\mu = \int_{\mathsf{X}} \mathsf{f} \, \mathrm{d}\mu + \int_{\mathsf{X}} \mathsf{g} \, \mathrm{d}\mu;$$

(ii) if $\boldsymbol{\mu}$ *is a vector measure taking values in* $\mathbb{R}^{\mathsf{n}}$ *and if* $\mathsf{f}, \mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$*, then* $\mathsf{f} + \mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$ *and*

$$\int_{\mathsf{X}} (\mathsf{f} + \mathsf{g}) \, \mathrm{d}\boldsymbol{\mu} = \int_{\mathsf{X}} \mathsf{f} \, \mathrm{d}\boldsymbol{\mu} + \int_{\mathsf{X}} \mathsf{g} \, \mathrm{d}\boldsymbol{\mu};$$

(iii) if μ *is a signed or complex measure, if* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *and if* $\alpha \in \mathbb{R}$*, then* $\alpha \mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *and*

$$\int_{\mathsf{X}} \alpha \mathsf{f} \, \mathrm{d}\mu = \alpha \int_{\mathsf{X}} \mathsf{f} \, \mathrm{d}\mu;$$

(iv) if $\boldsymbol{\mu}$ *is a vector measure taking values in* $\mathbb{R}^{\mathsf{n}}$*, if* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$*, and if* $\alpha \in \mathbb{R}$*, then* $\alpha \mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$ *and*

$$\int_{\mathsf{X}} \alpha \mathsf{f} \, \mathrm{d}\boldsymbol{\mu} = \alpha \int_{\mathsf{X}} \mathsf{f} \, \mathrm{d}\boldsymbol{\mu}.$$

Proof We first consider the case of a signed measure μ with Jordan decomposition $\mu = \mu_+ - \mu_-$. We then have

$$\begin{aligned} \int_X (f + g) \, \mathrm{d}\mu &= \int_X (f + g) \, \mathrm{d}\mu_+ - \int_X (f - g) \, \mathrm{d}\mu_- \\ &= \int_X f \, \mathrm{d}\mu_+ + \int_X g \, \mathrm{d}\mu_+ - \int_X f \, \mathrm{d}\mu_- - \int_X g \, \mathrm{d}\mu_- \\ &= \int_X f \, \mathrm{d}\mu + \int_X g \, \mathrm{d}\mu \end{aligned}$$

by Proposition 2.7.17. A similarly styled argument gives

$$\int_X \alpha f \,\mathrm{d}\mu = \alpha \int_X f \,\mathrm{d}\mu.$$

The result for vector measures then follows immediately from the result for signed measures by the definition of the integral with respect to a vector measure. The result for complex measures is a special case of the result for vector measures. ■

We can also break integrals with respect to signed, complex, and vector measures into separate integrals over disjoint sets.

2.7.62 Proposition (Breaking the integral in two) *For a measurable space* $(\mathsf{X}, \mathscr{A})$ *let* $\mathsf{A}, \mathsf{B} \in \mathscr{A}$ *be such that* $\mathsf{X} = \mathsf{A} \mathbin{\mathring{\cup}} \mathsf{B}$ *and let* $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}})$*. Then the following statements hold:*

(i) if μ *is a signed or complex measure on* $\mathscr{A}$*, then* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *if and only if* $\mathsf{f}|\mathsf{A} \in \mathsf{L}^{(1)}((\mathsf{A}, \mathscr{A}_\mathsf{A}, \mu|\mathscr{A}_\mathsf{A}); \overline{\mathbb{R}})$ *and* $\mathsf{f}|\mathsf{B} \in \mathsf{L}^{(1)}((\mathsf{B}, \mathscr{A}_\mathsf{B}, \mu|\mathscr{A}_\mathsf{B}); \overline{\mathbb{R}})$*, and if either of these two equivalent conditions holds,*

$$\int_\mathsf{X} \mathsf{f} \,\mathrm{d}\mu = \int_\mathsf{A} (\mathsf{f}|\mathsf{A}) \,\mathrm{d}\mu_\mathsf{A} + \int_\mathsf{B} (\mathsf{f}|\mathsf{B}) \,\mathrm{d}\mu_\mathsf{B};$$

(ii) if $\boldsymbol{\mu}$ *is a vector measure on* $\mathscr{A}$*, then* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$ *if and only if* $\mathsf{f}|\mathsf{A} \in \mathsf{L}^{(1)}((\mathsf{A}, \mathscr{A}_\mathsf{A}, \boldsymbol{\mu}|\mathscr{A}_\mathsf{A}); \overline{\mathbb{R}})$ *and* $\mathsf{f}|\mathsf{B} \in \mathsf{L}^{(1)}((\mathsf{B}, \mathscr{A}_\mathsf{B}, \boldsymbol{\mu}|\mathscr{A}_\mathsf{B}); \overline{\mathbb{R}})$*, and if either of these two equivalent conditions holds,*

$$\int_\mathsf{X} \mathsf{f} \,\mathrm{d}\boldsymbol{\mu} = \int_\mathsf{A} (\mathsf{f}|\mathsf{A}) \,\mathrm{d}\boldsymbol{\mu}_\mathsf{A} + \int_\mathsf{B} (\mathsf{f}|\mathsf{B}) \,\mathrm{d}\boldsymbol{\mu}_\mathsf{B};$$

Proof We first consider the case of a signed measure μ. By Proposition 2.7.22 it follows that $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, |\mu|); \overline{\mathbb{R}})$ if and only if $f|A \in \mathsf{L}^{(1)}((A, \mathscr{A}_A, |\mu||\mathscr{A}_A); \overline{\mathbb{R}})$ and $f|B \in \mathsf{L}^{(1)}((B, \mathscr{A}_B, |\mu||\mathscr{A}_B); \overline{\mathbb{R}})$. By Proposition 2.7.59 it follows that $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$ if and only if $f|A \in \mathsf{L}^{(1)}((A, \mathscr{A}_A, \boldsymbol{\mu}|\mathscr{A}_A); \overline{\mathbb{R}})$ and $f|B \in \mathsf{L}^{(1)}((B, \mathscr{A}_B, \boldsymbol{\mu}|\mathscr{A}_B); \overline{\mathbb{R}})$, as claimed. Moreover, writing $f = f\chi_A + f\chi_B$, we use Proposition 2.7.61 to give

$$\int_X f \,\mathrm{d}\mu = \int_A (f|A) \,\mathrm{d}\mu_A + \int_B (f|B) \,\mathrm{d}\mu_B.$$

The result for vector and complex measures follows immediately from the conclusion for signed measures, using the definition of the integral in these cases. ■

what about the completeness that is needed in the positive case?

2.7.63 Corollary (Breaking the integral almost in two) *For a measurable space* $(\mathsf{X}, \mathscr{A})$ *let* $\mathsf{A}, \mathsf{B} \in \mathscr{A}$ *be such that* $\mathsf{X} = \mathsf{A} \cup \mathsf{B}$ *and let* $\mathsf{f} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \overline{\mathbb{R}})$*. Then the following statements hold:*

(i) if μ *is a signed or complex measure on* $\mathscr{A}$ *and if* $|\mu|(\mathsf{A}) = 0$*, then* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *if and only if* $\mathsf{f}|\mathsf{A} \in \mathsf{L}^{(1)}((\mathsf{A}, \mathscr{A}_\mathsf{A}, \mu|\mathscr{A}_\mathsf{A}); \overline{\mathbb{R}})$ *and* $\mathsf{f}|\mathsf{B} \in \mathsf{L}^{(1)}((\mathsf{B}, \mathscr{A}_\mathsf{B}, \mu|\mathscr{A}_\mathsf{B}); \overline{\mathbb{R}})$*;*

(ii) *if* $\boldsymbol{\mu}$ *is a vector measure on* $\mathscr{A}$ *and if* $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}(A) = 0$*, then* $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$ *if and only if* $f|A \in \mathsf{L}^{(1)}((A, \mathscr{A}_A, \boldsymbol{\mu}|\mathscr{A}_A); \overline{\mathbb{R}})$ *and* $f|B \in \mathsf{L}^{(1)}((B, \mathscr{A}_B, \boldsymbol{\mu}|\mathscr{A}_B); \overline{\mathbb{R}})$*.*

Proof This follows from Propositions 2.7.60 2.7.62. ■

Finally, for signed, complex, and vector measures we have a version of the Dominated Convergence Theorem. Note here that a little care must be exercised in stating the hypotheses.

2.7.64 Theorem (Dominated Convergence Theorem for signed, complex, and vector measures) *For a measurable space* $(X, \mathscr{A})$ *and for a sequence* $(f_j)_{j\in\mathbb{Z}_{>0}}$ *in* $\mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ *the following statements hold:*

(i) *if* μ *is a signed or complex measure and if*

(a) *the limit* $f(x) = \lim_{j\to\infty} f_j(x)$ *exists for* $|\mu|$*-almost every* $x \in X$ *and if*

(b) *there exists* $g \in \mathsf{L}^{(1)}((X, \mathscr{A}, |\mu|); \overline{\mathbb{R}}_{\geq 0})$ *such that, for* $|\mu|$*-almost every* $x \in X$*,* $|f_j|(x) \leq g(x)$ *for every* $j \in \mathbb{Z}_{>0}$*,*

then $f, f_j \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$*,* $j \in \mathbb{Z}_{>0}$*, and*

$$\int_X f\,d\mu = \lim_{j\to\infty} \int_X f_j\,d\mu;$$

(ii) *if* $\boldsymbol{\mu}$ *is a vector measure taking values in* $\mathbb{R}^n$ *and if*

(a) *the limit* $f(x) = \lim_{j\to\infty} f_j(x)$ *exists for* $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}$*-almost every* $x \in X$ *and if*

(b) *there exists* $g \in \mathsf{L}^{(1)}((X, \mathscr{A}, \|\boldsymbol{\mu}\|_{\mathbb{R}^n}); \overline{\mathbb{R}}_{\geq 0})$ *such that, for* $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}$*-almost every* $x \in X$*,* $|f_j|(x) \leq g(x)$ *for every* $j \in \mathbb{Z}_{>0}$*,*

then $f, f_j \in \mathsf{L}^{(1)}((X, \mathscr{A}, \boldsymbol{\mu}); \overline{\mathbb{R}})$*,* $j \in \mathbb{Z}_{>0}$*, and*

$$\int_X f\,d\boldsymbol{\mu} = \lim_{j\to\infty} \int_X f_j\,d\boldsymbol{\mu}.$$

Proof We first consider the case of a signed measure μ with Jordan decomposition $\mu = \mu_+ - \mu_-$. The integrability of f and f_j, $j \in \mathbb{Z}_{>0}$, with respect to μ follows from their assumed integrability with respect to $|\mu|$, along with Proposition 2.7.59. Since $|\mu| = \mu_+ + \mu_-$, it follows that the limit $f(x) = \lim_{j\to\infty} f_j(x)$ exists for μ_+-almost every $x \in X$ and for μ_--almost every $x \in X$. Also, $|\mu|$-integrability of g implies μ_+- and μ_--integrability of g. Finally, we have $|f_j|(x) \leq g(x)$ for μ_+- and μ_--almost every $x \in X$ and for every $j \in \mathbb{Z}_{>0}$. Then we compute

$$\begin{aligned}\lim_{j\to\infty} \int_X f_j\,d\mu &= \lim_{j\to\infty}\Big(\int_X f_j\,\mu_+ - \int_X f_j\,d\mu_-\Big) \\ &= \lim_{j\to\infty} \int_X f_j\,\mu_+ - \lim_{j\to\infty} \int_X f_j\,d\mu_- \\ &= \int_X f\,d\mu_+ - \int_X f\,d\mu_- = \int_X f\,d\mu,\end{aligned}$$

using the Dominated Convergence Theorem for positive measures, along with the commutativity of limits with sums (Proposition I-2.3.23).

We next prove the theorem for the case of a vector measure, noting that the case of complex measures follows from this. As in (2.20), if Z has $\|\boldsymbol{\mu}\|_{\mathbb{R}^n}$-measure zero, then Z also has $|\mu_l|$-measure zero for each $l \in \{1, \dots, n\}$. Therefore, the hypotheses of the theorem give:

1. the limit $f(x) = \lim_{j\to\infty} f_j(x)$ exists for $|\mu_l|$-almost $x \in X$ for each $l \in \{1, \dots, l\}$;
2. $|f_j|(x) \leq g(x)$ for $|\mu_l|$-almost every $x \in X$ for each $j \in \mathbb{Z}_{>0}$ and $l \in \{1, \dots, n\}$.

As we saw in the proof of Proposition 2.7.59,

$$\int_X g \,\mathrm{d}|\mu_l| \leq \sqrt{n} \int_X g \,\mathrm{d}\|\boldsymbol{\mu}\|_{\mathbb{R}^n},$$

and so our hypotheses imply that $g \in \mathsf{L}^{(1)}((X, \mathscr{A}, |\mu_l|); \overline{\mathbb{R}}_{\geq 0})$ for each $l \in \mathbb{Z}_{>0}$. This all implies that the result from the first part of the theorem gives the result for vector measures. ■

We next show how signed, complex, and vector measures can be built from positive measures and integrable functions. This gives us a wealth of signed, complex, and vector measures. We shall see in , moreover, that an important class of measures arise *exactly* in the manner of the next result.

what

2.7.65 Proposition (Signed, complex, and vector measures from functions) *If* $(\mathrm{X}, \mathscr{A}, \mu)$ *is a measure space, then the following statements hold:*

(i) if $\mathrm{f} \in \mathsf{L}^{(1)}((\mathrm{X}, \mathscr{A}, \mu); \mathbb{R})$ *then* $\mathrm{f} \cdot \mu\colon \mathscr{A} \to \overline{\mathbb{R}}$ *defined by*

$$(\mathrm{f} \cdot \mu)(\mathrm{A}) = \int_{\mathrm{X}} \mathrm{f}\chi_{\mathrm{A}} \,\mathrm{d}\mu$$

is a finite signed measure on $\mathscr{A}$*;*

(ii) if $\mathrm{f} \in \mathsf{L}^{(1)}((\mathrm{X}, \mathscr{A}, \mu); \mathbb{C})$ *then* $\mathrm{f} \cdot \mu\colon \mathscr{A} \to \mathbb{C}$ *defined by*

$$(\mathrm{f} \cdot \mu)(\mathrm{A}) = \int_{\mathrm{X}} \mathrm{f}\chi_{\mathrm{A}} \,\mathrm{d}\mu$$

is a complex measure on $\mathscr{A}$*;*

(iii) if $\mathbf{f} \in \mathsf{L}^{(1)}((\mathrm{X}, \mathscr{A}, \mu); \mathbb{R}^{\mathrm{n}})$ *then* $\mathbf{f} \cdot \mu\colon \mathscr{A} \to \mathbb{R}^{\mathrm{n}}$ *defined by*

$$(\mathbf{f} \cdot \mu)(\mathrm{A}) = \int_{\mathrm{X}} \mathbf{f}\chi_{\mathrm{A}} \,\mathrm{d}\mu$$

is a vector measure on $\mathscr{A}$*.*

Proof We prove the statement for vector measures, since the other cases are a special case of this.

It is clear that $(\boldsymbol{f} \cdot \mu)(\varnothing) = 0$. Now let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a family of pairwise disjoint elements of $\mathscr{A}$ and let $A = \cup_{j\in\mathbb{Z}_{>0}} A_j$. If $g = \|\boldsymbol{f}\|_{\mathbb{R}^n}\chi_A$ then $g(x) \leq \|\boldsymbol{f}\|_{\mathbb{R}^n}(x)$ for every

$x \in X$ and so $g \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}_{\geq 0})$ by Proposition 2.7.47. If we define $B_k = \cup_{j=1}^k A_j$ and $f_k = f\chi_k$, $k \in \mathbb{Z}_{>0}$, then

$$\lim_{k\to\infty} f_k(x) = f(x)\chi_A(x), \qquad x \in X.$$

Therefore, by the Dominated Convergence Theorem, Theorem 2.7.56,

$$(f \cdot \mu)(A) = \int_X f\chi_A \,\mathrm{d}\mu = \lim_{k\to\infty} \int_X f_k \,\mathrm{d}\mu = \lim_{k\to\infty} \sum_{j=1}^k \int_X f\chi_{A_j} \,\mathrm{d}\mu = \sum_{j=1}^\infty (f \cdot \mu)(A_j),$$

giving countable-additivity of $f \cdot \mu$. ■

For the measures determined by integrable functions, as in Proposition 2.7.65, it is possible to explicitly characterise the integrals with respect to these measures. The notation from the previous proposition will be used in the statement of the next.

2.7.66 Proposition (Integration with respect to measures from functions) *If* $(\mathsf{X}, \mathscr{A}, \mu)$ *is a measure space and if* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \overline{\mathbb{R}})$, $\mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{C})$, *and* $\mathbf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R}^\mathsf{n})$, *then the following statements hold:*

(i) if $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R})$ *then* $\mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mathsf{f} \cdot \mu); \mathbb{R})$ *if and only if* $\mathsf{fg} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R})$, *and if either of these equivalent conditions holds,*

$$\int_\mathsf{X} \mathsf{g} \,\mathrm{d}(\mathsf{f} \cdot \mu) = \int_\mathsf{X} (\mathsf{fg}) \,\mathrm{d}\mu;$$

(ii) if $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{C})$ *then* $\mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mathsf{f} \cdot \mu); \mathbb{R})$ *if and only if* $\mathsf{fg} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{C})$, *and if either of these equivalent conditions holds,*

$$\int_\mathsf{X} \mathsf{g} \,\mathrm{d}(\mathsf{f} \cdot \mu) = \int_\mathsf{X} (\mathsf{fg}) \,\mathrm{d}\mu;$$

(iii) if $\mathbf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R}^\mathsf{n})$ *then* $\mathsf{g} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mathbf{f} \cdot \mu); \mathbb{R})$ *if and only if* $\mathsf{g}\mathbf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R}^\mathsf{n})$, *and if either of these equivalent conditions holds,*

$$\int_\mathsf{X} \mathsf{g} \,\mathrm{d}(\mathbf{f} \cdot \mu) = \int_\mathsf{X} (\mathsf{g}\mathbf{f}) \,\mathrm{d}\mu.$$

Proof Let us first consider the case where $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{R})$. Let us define

$$P = \{x \in X \mid f(x) \geq 0\}, \quad N = X \setminus P,$$

noting that P (and so N) is measurable by Proposition 2.6.16. Clearly (P, N) is a Hahn decomposition for $(X, \mathscr{A}, f \cdot \mu)$. Moreover, the corresponding Jordan decomposition is

$$f \cdot \mu = f_+ \cdot \mu - f_- \cdot \mu,$$

where, as usual, $f_+(x) = \max\{f(x), 0\}$ and $f_-(x) = \max\{-f(x), 0\}$. Noting that gf is integrable if and only if both gf_+ and gf_- are integrable, and computing

$$\int_X (fg)\,\mathrm{d}\mu = \int_X (f_+g)\,\mathrm{d}\mu - \int_X (f_-g)\,\mathrm{d}\mu = \int_X g\,\mathrm{d}(f_+\cdot\mu) - \int_X g\,\mathrm{d}(f_-\cdot\mu) = \int_X g\,\mathrm{d}(f\cdot\mu),$$

the result for signed measures follows.

To complete the proof, we suppose that $\boldsymbol{f} \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$, and prove the last assertion in the statement of the proposition. The proof of the second assertion is a consequence of this. For $A \in \mathscr{A}$ and $l \in \{1, \dots, n\}$ we have

$$(\boldsymbol{f}\cdot\mu)_l(A) = \mathrm{pr}_l\Big(\int_X \boldsymbol{f}\chi_A\,\mathrm{d}\mu\Big) = \int_X f_l\chi_A\,\mathrm{d}\mu = (f_l\cdot\mu)(A),$$

where $\mathrm{pr}_l\colon \mathbb{R}^n \to \mathbb{R}$ is the projection onto the lth component. Given this, and the definitions of the integral with respect to a vector measure and the integral of a vector-valued function, the result follows from the result proved above for $\mathbb{R}$-valued functions. ■

2.7.9 Notes

There is no standard convention on what Beppo Levi's Theorem is. Sometimes what we call the Monotone Convergence Theorem is called Beppo Levi's Theorem.

Exercises

2.7.1 Let $(X, \mathscr{A}, \mu)$ be a measure space and let $f, g \in \mathsf{S}(X; \overline{\mathbb{R}}_{\geq 0})$ satisfy $f(x) \leq g(x)$ for each $x \in X$. Show that

$$\int_X f\,\mathrm{d}\mu \leq \int_X g\,\mathrm{d}\mu.$$

2.7.2 Let $(X, \mathscr{A}, \mu)$ be a measure space and let $f \in \mathsf{S}(X; \overline{\mathbb{R}})$. For $A \in \mathscr{A}$ define $f_A\colon X \to \overline{\mathbb{R}}$ by $f_A = f\chi_A$. Show that

$$\int_X f_A\,\mathrm{d}\mu = \int_A (f|A)\,\mathrm{d}\mu_A.$$

2.7.3 Let $X = \mathbb{Z}_{>0}$, let $\mathscr{A} = \mathbf{2}^{\mathbb{Z}_{>0}}$, and let $\mu_\Sigma\colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ be the counting measure:

$$\mu_\Sigma(A) = \begin{cases} \mathrm{card}(A), & \mathrm{card}(A) < \infty, \\ \infty, & \text{otherwise.} \end{cases}$$

Verify the following statements using only the definition of the integral, i.e., do not use the general constructions of Examples 2.7.7 and 2.7.10.

(a) A function $f\colon \mathbb{Z}_{>0} \to \mathbb{R}$ is integrable if and only if the series $\sum_{j=1}^{\infty} f(j)$ is absolutely convergent.

(b) If f is integrable then

$$\int_{\mathbb{Z}_{>0}} f \, \mathrm{d}\mu_\Sigma = \sum_{j=1}^{\infty} f(j).$$

2.7.4 For a measure space $(X, \mathscr{A}, \mu)$ and for $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$, show that $f \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ if and only if $|f| \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$.

2.7.5 Let $X = \mathbb{Z}_{>0}$, $\mathscr{A} = \mathbf{2}^X$, and let μ_Σ be the counting measure on $\mathscr{A}$. Define $f \colon \mathbb{Z}_{>0} \to \mathbb{R}$ by $f(j) = j$. Use the Monotone Convergence Theorem to show that $f \notin \mathsf{L}^{(1)}((\mathbb{Z}_{>0}, \mathbf{2}^{\mathbb{Z}}, \mu_\Sigma); \mathbb{R})$.

The following exercise requires the notion of the concept of a norm which will be introduced in Section 3.1.

2.7.6 Let $(X, \mathscr{A}, \mu)$ be a measure space, let $\boldsymbol{f} \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu), \mathbb{R}^n)$, and let $\|\cdot\|$ be a norm on $\mathbb{R}^n$. Show that

$$\left\| \int_X \boldsymbol{f} \, \mathrm{d}\mu \right\| \leq \int_X \|\boldsymbol{f}\| \, \mathrm{d}\mu.$$

Hint: *Use Proposition 2.7.47 and Theorem 3.1.15.* figure this out

Section 2.8

Integration on products

In Section II-1.6.7 we presented Fubini's Theorem for the Riemann integral which showed how the n-dimensional Riemann integral could be computed by means of one-dimensional integrals. In Section 2.3.6 we introduced the product measure on a finite product of measure spaces. Understanding these two things, it is then naturally ask whether the integral for a product measure can be understood in terms of the measure of the component measure spaces. The result is the general version of Fubini's Theorem. As part of our treatment of Fubini's Theorem, we give an alternative characterisation of the product measure.

Do I need to read this section? We shall make frequent use of Fubini's Theorem. That being said, to make use of Fubini's Theorem it is not necessary to understand all of the details we present here. What is most important is to understand the hypotheses of Fubini's Theorem. •

2.8.1 The product measure by integration

In Section 2.3.6 we defined a unique measure on a product of measure spaces that had a natural property in terms of the measure of measurable rectangles. In this section we retrieve this measure in another way, using the integral. This construction has the benefit of being simpler than that in Section 2.3.6, but only after one has the integral at hand.

In Section 2.3.6 we defined product measures for arbitrary finite products. However, it is notationally easier to deal with a product with two factors, and then use induction to arrive at the general case. Thus we consider two measure spaces $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$. As in Section 2.2.3, a ***measurable rectangle*** is a subset $A \times B \subseteq X \times Y$ where $A \in \mathscr{A}$ and $B \in \mathscr{B}$. We denote by $\sigma(\mathscr{A} \times \mathscr{B})$ the σ-algebra generated by the collection of measurable rectangles. For a set $E \subseteq X \times Y$ and for $(x, y) \in X \times Y$ we define subsets $E_x \subseteq Y$ and $E^y \subseteq X$ by

$$E_x = \{y' \in Y \mid (x, y') \in E\}, \qquad E^y = \{x' \in X \mid (x', y) \in E\}.$$

One calls the sets E_x and E^y ***sections*** of the set E.

The following result begins our construction of the product measure using the integral. The reader will hopefully recognise something Fubini-like in this result.

2.8.1 Lemma (Integrals of sections) *For σ-finite measure spaces $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ and for $E \in \sigma(\mathscr{A} \times \mathscr{B})$, define*

$$\begin{aligned} \phi_E \colon X &\to \overline{\mathbb{R}} \\ x &\mapsto \nu(E_x), \end{aligned} \qquad \begin{aligned} \psi_E \colon Y &\to \overline{\mathbb{R}} \\ y &\mapsto \mu(E^y). \end{aligned}$$

Then ϕ_E and ψ_E are $\mathscr{A}$-measurable and $\mathscr{B}$-measurable, respectively. Moreover,

$$\int_X \phi_E \, d\mu = \int_Y \psi_E \, d\nu.$$

Proof Denote by $\mathscr{M}(X \times Y)$ the collection of all sets E for which the conclusions of the lemma hold. We shall show that $\mathscr{M}(X \times Y)$ is a monotone class containing the set of measurable rectangles.

For $A \in \mathscr{A}$ and $B \in \mathscr{B}$ we have

$$\phi_{A\times B}(x) = \nu(B)\chi_A(x), \quad \psi_{A\times B}(y) = \mu(A)\chi_B(y),$$

which shows that $A \times B \in \mathscr{M}(X \times Y)$. Therefore, $\phi_{A\times B}$ and $\psi_{A\times B}$ are measurable (by Example 2.6.8–2) and

$$\int_X \phi_{A\times B} \, d\mu = \int_Y \psi_{A\times B} \, d\nu = \mu(A)\nu(B).$$

Thus $\mathscr{M}(X \times Y)$ contains the measurable rectangles.

Now let $(E_j)_{j\in\mathbb{Z}_{>0}}$ be a collection of subsets of $\mathscr{M}(A\times B)$ for which $E_j \subseteq E_{j+1}$, $j \in \mathbb{Z}_{>0}$. Then, denoting $E = \cup_{j\in\mathbb{Z}_{>0}} E_j$,

$$\lim_{j\to\infty} \phi_{E_j}(x) = \phi_E(x), \quad \lim_{j\to\infty} \psi_{E_j}(y) = \psi_E(y).$$

Thus $\phi_E \in \mathsf{L}^{(0)}((X, \mathscr{A}), \overline{\mathbb{R}})$ and $\psi_E \in \mathsf{L}^{(0)}((Y, \mathscr{B}); \overline{\mathbb{R}})$ by Proposition 2.6.18. Note that the sequences $(\phi_{E_j}(x))_{j\in\mathbb{Z}_{>0}}$ and $(\psi_{E_j}(x))_{j\in\mathbb{Z}_{>0}}$ are monotonically increasing, so the Monotone Convergence Theorem gives

$$\int_X \phi_E \, d\mu = \lim_{j\to\infty} \int_X \phi_{E_j} \, d\mu = \lim_{j\to\infty} \int_Y \psi_{E_j} \, d\nu = \int_Y \psi_E \, d\nu.$$

Therefore, $E \in \mathscr{M}(X \times Y)$, which is part (i) of the definition of a monotone class.

Now, for the moment, suppose that $\mu(X)$ and $\nu(Y)$ are finite. Let $(E_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\sigma(\mathscr{A} \times \mathscr{B})$ such that $E_j \supseteq E_{j+1}$, $j \in \mathbb{Z}_{>0}$. Define $E = \cap_{j\in\mathbb{Z}_{>0}} E_j$ and note that

$$\lim_{j\to\infty} \phi_{E_j}(x) = \phi_E(x), \quad \lim_{j\to\infty} \psi_{E_j}(y) = \psi_E(y).$$

Thus $\phi_E \in \mathsf{L}^{(0)}((X, \mathscr{A}), \overline{\mathbb{R}})$ and $\psi_E \in \mathsf{L}^{(0)}((Y, \mathscr{B}); \overline{\mathbb{R}})$ by Proposition 2.6.18. Note that we obviously have

$$\phi_{E_j}(x) \le \nu(Y)\chi_X(x),\ \phi_E(x) \le \nu(Y)\chi_X(x), \qquad \psi_{E_j}(y) \le \mu(X)\chi_Y(y),\ \psi_E(y) \le \mu(X)\chi_Y(y)$$

for every $(x, y) \in X, Y$. Moreover, since we are assuming that X and Y have finite measure we have $\chi_X \in \mathsf{L}^{(1)}((X, \mathscr{A}, \mu); \mathbb{R})$ and $\chi_Y \in \mathsf{L}^{(1)}((Y, \mathscr{B}, \nu); \mathbb{R})$. Therefore, the hypotheses of the Dominated Convergence Theorem hold and we have

$$\int_X \phi_E \, d\mu = \lim_{j\to\infty} \int_X \phi_{E_j} \, d\mu = \lim_{j\to\infty} \int_Y \psi_{E_j} \, d\nu = \int_Y \psi_E \, d\nu,$$

from which we conclude that $E \in \mathscr{M}(X \times X)$. This verifies part (ii) of Definition 2.2.11 in this case. Thus this shows that, when $\mu(X), \nu(Y) < \infty$, $\mathscr{M}(X \times Y)$ is a monotone class containing the measurable rectangles. From Theorem 2.2.13 it then follows that $\sigma(\mathscr{A} \times \mathscr{B}) \subseteq \mathscr{M}(X \times Y)$. Thus the lemma holds in this case.

Now let us suppose that $\mu(X)$ and $\nu(Y)$ are not necessarily finite, but that using our assumption of σ-additivity we can write $X = \cup_{k \in \mathbb{Z}_{>0}} X_k$ and $Y = \cup_{k \in \mathbb{Z}_{>0}} Y_k$ where $\mu(X_k), \nu(Y_k) < \infty$, $k \in \mathbb{Z}_{>0}$, and where $(X_k)_{k \in \mathbb{Z}_{>0}}$ and $(Y_k)_{k \in \mathbb{Z}_{>0}}$ are pairwise disjoint measurable sets. Thus $X \times Y$ is the disjoint union of the measurable rectangles $X_{\times} Y_l$, $k, l \in \mathbb{Z}_{>0}$. Let $f \colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$ be a bijection and, for $m \in \mathbb{Z}_{>0}$, define $Z_m = X_k \times Y_l$ where $\phi(m) = (k, l)$. Now $X \times Y$ is a disjoint union of the measurable sets Z_m, $m \in \mathbb{Z}_{>0}$. Finally, define $S_n = \cup_{m=1}^{n} Z_m$ so that $X \times Y$ is a union of the measurable sets S_n, $n \in \mathbb{Z}_{>0}$, where $S_n \subseteq S_{n+1}$. Note that $\mu(S_n) < \infty$ for every $n \in \mathbb{Z}_{>0}$ since S_n is a finite union of sets of finite measure.

Now let $E \in \sigma(\mathscr{A} \times \mathscr{B})$ and denote $E_n = E \cap S_n$. From our argument above, $E_n \in \sigma(\mathscr{A} \times \mathscr{B})$ and

$$\int_X \phi_{E_n} \,\mathrm{d}\mu = \int_Y \psi_{E_n} \,\mathrm{d}\nu.$$

We also have

$$\lim_{n \to \infty} \phi_{E_n}(x) = \phi_E(x), \quad \lim_{n \to \infty} \psi_{E_n}(y) = \psi_E(y)$$

for every $(x, y) \in X \times Y$. Since $S_n \subseteq S_{n+1}$ for every $n \in \mathbb{Z}_{>0}$, the sequences $(\phi_{E_n}(x))_{n \in \mathbb{Z}_{>0}}$ and $(\psi_{E_n}(y))_{n \in \mathbb{Z}_{>0}}$ are increasing for every $(x, y) \in X \times Y$. Therefore, by the Monotone Convergence Theorem,

$$\int_X \phi_E \,\mathrm{d}\mu = \lim_{n \to \infty} \int_X \phi_{E_n} \,\mathrm{d}\mu = \lim_{n \to \infty} \int_Y \psi_{E_n} \,\mathrm{d}\nu = \int_Y \psi_E \,\mathrm{d}\nu,$$

giving the lemma. ■

With the preceding, we can fairly easily derive the product measure using the integral.

2.8.2 Theorem (The product measure using the integral) *For σ-finite measure spaces $(\mathrm{X}, \mathscr{A}, \mu)$ and $(\mathrm{Y}, \mathscr{B}, \nu)$, the map $\mu \times \nu \colon \sigma(\mathscr{A} \times \mathscr{B}) \to \overline{\mathbb{R}}_{\geq 0}$ defined by*

$$\mu \times \nu(\mathrm{E}) = \int_{\mathrm{X}} \phi_{\mathrm{E}} \,\mathrm{d}\mu = \int_{\mathrm{Y}} \psi_{\mathrm{E}} \,\mathrm{d}\nu$$

makes $(\mathrm{X} \times \mathrm{Y}, \sigma(\mathscr{A} \times \mathscr{B}), \mu \times \nu)$ a σ-finite measure space. Moreover, the measure $\mu \times \nu$ is the product measure as defined in Definition 2.3.34.

Proof It is clear that $\mu \times \lambda(\varnothing) = 0$ since $\varnothing = \varnothing \times \varnothing$ is a measurable rectangle, being the product of two sets with zero measure. For a sequence $(E_j)_{j \in \mathbb{Z}_{>0}}$ of disjoint subsets of $\sigma(\mathscr{A} \times \mathscr{B})$ define $E = \cup_{j \in \mathbb{Z}_{>0}} E_j$. Note that

$$\phi_E(x) = \sum_{j=1}^{\infty} \phi_{E_j}(x),$$

and so Beppo Levi's Theorem gives

$$\mu \times \nu(E) = \int_X \phi_E \,\mathrm{d}\mu = \sum_{j=1}^{\infty} \int_X \phi_{E_j} \,\mathrm{d}\mu = \sum_{j=1}^{\infty} \mu \times \nu(E_j),$$

as desired.

That $\mu \times \nu$ is the product measure follows from Theorem 2.3.33, along with the fact that we showed in the proof of Lemma 2.8.1 that $\mu \times \nu(A \times B) = \mu(A)\nu(B)$ for $A \in \mathscr{A}$ and $B \in \mathscr{B}$. ■

Now that we have established the product measure using the integral for a product with two factors, it is more or less a straightforward induction to do the same for products with three or more factors. Indeed, suppose we have σ-finite measure spaces $(X_j, \mathscr{A}_j, \mu_j)$, $j \in \{1, \dots, k\}$. For $E \subseteq X_1 \times \cdots \times X_k$ and for $x_k \in X_k$, denote

$$E_{x_k} = \{(x_1, \dots, x_{k-1}) \in X_1 \times \cdots \times X_{k-1} \mid (x_1, \dots, x_{k-1}, x_k) \in E\}.$$

Suppose that we have defined the product measure $\mu_1 \times \cdots \times \mu_{k-1}$ on $X_1 \times \cdots \times X_{k-1}$. Then define $\phi_E \colon X_k \to \overline{\mathbb{R}}$ by

$$\phi_E(x_k) = \mu_1 \times \cdots \times \mu_{k-1}(E_{x_k}).$$

We then have

$$\mu_1 \times \cdots \times \mu_k(E) = \int_{X_k} \phi_{E_k} \,\mathrm{d}\mu_k,$$

which is the product measure.

2.8.2 The integral on product spaces

Either by the construction of the previous section, or by the construction of Section 2.3.6, we have defined on the product $X_1 \times \cdots \times X_k$, for measure spaces $(X_j, \mathscr{A}_j, \mu_j)$, $j \in \{1, \dots, k\}$, a natural measure. One can then apply the construction of the integral from Section 2.7 to define the integral of measurable functions on the product. There is a slight hitch here that one needs to account for if one is to use this theory for the n-dimensional Lebesgue integral. To wit, in Section 2.5.4 we observed that the n-dimensional Lebesgue measure is not the product of the 1-dimensional Lebesgue measures on $\mathbb{R} \times \cdots \times \mathbb{R}$, but is the completion of this measure. Thus we should develop integration for, not just the product measure, but its completion. This is not particularly difficult, but just requires a few additional words.

As in the preceding section, for simplicity we start with two measure spaces $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$. As in the preceding section, we denote by $\sigma(\mathscr{A} \times \mathscr{B})$ the natural product σ-algebra on $X \times Y$, i.e., the σ-algebra generated by the measurable rectangles. By $\mu \times \nu$ we denote the product measure. As we saw in Section 2.5.4 (and more generally in Remark 2.3.36), there are cases where the measure $\mu \times \nu$ is not complete (although there are also cases where the product measure *is* complete). Thus we denote by $(X\times Y, \overline{\sigma}(\mathscr{A}\times\mathscr{B}), \overline{\mu\times\nu})$ the completion of $(X\times Y, \sigma(\mathscr{A}\times\mathscr{B}), \mu\times\nu)$.

In the previous section we defined the notion of the sections for a subset $E \subseteq X\times Y$. This can also be done for functions. For a function $f\colon X\times Y \to \overline{\mathbb{R}}$, we define functions $f_x\colon Y \to \overline{\mathbb{R}}$ and $f^y\colon X \to \overline{\mathbb{R}}$ by

$$f_x(y) = f^y(x) = f(x, y).$$

One calls the functions f_x and f^y ***sections*** of the function f. The following result give the measurability properties of sections of sets and functions.

2.8.3 Lemma (Measurability of sections) *For measure spaces $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$, the following statements hold:*

(i) if $E \subseteq X\times Y$ is $\sigma(\mathscr{A}\times\mathscr{B})$-measurable, then $E_x \in \mathscr{B}$ for every $x \in X$ and $E^y \in \mathscr{A}$ for every $y \in Y$;

(ii) if $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ are complete and if $E \subseteq X\times Y$ is $\overline{\sigma}(\mathscr{A}\times\mathscr{B})$-measurable, then $E_x \in \mathscr{B}$ for every $x \in X$ and $E^y \in \mathscr{A}$ for every $y \in Y$;

(iii) if $f\colon X\times Y \to \overline{\mathbb{R}}$ is $\sigma(\mathscr{A}\times\mathscr{B})$-measurable, then $f_x \in \mathsf{L}^{(0)}((Y, \mathscr{B}); \overline{\mathbb{R}})$ for almost every $x \in X$ and $f^y \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ for almost every $y \in Y$;

(iv) if $(X, \mathscr{A}, \mu)$ and $(Y, \mathscr{B}, \nu)$ are complete and if $f\colon X\times Y \to \overline{\mathbb{R}}$ is $\overline{\sigma}(\mathscr{A}\times\mathscr{B})$-measurable, then $f_x \in \mathsf{L}^{(0)}((Y sB); \overline{\mathbb{R}})$ and $f^y \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$.

Proof (i) This is a special case of Proposition 2.2.18.

(ii) Next suppose that $E \in \overline{\sigma}(\mathscr{A}\times\mathscr{B})$. We let $U \subseteq E \subseteq L$ have the property that $U, L \in \sigma(\mathscr{A}\times\mathscr{B})$ and $\mu\times\nu(U \setminus L) = 0$. We may apply the first part of the proof to U to assert that U_x and L_x are measurable for all $x \in X$. Since $(U_x \setminus E_x) \subseteq (U_x \setminus L_x)$ and since $U_x \setminus L_x$ has measure zero, it follows that $U_x \setminus E_x$ is measurable by completeness of $\mathscr{A}$. Thus E_x is $\overline{\sigma}(\mathscr{A}\times\mathscr{B})$-measurable. Similarly, E^y is also $\overline{\sigma}(\mathscr{A}\times\mathscr{B})$-measurable.

(iii) Note that for $S \subseteq \mathbb{R}$ we have $f_x^{-1}(S) = (f^{-1}(S))_x$ and $(f^y)^{-1}(S) = (f^{-1}(S))^y$. This part of the lemma now follows from part (i).

(iv) By Proposition 2.7.15 we may find g that is $\sigma(\mathscr{A}\times\mathscr{B})$-measurable and for which $f(x, y) = g(x, y)$ except on a set that has zero measure relative to $\mu\times\nu$. Thus $h = f - g$ is zero except on a set that has zero measure relative to $\mu\times\nu$. This part of the lemma will follow from part (iii) if we can show that h_x and h^y are measurable for almost every $x \in X$ and $y \in Y$. If E is the set of points in $X\times Y$ where h does not vanish then $E \in \overline{\sigma}(\mathscr{A}\times\mathscr{B})$. Thus we may find $E \subseteq U$ with $U \in \sigma(\mathscr{A}\times\mathscr{B})$ with $(\mu\times\nu)(U) = 0$. By Lemma 2.8.1 we have

$$\int_X \phi_U \,\mathrm{d}\mu = 0.$$

Now let $Z = \{x \in X \mid \phi_U(x) \neq 0\}$. We must have $\mu(Z) = 0$. Thus, for almost every $x \in X$ we have $\mu(U_x) = 0$. Since $E_x \subseteq U_x$ and since μ is complete, it follows that E_x is $\mathscr{B}$-measurable for almost every $x \in X$. If $y \notin E_x$ then we must have $h_x(y) = 0$. This implies that, as long as $x \notin Z$ then h_x is measurable and zero almost everywhere. This completes the proof. ■

2.8.3 Fubini's Theorem

Now let us investigate swapping the order of integration in computing integrals on products. Let us see what we might mean by this. If $f\colon X\times Y\to\overline{\mathbb{R}}$ is $\sigma(\mathscr{A}\times\mathscr{B})$-measurable or $\overline{\sigma}(\mathscr{A}\times\mathscr{B})$-measurable, then we define

$$\phi_f(x)=\begin{cases}\int_Y f_x\,\mathrm{d}\nu, & \text{the integral exists,}\\ 0, & \text{otherwise,}\end{cases}\qquad \psi_f(y)=\begin{cases}\int_X f^y\,\mathrm{d}\mu, & \text{the integral exists,}\\ 0, & \text{otherwise.}\end{cases}$$

We may then ask when it holds that

$$\int_X \phi_f\,\mathrm{d}\mu=\int_Y \psi_f\,\mathrm{d}\nu,$$

and when, if the preceding equality holds, both sides are, in fact, the integral of f with respect to the product measure. We have two more or less identical theorems, one for the product measure and one for its completion.

The first theorem deals with the product measure on $A\times B$.

2.8.4 Theorem (Fubini's Theorem for the product measure) *Let $(\mathsf{X},\mathscr{A},\mu)$ and $(\mathsf{Y},\mathscr{B},\nu)$ be σ-finite measure spaces and let $\mathrm{f}\colon\mathsf{A}\times\mathsf{B}\to\overline{\mathbb{R}}$ be $\sigma(\mathscr{A}\times\mathscr{B})$-measurable. Then the following statements hold:*

(i) if f is $\overline{\mathbb{R}}_{\geq 0}$-valued then ϕ_{f} and ψ_{f} are measurable and

$$\int_{\mathsf{X}}\phi_{\mathrm{f}}\,\mathrm{d}\mu=\int_{\mathsf{Y}}\psi_{\mathrm{f}}\,\mathrm{d}\nu=\int_{\mathsf{X}\times\mathsf{Y}}\mathrm{f}\,\mathrm{d}(\mu\times\nu);$$

(ii) if $\phi_{|\mathrm{f}|}\in\mathsf{L}^{(1)}((\mathsf{X},\mathscr{A},\mu);\overline{\mathbb{R}})$ or if $\psi_{|\mathrm{f}|}\in\mathsf{L}^{(1)}((\mathsf{Y},\mathscr{B},\nu),\overline{\mathbb{R}})$, then

$$\mathrm{f}\in\mathsf{L}^{(1)}((\mathsf{X}\times\mathsf{Y},\sigma(\mathscr{A}\times\mathscr{B}),\mu\times\nu);\overline{\mathbb{R}})$$

and

$$\int_{\mathsf{X}}\phi_{\mathrm{f}}\,\mathrm{d}\mu=\int_{\mathsf{Y}}\psi_{\mathrm{f}}\,\mathrm{d}\nu=\int_{\mathsf{X}\times\mathsf{Y}}\mathrm{f}\,\mathrm{d}(\mu\times\nu);$$

(iii) if $\mathrm{f}\in\mathsf{L}^{(1)}((\mathsf{X}\times\mathsf{Y},\sigma(\mathscr{A}\times\mathscr{B}),\mu\times\nu);\overline{\mathbb{R}})$ then

(a) $\mathrm{f}_{\mathrm{x}}\in\mathsf{L}^{(1)}((\mathsf{Y},\mathscr{B},\nu);\overline{\mathbb{R}})$ and $\mathrm{f}^{\mathrm{y}}\in\mathsf{L}^{(1)}((\mathsf{X},\mathscr{A},\mu);\overline{\mathbb{R}})$ for almost every $\mathrm{x}\in\mathsf{X}$ and $\mathrm{y}\in\mathsf{Y}$,

(b) $\phi_{\mathrm{f}}\in\mathsf{L}^{(1)}((\mathsf{X},\mathscr{A},\mu);\overline{\mathbb{R}})$ and $\psi_{\mathrm{f}}\in\mathsf{L}^{(1)}((\mathsf{Y},\mathscr{B},\nu);\overline{\mathbb{R}})$, and

(c) it holds that

$$\int_{\mathsf{X}}\phi_{\mathrm{f}}\,\mathrm{d}\mu=\int_{\mathsf{Y}}\psi_{\mathrm{f}}\,\mathrm{d}\nu=\int_{\mathsf{X}\times\mathsf{Y}}\mathrm{f}\,\mathrm{d}(\mu\times\nu).$$

Proof (i) By Lemma 2.8.3 the functions ϕ_f and ψ_f are everywhere defined since the integral of a nonnegative-valued measurable function always exists. By Lemma 2.8.1 this part of the theorem holds for characteristic functions of $\mathscr{L}(A)\times\mathscr{L}(B)$-measurable sets. Therefore, it also holds for simple functions by Proposition 2.7.16 since simple functions are finite linear combinations of characteristic functions. By Proposition 2.6.39 let $(g_j)_{j\in\mathbb{Z}_{>0}}$ be a monotonically increasing sequence of simple functions such that $f(x,y)=\lim_{j\to\infty} g_j(x,y)$ for each $(x,y)\in X\times Y$. By the Monotone Convergence Theorem we have

$$\int_X \phi_f \,\mathrm{d}\mu = \lim_{j\to\infty}\int_X \phi_{g_j}\,\mathrm{d}\mu = \lim_{j\to\infty}\int_{X\times Y} g_j\,\mathrm{d}(\mu\times\nu) = \int_{X\times Y} f\,\mathrm{d}(\mu\times\nu),$$

and similarly for ψ_f. This gives the result.

(ii) By part (i) we have

$$\int_X \phi_{|f|}\,\mathrm{d}\mu = \int_X \psi_{|f|}\,\mathrm{d}\nu = \int_{X\times Y}|f|\,\mathrm{d}(\mu\times\nu) < \infty.$$

Thus f is $\mu\times\nu$-integrable, as desired. Note, then, that f_+, f_- are $\mu\times\nu$-integrable. Thus $f\in \mathsf{L}^{(1)}((X\times Y),\mathscr{A}\times\mathscr{B},\mu\times\nu);\overline{\mathbb{R}})$ by Exercise 2.7.4. By part (i) we have

$$\int_X \phi_{f_+}\,\mathrm{d}\mu = \int_X \psi_{f_+}\,\mathrm{d}\nu = \int_{X\times Y} f_+\,\mathrm{d}(\mu\times\nu),$$

and similarly for f_-. By Proposition 2.7.17 it then follows that

$$\int_X \phi_f\,\mathrm{d}\mu = \int_X \psi_f\,\mathrm{d}\nu = \int_{X\times Y} f\,\mathrm{d}(\mu\times\nu),$$

as desired.

(iii) Write $f = f_+ - f_-$ and note that f_x, $f_{+,x}$, and $f_{-,x}$ are $\mathscr{B}$-measurable by Lemma 2.8.3. By part (i) the functions ϕ_{f_+} and ϕ_{f_-} are $\mathscr{A}$-measurable. Also by part (i) we have

$$\int_X \phi_{f_+}\,\mathrm{d}\mu = \int_X \psi_{f_+}\,\mathrm{d}\nu = \int_{X\times Y} f_+\,\mathrm{d}(\mu\times\nu),$$

and similarly for f_-. Therefore, ϕ_{f_+} and ϕ_{f_-} are integrable with respect to μ. Therefore, ϕ_{f_+} and ϕ_{f_-} are finite for almost all $x\in X$ by Proposition 2.7.12. If

$$Z = \{x\in X \mid \phi_{f_+}(x)=\infty\}\cup\{x\in X \mid \phi_{f_-}(x)=\infty\}$$

then $Z\in\mathscr{A}$ by Proposition 2.6.6 and $\mu(Z)=0$. If $x\notin Z$ then we have

$$\phi_f(x) = \int_X f_+\,\mathrm{d}\mu - \int_X f_-\,\mathrm{d}\mu = \phi_{f_+}(x) - \phi_{f_-}(x)$$

and if $x\in Z$ we have $\phi_f(x)=0$. Thus ϕ_f almost everywhere agrees with $\phi_{f_+}-\phi_{f_-}$. By Propositions 2.7.11 and 2.7.17 we have

$$\begin{aligned}\int_X \phi_f\,\mathrm{d}\mu &= \int_X \phi_{f_+}\,\mathrm{d}\mu - \int_X \phi_{f_-}\,\mathrm{d}\mu\\ &= \int_{X\times Y} f_+\,\mathrm{d}(\mu\times\nu) - \int_{X\times Y} f_-\,\mathrm{d}(\mu\times\nu)\\ &= \int_{X\times Y} f\,\mathrm{d}(\mu\times\nu),\end{aligned}$$

as desired. A similar argument gives

$$\int_Y \psi_f \, d\nu = \int_{X\times Y} f \, d(\mu \times \nu)$$

which completes the proof. ■

We shall also use the following result, which follows from the previous theorem, along with the definition of the integral for vector-valued functions. In the statement of the theorem, we use the obvious definitions for f_x and f^y for a function $f\colon X \times Y \to \mathbb{R}^n$ and for functions $\phi_f\colon X \to \mathbb{R}^n$ and $\psi_f\colon Y \to \mathbb{R}^n$.

2.8.5 Corollary (Vector-valued Fubini's Theorem for the product measure) *Let* $(X, \mathscr{A}, \mu)$ *and* $(Y, \mathscr{B}, \nu)$ *be σ-finite measure spaces and let* $\mathbf{f}\colon A \times B \to \mathbb{R}^n$ *be* $\sigma(\mathscr{A} \times \mathscr{B})$*-measurable. Then the following statements hold:*

(i) if $\phi_{\|\mathbf{f}\|_{\mathbb{R}^n}} \in L^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$ *or if* $\psi_{\|\mathbf{f}\|_{\mathbb{R}^n}} \in L^{(1)}((Y, \mathscr{B}, \nu), \mathbb{R}^n)$*, then*

$$\mathbf{f} \in L^{(1)}((X \times Y, \sigma(\mathscr{A} \times \mathscr{B}), \mu \times \nu); \mathbb{R}^n)$$

and

$$\int_X \phi_{\mathbf{f}} \, d\mu = \int_Y \psi_{\mathbf{f}} \, d\nu = \int_{X\times Y} \mathbf{f} \, d(\mu \times \nu);$$

(ii) if $\mathbf{f} \in L^{(1)}((X \times Y, \sigma(\mathscr{A} \times \mathscr{B}), \mu \times \nu); \mathbb{R}^n)$ *then*

(a) $\mathbf{f}_x \in L^{(1)}((Y, \mathscr{B}, \nu); \mathbb{R}^n)$ *and* $\mathbf{f}^y \in L^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$ *for almost every* $x \in X$ *and* $y \in Y$*,*

(b) $\phi_{\mathbf{f}} \in L^{(1)}((X, \mathscr{A}, \mu); \mathbb{R}^n)$ *and* $\psi_{\mathbf{f}} \in L^{(1)}((Y, \mathscr{B}, \nu); \mathbb{R}^n)$*, and*

(c) it holds that

$$\int_X \phi_{\mathbf{f}} \, d\mu = \int_Y \psi_{\mathbf{f}} \, d\nu = \int_{X\times Y} \mathbf{f} \, d(\mu \times \nu).$$

Of course, the theorem applies to the space case of $\mathbb{R}^2$ and so to $\mathbb{C}$-valued functions.

Let us give some examples that illustrate how to use Fubini's Theorem, as well as some of the caveats one must be aware of when applying the theorem.

2.8.6 Examples (Fubini's Theorem)

1. Let us take $X = Y = \mathbb{Z}_{>0}$, $\mathscr{A} = \mathscr{B} = \mathbf{2}^{\mathbb{Z}_{>0}}$, and $\mu = \nu = \mu_\Sigma$, where we recall from Example 2.3.9–3 that μ_Σ denotes the counting measure. For $f\colon \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{R}$ and $m \in \mathbb{Z}_{>0}$ define $f_m\colon \mathbb{Z}_{>0} \times \mathbb{Z}_{>0} \to \mathbb{R}$ by

$$f_m(j,k) = \begin{cases} f(j,k), & j,k \in \{1,\dots,m\}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that

$$\int_{\mathbb{Z}_{>0}\times\mathbb{Z}_{>0}} f_m \, d(\mu_\Sigma \times \mu_\Sigma) = \sum_{j=1}^{m} \sum_{k=1}^{m} f(j,k)$$

since f_m is a simple function. Clearly, $f(j,k) = \lim_{m\to\infty} f_m(j,k)$ for every $(j,k) \in \mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$. Thus, by the Monotone Convergence Theorem,

$$\int_{\mathbb{Z}_{>0}\times\mathbb{Z}_{>0}} |f| \,\mathrm{d}\mu_\Sigma \times \mu_\Sigma = \lim_{m\to\infty} \sum_{j=1}^{m} \sum_{j=1}^{m} |f(j,k)|.$$

In other words, $f \in \mathsf{L}^{(1)}((\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}), \mathbf{2}^{\mathbb{Z}_{>0}} \times \mathbf{2}^{\mathbb{Z}_{>0}}, \mu_\Sigma \times \mu_\Sigma); \mathbb{R})$ if and only if

$$\sum_{j,k=1}^{\infty} |f(j,k)| < \infty,$$

noting that the doubly infinite sum is unambiguously defined since it is a sum of positive terms, cf. Theorem I-2.4.5.

Now, Fubini's Theorem in this case tells us that when $f \in \mathsf{L}^{(1)}((\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}, \mathbf{2}^{\mathbb{Z}_{>0}} \times \mathbf{2}^{\mathbb{Z}_{>0}}, \mu_\Sigma \times \mu_\Sigma); \mathbb{R})$ then

$$\sum_{j=1}^{\infty} \sum_{k=1}^{\infty} f(j,k) = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} f(j,k) = \sum_{j,k=1}^{\infty} f(j,k),$$

i.e., the order of summation can be swapped.

2. We take $X = Y = \mathbb{Z}$, $\mathscr{A} = \mathscr{B} = \mathbf{2}^{\mathbb{Z}}$, and $\mu = \nu = \mu_\Sigma$. We define $f\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ by

$$f(j,k) = \begin{cases} 1, & j \in \mathbb{Z}_{\geq 0},\ k = j, \\ -1, & j \in \mathbb{Z}_{\geq 0},\ k = j+1, \\ 0, & \text{otherwise.} \end{cases}$$

We directly compute

$$\phi_f(j) = 0, \qquad \psi_f(k) = \begin{cases} 1, & k = 0, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$\int_{\mathbb{Z}} \phi_f \,\mathrm{d}\mu_\Sigma = 0, \qquad \int_{\mathbb{Z}} \psi_f \,\mathrm{d}\mu_\Sigma = 1,$$

which shows that the order of integration (order of summation, in this case) cannot be swapped. This does not contradict Theorem 2.8.4, however. Indeed, note that

$$\phi_{|f|}(j) = \begin{cases} 2, & j \in \mathbb{Z}_{\geq 0}, \\ 0, & \text{otherwise,} \end{cases} \qquad \psi_{|f|} = \begin{cases} 1, & j = 0, \\ 2, & j \in \mathbb{Z}_{>0}, \\ 0, & \text{otherwise.} \end{cases}$$

Since neither of these functions is integrable, part (ii) of Theorem 2.8.4 cannot be applied.

3. One might wonder whether the fact that the measure spaces are infinite in the preceding example is the reason for the failure of Fubini's Theorem. In this example, we shall show that this is not the case. Here we shall use the Lebesgue integral, which is defined using the Lebesgue measure. Although we do not discuss this in detail until Sections 2.9 and 2.10, this should not cause problems since for this example it suffices to consider the functions as being Riemann integrable.
We take $X = Y = [0,1]$, $\mathscr{A} = \mathscr{B} = \mathscr{L}([0,1])$, and $\mu = \nu = \lambda_{[0,1]}$. Define $\xi_j = 1-\frac{1}{j+1}$, $j \in \mathbb{Z}_{>0}$, and let $g_j\colon [0,1] \to \mathbb{R}$ be a positive continuous function such that $\int_{[0,1]} g_j \, \mathrm{d}\lambda_{[0,1]} = 1$ and such that $\operatorname{supp}(g_j) \subseteq (\xi_j, \xi_{j+1})$ (for example, a "triangular" function of the right height and base). Then define $f\colon [0,1] \times [0,1] \to \mathbb{R}$ by

$$f(x,y) = \sum_{j=1}^{\infty} (g_j(x) - g_{j+1}(x)) g_j(y).$$

It is clear that for each $(x,y) \in [0,1] \times [0,1]$ this sum has at most one nonzero term, and so is well-defined. By construction, we have

$$\phi_f(x) = \sum_{j=1}^{\infty} (g_j(x) - g_{j+1}(x)) \int_{[0,1]} g_j \, \mathrm{d}\lambda_{[0,1]} = \sum_{j=1}^{\infty} (g_j(x) - g_{j+1}(x))$$

and

$$\psi_f(y) = \sum_{j=1}^{\infty} g_j(y) \int_{[0,1]} (g_j - g_{j+1}) \, \mathrm{d}\lambda_{[0,1]} = 0.$$

Therefore, observing that

$$\sum_{j=1}^{\infty} (g_j(x) - g_{j+1}(x)) = g_1(x),$$

we have

$$\int_{[0,1]} \phi_f \, \mathrm{d}\lambda_{[0,1]} = 1, \quad \int_{[0,1]} \psi_f \, \mathrm{d}\lambda_{[0,1]} = 0,$$

showing that the order of integration cannot be swapped. But this does not contradict part (ii) of Theorem 2.8.4 since

$$\phi_{|f|}(x) = \sum_{j=1}^{\infty} |g_j(x) - g_{j+1}(x)| \int_{[0,1]} g_j \, \mathrm{d}\lambda_{[0,1]} = \sum_{j=1}^{\infty} (g_j(x) + g_{j+1}(x))$$

$$\implies \quad \int_{[0,1]} \phi_{|f|} \, \mathrm{d}\lambda_{[0,1]} = \infty,$$

using the fact that the functions g_j, $j \in \mathbb{Z}_{>0}$, are positive and have pairwise disjoint support. Thus the hypotheses of part (ii) of Theorem 2.8.4 do not hold.

4. Let us consider now a case where Fubini's Theorem can fail for a positive-valued function. Again, we make use of the Lebesgue integral. We take $X = Y = [0,1]$, $\mathscr{A} = \mathbf{2}^{[0,1]a}$, $\mathscr{B} = \mathscr{L}([0,1])$, and $\mu = \mu_\Sigma$ and $\nu = \lambda_{[0,1]}$. We define $f\colon \mathbb{Z}\times\mathbb{R} \to \mathbb{R}$ by

$$f(x,y) = \begin{cases} 1, & x = y, \\ 0, & \text{otherwise.} \end{cases}$$

Then we compute

$$\phi_f(x) = 0, \quad \psi_f(y) = 1$$

for all $(x,y) \in [0,1]\times[0,1]$. Therefore,

$$\int_{[0,1]} \phi_f \, \mathrm{d}\mu_\Sigma = 0, \quad \int_{[0,1]} \psi_f \, \mathrm{d}\lambda_{[0,1]} = 1.$$

Again, the order of integration cannot be swapped. In this case, the issue cannot be with the hypotheses of part (ii) of Theorem 2.8.4 since f is nonnegative-valued, and so it is part (i) that should be applied. However, the problem with this example is that the measure space $([0,1], \mathbf{2}^{[0,1]}, \mu_\Sigma)$ is not σ-finite.

finish

5. •

Next we state the version of Fubini's Theorem for the completion of the product measure. This is actually the version of Fubini's Theorem that gets the most use since it applies to the Lebesgue integral on $\mathbb{R}^n$ as a product measure. Fortunately, it differs from Theorem 2.8.4 only in the use of the completed measure in the statement.

2.8.7 Theorem (Fubini's Theorem for the completion of the product measure) *Let* $(\mathsf{X},\mathscr{A},\mu)$ *and* $(\mathsf{Y},\mathscr{B},\nu)$ *be σ-finite measure spaces and let* $\mathsf{f}\colon \mathsf{A}\times\mathsf{B}\to\overline{\mathbb{R}}$ *be* $\overline{\sigma}(\mathscr{A}\times\mathscr{B})$*-measurable. Then the following statements hold:*

(i) if f *is* $\overline{\mathbb{R}}_{\ge 0}$*-valued then* ϕ_f *and* ψ_f *are measurable and*

$$\int_\mathsf{X} \phi_\mathsf{f}\,\mathrm{d}\mu = \int_\mathsf{Y} \psi_\mathsf{f}\,\mathrm{d}\nu = \int_{\mathsf{X}\times\mathsf{Y}} \mathsf{f}\,\mathrm{d}(\overline{\mu\times\nu});$$

(ii) if $\phi_\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X},\mathscr{A},\mu);\overline{\mathbb{R}})$ *or if* $\psi_\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{Y},\mathscr{B},\nu),\overline{\mathbb{R}})$*, then*

$$\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}\times\mathsf{Y},\overline{\sigma}(\mathscr{A}\times\mathscr{B}),\overline{\mu\times\nu});\overline{\mathbb{R}})$$

and

$$\int_\mathsf{X} \phi_\mathsf{f}\,\mathrm{d}\mu = \int_\mathsf{Y} \psi_\mathsf{f}\,\mathrm{d}\nu = \int_{\mathsf{X}\times\mathsf{Y}} \mathsf{f}\,\mathrm{d}(\overline{\mu\times\nu});$$

(iii) if $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}\times\mathsf{Y},\overline{\sigma}(\mathscr{A}\times\mathscr{B}),\overline{\mu\times\nu});\overline{\mathbb{R}})$ *then*

(a) $f_x \in L^{(1)}((Y, \mathscr{B}, \nu); \overline{\mathbb{R}})$ *and* $f^y \in L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *for almost every* $x \in X$ *and* $y \in Y$,

(b) $\phi_f \in L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *and* $\psi_f \in L^{(1)}((Y, \mathscr{B}, \nu); \overline{\mathbb{R}})$*, and*

(c) *it holds that*

$$\int_X \phi_f \, d\mu = \int_Y \psi_f \, d\nu = \int_{X \times Y} f \, d(\overline{\mu \times \nu}).$$

Proof By Proposition 2.7.15 we can write $f = g + h$ where

$$\mu \times \nu(\{(x, y) \in X \times Y \mid h(x, y) \neq 0\}) = 0$$

and where g is $\sigma(\mathscr{A} \times \mathscr{B})$-measurable. One now applies Theorem 2.8.4 to g, notes that $f_x = g_x$ for almost every x by Lemma 2.8.3, and therefore deduces from Proposition 2.7.15 that

$$\int_{X \times Y} f \, d(\overline{\mu \times \nu}) = \int_X \phi_f \, d\mu = \int_X \phi_g \, d\mu = \int_{X \times Y} g \, d(\mu \times \nu),$$

provided that all integrals exist. A similar conclusion holds using f^y, g^y, ψ_f, and ψ_g. The theorem follows directly from this. ■

The next result deals with a situation we will commonly encounter when using Fubini's theorem.

2.8.8 Corollary (A special case of Fubini's Theorem) *Let* $(X, \mathscr{A}, \mu)$ *and* $(Y, \mathscr{B}, \nu)$ *be σ-finite measure spaces, let* $f \in L^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}})$ *and* $g \in L^{(0)}((Y, \mathscr{B}); \overline{\mathbb{R}})$*, and define* $F: X \times Y \to \overline{\mathbb{R}}$ *by* $F(x, y) = f(x)g(y)$*. Then*

(i) F *is both* $\sigma(\mathscr{A} \times \mathscr{B})$*- and* $\overline{\sigma}(\mathscr{A} \times \mathscr{B})$*-measurable and*

(ii) F *is integrable with respect to both* $\mu \times \nu$ *and* $\overline{\mu \times \nu}$ *if* $f \in L^{(1)}((X, \mathscr{A}, \mu); \overline{\mathbb{R}})$ *and* $g \in L^{(1)}((Y, \mathscr{B}, \nu); \overline{\mathbb{R}})$.

Proof (i) Denote $\tilde{f}, \tilde{g}: X \times Y \to \mathbb{R}$ by $\tilde{f}(x, y) = f(x)$ and $\tilde{g}(x, y) = g(y)$. Then

$$\tilde{f}^{-1}([a, \infty]) = f^{-1}([a, \infty]) \times Y \in \mathscr{A} \times \mathscr{B},$$

and so both $\tilde{f}$ is $\sigma(\mathscr{A} \times \mathscr{B})$- and $\overline{\sigma}(\mathscr{A} \times \mathscr{B})$-measurable. Similarly, $\tilde{g}$ is both $\sigma(\mathscr{A} \times \mathscr{B})$- and $\overline{\sigma}(\mathscr{A} \times \mathscr{B})$-measurable. Therefore, by Proposition 2.6.11, $\tilde{f}\tilde{g}$ is both $\sigma(\mathscr{L}(A) \times \mathscr{L}(B))$- and $\overline{\sigma}(\mathscr{A} \times \mathscr{B})$-measurable. This part of the result follows since $F = \tilde{f}\tilde{g}$.

(ii) By part (i) of Theorem 2.8.4 we compute

$$\int_{X \times Y} |F| \, d(\mu \times \nu) = \int_Y |g| \Big(\int_X |f| \, d\mu \Big) \, d\nu = \Big(\int_X |f| \, d\mu \Big) \Big(\int_Y |g| \, d\nu \Big) < \infty.$$

The result now follows from part (ii) of Theorem 2.8.4. ■

Move this to Lebesgue section?

2.8.9 Example (Fubini's Theorem for the Lebesgue measure) Let us consider $X = Y = \mathbb{R}$, $\mathscr{A} = \mathscr{B} = \mathscr{L}(\mathbb{R})$, and $\mu = \nu = \lambda$. Define $f\colon \mathbb{R}\times\mathbb{R}\to\mathbb{R}$ by

$$f(x,y) = \begin{cases} 1, & x\in\mathbb{R}_{\geq 0},\ y\in[x,x+1],\\ -1, & x\in\mathbb{R}_{>0},\ y\in[x+1,x+2],\\ 0, & \text{otherwise.}\end{cases}$$

In Sections 2.9 and 2.10 we shall show that the Lebesgue integral agrees with the Riemann integral in cases where the latter is defined. Therefore, to work out this example, it suffices to perform integration using the usual Riemann integral. Let us then denote the integral with respect to the first factor by $\int \mathrm{d}x$ and the integral with respect to the second factor by $\int \mathrm{d}y$. In Figure 2.5 we depict the function.

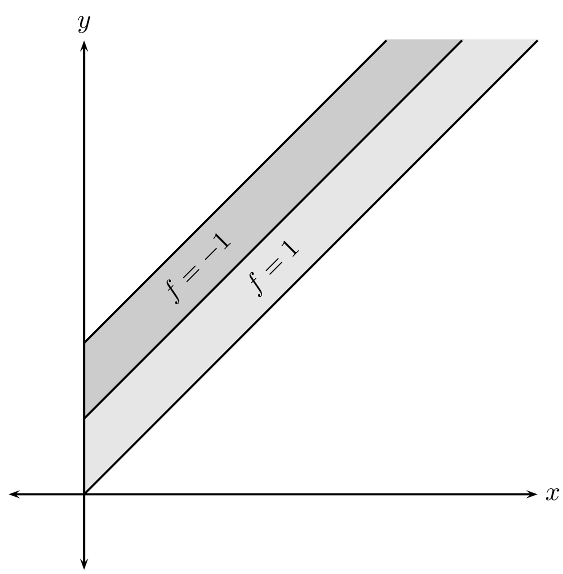


Figure 2.5 A function for which Fubini's Theorem does not hold

With this figure in mind we compute

$$\begin{aligned}\int_{\mathbb{R}}\Big(\int_{\mathbb{R}} f(x,y)\,\mathrm{d}x\Big)\,\mathrm{d}y &= \int_0^1\Big(\int_0^y \mathrm{d}x\Big)\,\mathrm{d}y + \int_1^2\Big(\int_{y-1}^y \mathrm{d}x - \int_0^{y-1}\Big)\,\mathrm{d}y\\ &\quad + \int_2^\infty\Big(\int_{y-1}^y \mathrm{d}x - \int_{y-2}^{y-1}\mathrm{d}x\Big)\,\mathrm{d}y\\ &= \frac{1}{2}+\frac{1}{2}+0 = 1\end{aligned}$$

and

$$\int_{\mathbb{R}}\Big(\int_{\mathbb{R}} f(x,y)\,\mathrm{d}y\Big)\,\mathrm{d}x = \int_0^\infty\Big(\int_x^{x+1}\mathrm{d}y - \int_{x+1}^{x+2}\Big)\,\mathrm{d}x = 0.$$

Thus both integrals

$$\int_{\mathbb{R}}\Big(\int_{\mathbb{R}} f(x,y)\,\mathrm{d}x\Big)\,\mathrm{d}y, \qquad \int_{\mathbb{R}}\Big(\int_{\mathbb{R}} f(x,y)\,\mathrm{d}y\Big)\,\mathrm{d}x$$

exist, but they are not equal to one another. However, this does not contradict Theorem 2.8.7. To see this, note that

$$\phi_{|f|}(x) = \begin{cases} 2, & x \in \mathbb{R}_{\geq 0}, \\ 0, & \text{otherwise}, \end{cases} \qquad \psi_{|f|}(y) = \begin{cases} y, & y \in [0,2], \\ y-2, & y \in (2,\infty), \\ 0, & \text{otherwise}. \end{cases}$$

Since neither of these functions is integrable, part (ii) of Theorem 2.8.7 does not apply. More directly, $|f|$ is the characteristic function of the union of the shaded regions in Figure 2.5. Therefore, the integral of $|f|$ is the area of this region which is infinity. Thus f is not integrable. •

Exercises

2.8.1

Section 2.9

The single-variable Lebesgue integral

The Lebesgue integral on $\mathbb{R}$ is nothing but the integral defined in Section 2.7.1 when the measure space is $(\mathbb{R}, \mathscr{L}(\mathbb{R}), \lambda)$. We shall not develop the *definition* of Lebesgue integral beyond this observation, so the reader looking to understand this definition will have to read Section 2.4 and then read Section 2.7 replacing all occurrences of $(X, \mathscr{A}, \mu)$ with $(\mathbb{R}, \mathscr{L}(\mathbb{R}), \lambda)$. This will give the reader most of what they will need to use the Lebesgue integral effectively. In this section we gather a few results and observations that are particuar to the Lebesgue integral on $\mathbb{R}$.

Do I need to read this section? The reader looking for the definition of the Lebesgue integral and some of its basic properties will get that by reading Sections 2.4 and 2.7 as described above. If this is all one is interested in, then this section can be bypassed, and the results consulted when needed. One topic in this section that may be of interest, and which is not contained in Sections 2.4 and 2.7, is the relationship between the Lebesgue integral and the Riemann integral. This, after all, is how we motivated the constructions that have gotten us to where we are. •

2.9.1 Lebesgue measurable functions

We begin by studying the character of Lebesgue measurable functions on $\mathbb{R}$. In this case, the additional structure of $\mathbb{R}$ allows us to give some further refinements of the properties of measurable functions.

Let us introduce the common terminology for the particular measurable functions we discuss in this section.

2.9.1 Definition (Borel measurable, Lebesgue measurable) Let $A \subseteq \mathbb{R}$. A function $f\colon A \to \overline{\mathbb{R}}$ is

(i) ***Borel measurable*** if $A \in \mathscr{B}(\mathbb{R})$ and if f is $\mathscr{B}(A)$-measurable and

(ii) ***Lebesgue measurable*** if $A \in \mathscr{L}(\mathbb{R})$ and if f is $\mathscr{L}(A)$-measurable.

We shall almost always write $\mathsf{L}^{(0)}(A; \overline{\mathbb{R}})$ for the Lebesgue measurable functions on A, rather than $\mathsf{L}^{(0)}((\mathbb{R}, \mathscr{L}(A)); \overline{\mathbb{R}})$. •

Now let us consider the approximation of measurable functions by "nice" functions like step functions and continuous functions. We recall from Section I-3.4.1 the notion of a step function defined on a compact interval.

2.9.2 Theorem (Lebesgue measurable functions are approximated by step functions) *If* $\mathrm{I} = [\mathrm{a}, \mathrm{b}]$ *is a compact interval, if* $\mathrm{f}\colon \mathrm{I} \to \overline{\mathbb{R}}$ *is measurable and satisfies*

$$\lambda(\{\mathrm{x} \in \mathrm{I} \mid \mathrm{f}(\mathrm{x}) \in \{-\infty, \infty\}\}) = 0,$$

and if $\epsilon_1, \epsilon_2 \in \mathbb{R}_{>0}$, *then there exists a step function* $g\colon I \to \mathbb{R}_{\geq 0}$ *such that*

$$\lambda(\{x \in I \mid |f(x) - g(x)| \geq \epsilon_1\}) < \epsilon_2.$$

Proof It suffices to prove the theorem when $\epsilon_1 = \epsilon_2 = \epsilon$. Thus we take $\epsilon \in \mathbb{R}_{>0}$.

For $k \in \mathbb{Z}_{>0}$ define

$$A_k = \{x \in I \mid |f(x)| \geq k\},$$

and note that the sequence $(\lambda(I \setminus A_k))_{k\in\mathbb{Z}_{>0}}$ is monotonically increasing and bounded above by $b - a$. Thus it is convergent by Theorem I-2.3.8. Moreover, it converges to $b - a$. Indeed, if the sequence converges to $\ell < b - a$ then this would imply, by Proposition 2.3.3, that

$$\lim_{k\to\infty} \lambda(I \setminus A_k) = \lambda(I \setminus \cup_{k\in\mathbb{Z}_{>0}} A_k) < b - a.$$

Thus there exists a set $B \subseteq I$ of positive measure such that $I = (\cup_{k\in\mathbb{Z}_{>0}A_k} \overset{\circ}{\cup} B)$. Note if $x \in B$ then $|f(x)| = \infty$, contradicting our assumptions on f. Thus we indeed have $\lim_{k\to\infty} \lambda(I \setminus A_k) = b - a$. Thus there exists $M \in \mathbb{Z}_{>0}$ such that $\lambda(I \setminus A_M) < b - a - \frac{\epsilon}{2}$, i.e., $\lambda(A_M) < \frac{\epsilon}{2}$. Therefore,

$$\lambda(\{x \in I \mid |f(x)| \geq M\}) < \tfrac{\epsilon}{2}.$$

Then define $f_M\colon I \to \mathbb{R}$ by

$$f_M(x) = \begin{cases} f(x), & |f(x)| < M, \\ M, & |f(x)| \geq M, \\ -M, & f(x) < -M. \end{cases}$$

Note that f_M is measurable by Proposition 2.6.16.

Now take $K \in \mathbb{Z}_{>0}$ such that $2^{-K} < \epsilon$ and such that $K \geq M$. If we follow the construction in the proof of Proposition 2.6.39 then we define

$$A_{+,K,j} = \{x \in I \mid 2^{-K}(j-1) \leq f_M(x) < 2^{-K}j\}$$

and

$$A_{-,K,j} = \{x \in I \mid -2^{-K}j \leq f_M(x) < -2^{-K}(j-1)\}$$

for $j \in \{1, \dots, K2^K\}$. Since $K \geq M$ we have

$$I = (\cup_{j=1}^{K2^K} A_{+,K,j}) \cup (\cup_{j=1}^{K2^K} A_{-,K,j}).$$

Moreover, if we define a simple function $h\colon I \to \mathbb{R}$ by

$$h(x) = \begin{cases} 2^{-K}(j-1), & x \in A_{+,K,j}, \\ -2^{-K}j, & x \in A_{-,K,j}, \end{cases}$$

then we have $|h(x) - f_M(x)| < \epsilon$ for every $x \in I$.

Now that we have a $\mathbb{R}$-valued simple function h that approximates f_M to within ϵ on I, let us dispense with the cumbersome notation above we introduced to define

h, and instead write $h = \sum_{j=1}^k a_j \chi_{A_j}$ for $a_1, \dots, a_k \in \mathbb{R}$ and for a partition $(A_1, \dots, A_k)$ of I into Lebesgue measurable sets. Fix $j \in \{1, \dots, k\}$. Since A_j is measurable, by Corollary 2.4.20 we can write $A_j = U_j \setminus B_j$ where U_j is open and where $B_j \subseteq U_j$ satisfies $\lambda(B_j) < \frac{\epsilon}{8k}$. Since U_j is open, it is a countable union of disjoint open intervals by Proposition I-2.5.6. If U_j is in fact a finite union of open intervals then denote $V_j = U_j$. If any of the intervals comprising V_j have common endpoints, then these intervals may be shrunk so that their complement in A_j has measure at most $\frac{\epsilon}{2k}$. Next suppose that U_j is a countable union of open intervals $(J_{j,l})_{l \in \mathbb{Z}_{>0}}$. Since U_j is bounded we must have $\sum_{l=1}^\infty \lambda(J_{j,l}) < \infty$. Therefore, there exists $N_j \in \mathbb{Z}_{>0}$ such that $\sum_{j=N_j+1}^\infty \lambda(J_{l,j}) < \frac{\epsilon}{8k}$. We then define $V_j = \cup_{l=1}^{N_j} J_{j,l}$. If any of the intervals $J_{1,j}, \dots, J_{N_j+1,j}$ have common endpoints, they can be shrunk while maintaining the fact that the measure of their complement in A_j is at most $\frac{\epsilon}{2k}$. Define $g\colon I \to \mathbb{R}$ on V_j by asking that $g(x) = a_j$ for $x \in V_j$. Doing this for each $j \in \{1, \dots, k\}$ defines $g\colon I \to \mathbb{R}$ on the set $\cup_{j=1}^k V_j$ which is a finite union of open intervals whose complement has measure at most $\frac{\epsilon}{2}$. The complement to $\cup_{j=1}^k V_j$ is a union of intervals, and on these intervals define g to be, say, 0. Note that g as constructed is a step function, and that $g(x) = h(x)$ for $x \in \cup_{j=1}^k V_j$.

Note that if $x \in (\cup_{j=1}^k V_j) \cup (I \setminus A_M)$ we have

$$|g(x) - f(x)| = |h(x) - f_M(x)| < \epsilon.$$

Therefore,

$$\lambda(\{x \in I \mid f(x) - g(x) \geq \epsilon\}) \subseteq I \setminus ((\cup_{j=1}^k V_j) \cup (I \setminus A_M)),$$

and

$$\lambda(I \setminus ((\cup_{j=1}^k V_j) \cup (I \setminus A_M))) < \epsilon,$$

giving the result. ■

A similar sort of result holds for approximations of measurable functions by continuous functions.

2.9.3 Theorem (Lebesgue measurable functions are approximated by continuous functions) *If* $\mathrm{I} = [\mathrm{a}, \mathrm{b}]$ *is a compact interval, if* $\mathrm{f}\colon \mathrm{I} \to \overline{\mathbb{R}}$ *is measurable and satisfies*

$$\lambda(\{\mathrm{x} \in \mathrm{I} \mid \mathrm{f}(\mathrm{x}) \in \{-\infty, \infty\}\} = 0,$$

and if $\epsilon_1, \epsilon_2 \in \mathbb{R}_{>0}$*, then there exists a continuous function* $\mathrm{h}\colon \mathrm{I} \to \mathbb{R}_{\geq 0}$ *such that*

$$\lambda(\{\mathrm{x} \in \mathrm{I} \mid |\mathrm{f}(\mathrm{x}) - \mathrm{h}(\mathrm{x})| \geq \epsilon_1\}) < \epsilon_2.$$

Proof We shall merely outline how this works, since this is "obvious" once one has the basic idea at hand. We assume that $\epsilon_1 = \epsilon_2 = \epsilon$. By the method of Theorem 2.9.2, we approximate f with a step function g such that

$$\lambda(\{x \in I \mid |f(x) - g(x)| \geq \epsilon\}) < \epsilon.$$

Note that the set of points in I where $|f(x) - g(x)| < \epsilon$ is a finite union of intervals with pairwise disjoint closures on each of which g is constant. The value of g on the intervals complementary to these intervals is of no consequence. To define the continuous function h we ask that h agree with g on the intervals upon which g is constant, and between these intervals we ask that h be a linear function that interpolates between the values of h at the two endpoints. The resulting function clearly satisfies the conclusions of the theorem. ■

In Definition 3.8.28 we will define the support for continuous functions as the closure of the set of points where the function is nonzero. For continuous functions, this is a satisfactory definition. For more general classes of functions, this is not so. For example, if $f\colon \mathbb{R} \to \mathbb{R}$ is the characteristic function of $\mathbb{Q}$, then the definition of support for continuous functions, when applied to f, gives $\operatorname{supp}(f) = \mathbb{R}$. However, this does not reflect the fact that f is zero almost everywhere. So we adapt the notion of support for continuous functions to measurable functions as follows. still ''will''?

2.9.4 Definition (Support of a measurable function) Let $f \in \mathsf{L}^{(0)}(\mathbb{R}; \overline{\mathbb{R}})$ and define

$$\mathscr{O}_f = \{U \subseteq \mathbb{R} \mid U \text{ open and } f(x) = 0 \text{ for almost every } x \in U\}.$$

Then the ***support*** of f is $\operatorname{supp}(f) = \mathbb{R} \setminus (\cup_{U \in \mathscr{O}_f} U)$. •

Being the complement of an open set, the support of a measurable function is closed. The following result gives the essential property of closure.

2.9.5 Proposition (Characterisation of support) *For* $f, g \in \mathsf{L}^{(0)}(\mathbb{R}; \overline{\mathbb{R}})$*, the following two statements hold:*

(i) $f(x) = 0$ *for almost every* $x \in \mathbb{R} \setminus \operatorname{supp}(f)$*;*

(ii) if $f(x) = g(x)$ *for almost every* $x \in \mathbb{R}$ *then* $\operatorname{supp}(f) = \operatorname{supp}(g)$*.*

Proof (i) We have $\mathbb{R} \setminus \operatorname{supp}(f) = \mathscr{O}_f$ in the notation of Definition 2.9.4. Recall from Definition II-1.10.23 that the distance between $x \in \mathbb{R}$ and $A \subseteq \mathbb{R}$ is denoted by

$$\operatorname{dist}(x, A) = \inf\{|y - x| \mid y \in A\}.$$

Let $k \in \mathbb{Z}_{>0}$ and define

$$K_k = \{x \in \mathscr{O}_f \mid \operatorname{dist}(x, \operatorname{supp}(f)) \geq \tfrac{1}{k},\ |x| \leq k\}.$$

By Proposition II-1.10.24, the function $x \mapsto \operatorname{dist}(x, A)$ is continuous. By Corollary I-3.1.4, since the set $[\frac{1}{k}, \infty)$ is closed and since $\overline{\mathsf{B}}(k, 0)$ is closed, K_k is the intersection of closed sets, and so closed. Therefore, since it is also bounded, it is compact. Since $K_k \subseteq \mathscr{O}_f$ and since $\mathscr{O}_f$ is a union of open sets, by the Heine–Borel Theorem, K_k is a finite union of open sets from $\mathscr{O}_f$, say $K_k = \cup_{j=1}^{m_k} U_{k,j}$. Denote

$$Z_{k,j} = \{x \in U_{k,j} \mid f(x) \neq 0\}, \qquad j \in \{1, \dots, m_k\}.$$

Since $f(x) = 0$ for almost every $x \in U_{j_k}$ for each $j_k \in \{1, \dots, m_k\}$, it follows that $\lambda(Z_{k,j}) = 0$. Therefore, since the set of points in K_k at which f is nonzero is $\cup_{j=1}^{m_k} Z_{k,j}$, it follows that $f(x) = 0$ for almost every $x \in K_k$. Now note that $\mathscr{O}_f = \cup_{k \in \mathbb{Z}_{>0}} K_k$. Thus the set of points

$x \in \mathscr{O}_f$ such that $f(x) \neq 0$ is a countable union of sets of measure zero, and so has measure zero. That is, $f(x) = 0$ for almost every $x \in \mathscr{O}_f$.

(ii) We claim that if f and g agree almost everywhere, then $\mathscr{O}_f = \mathscr{O}_g$. Indeed, suppose that $U \in \mathscr{O}_f$ so that $f(x) = 0$ for almost every $x \in U$. Define

$$Z_1 = \{x \in U \mid f(x) \neq 0\}, \qquad Z_2 = \{x \in U \mid g(x) \neq f(x)\}.$$

Note that Z_1 and Z_2 have measure zero and so $Z_1 \cup Z_2$ also has measure zero. Moreover, if $x \in U \setminus (Z_1 \cup Z_2)$ then $g(x) = f(x) = 0$. Thus $U \in \mathscr{O}_g$ and so $\mathscr{O}_f \subseteq \mathscr{O}_g$. Reversing the argument shows that $\mathscr{O}_g \subseteq \mathscr{O}_f$. It then immediately follows that $\mathrm{supp}(f) = \mathrm{supp}(g)$. ■

Let us give an example which shows that the notion of support must be treated with some care, the previous result notwithstanding.

2.9.6 Example (A caveat concerning the support of a function) Note that $\mathbb{Q} \subseteq \mathbb{R}$ has Lebesgue measure zero. It follows, by definition of measure zero, that there exists a countable collection of intervals $((a_j, b_j))_{j \in \mathbb{Z}_{>0}}$ such that

$$\sum_{j=1}^{\infty} |b_j - a_j| < 1$$

and such that $\mathbb{Q} \subseteq \cup_{j \in \mathbb{Z}_{>0}} (a_j, b_j)$. Let us define $A = \cup_{j \in \mathbb{Z}_{>0}} (a_j, b_j)$. By countable-subadditivity of the Lebesgue measure we have $\lambda(A) \leq 1$. We claim that $\mathrm{supp}(\chi_A) = \mathbb{R}$. Indeed, if $U \in \mathscr{O}_{\chi_A}$ then $\lambda(A \cap U) = o$. If U is nonempty then it contains an interval, say (a, b). Note that A is a nonempty open set by Exercise I-2.5.1. Moreover, since there are rational numbers in (a, b) by Proposition I-2.2.15, it follows that $A \cap U$ is a nonempty open set, and so has positive Lebesgue measure. We conclude, therefore, that if $U \in \mathscr{O}_{\chi_A}$ then $U = \emptyset$. Thus $\mathrm{supp}(\chi_A) = \mathbb{R}$, as claimed. The point is that we have

$$\lambda(A) \leq 1 < \lambda(\mathrm{supp}(A)) = \infty.$$

Thus the measure of the support of a function can far exceed the measure of the set of points where the function is nonzero. This is a consequence of our asking that the support be a closed set. •

For continuous functions, the preceding definition of support reduces to the usual one, i.e., the one used in Definition 3.8.28.

2.9.7 Proposition (The support of a continuous function) *If* $f\colon \mathbb{R} \to \mathbb{R}$ *is continuous then*

$$\mathrm{supp}(f) = \mathrm{cl}(\{x \in \mathbb{R} \mid f(x) \neq 0\}).$$

Proof Let $x_0 \in \mathbb{R} \setminus \mathrm{supp}(f)$. Then there exists $U \in \mathscr{O}_f$ such that $x_0 \in U$. By Exercise I-3.1.12 we have $f(x) = 0$ for every $x \in U$. In particular, $f(x_0) = 0$ and, moreover, $f(x) = 0$ in the neighbourhood U of x_0. Thus x_0 cannot be a limit $\lim_{j \to \infty} x_j$ with $f(x_j) \neq 0$. That is,

$$x_0 \notin \mathrm{cl}(\{x \in \mathbb{R} \mid f(x) \neq 0\}).$$

Conversely, suppose that $x_0 \in \mathbb{R} \setminus \mathrm{cl}(\{x \in \mathbb{R} \mid f(x) \neq 0\})$. Then there must be a neighbourhood U of x_0 such that $f(x) = 0$ for every $x \in U$. Thus $U \subseteq \mathscr{O}_f$ and so $x \in \mathbb{R} \setminus \mathrm{supp}(f)$. ■

2.9.2 The (conditional) Lebesgue integral

Let $\mathscr{L}(\mathbb{R})$ be the collection of Lebesgue measurable subsets of $\mathbb{R}$ (see Definition 2.4.4) and let $\lambda\colon \mathscr{L}(\mathbb{R}) \to \overline{\mathbb{R}}_{\geq 0}$ be the Lebesgue measure (see Definition 2.4.4). From Proposition 2.4.6, recall also that if $A \in \mathscr{L}(\mathbb{R})$ then we denote by $\mathscr{L}(A)$ the Lebesgue measurable subsets of A and by λ_A the restriction of λ to $\mathscr{L}(A)$.

Although it is pretty clear if you have been reading this chapter from the beginning, perhaps the following definition ought to be made for those who "skipped to the good bit."

2.9.8 Definition (Lebesgue integral on $\mathbb{R}$) If $f \in \mathsf{L}^{(0)}(\mathbb{R};\overline{\mathbb{R}})$ then f is ***Lebesgue integrable*** and the ***Lebesgue integral*** of f is the integral of f with respect to the Lebesgue measure when the integral exists:

$$\int_{\mathbb{R}} f\,\mathrm{d}\lambda.$$

If $f \in \mathsf{L}^{(0)}(A;\overline{\mathbb{R}})$, then f is ***Lebesgue integrable*** and the ***Lebesgue integral*** of f is the integral of f with respect to the Lebesgue measure when the integral exists:

$$\int_A f\,\mathrm{d}\lambda_A.$$

We shall almost always denote the Lebesgue integrable functions on A by $\mathsf{L}^{(1)}(A;\overline{\mathbb{R}})$ rather than $\mathsf{L}^{(1)}((A,\mathscr{L}(A),\lambda_A);\overline{\mathbb{R}})$. •

Of course, if $A \in \mathscr{L}(\mathbb{R})$ and if $f \in \mathsf{L}^{(0)}(A;\overline{\mathbb{R}})$, we can think of f as being in $\mathsf{L}^{(0)}(\mathbb{R};\overline{\mathbb{R}})$ by making it zero outside A. The resulting function can be directly verified to be measurable (cf. Exercise 2.6.3). We can, therefore, write

$$\int_A f\,\mathrm{d}\lambda_A = \int_{\mathbb{R}} f\,\mathrm{d}\lambda$$

without risk of confusion. When it is convenient to do so, we shall do this. We will also omit the subscript "A" in "$\mathrm{d}\lambda_A$" when the resulting compactness of notation is desired. Thus, we will use the symbols

$$\int_A f\,\mathrm{d}\lambda_A, \quad \int_A f\,\mathrm{d}\lambda, \quad \int_{\mathbb{R}} f\,\mathrm{d}\lambda$$

to stand for the same thing when it is clear from context what is meant.

It is worth making some connections at this point with how we defined the single-variable Riemann integral in Section I-3.4. For the Riemann integral we had two constructions which we showed were equivalent when the domain of the function was a compact interval. However, the so-called conditional Riemann integral generalises the Riemann integral when the domain of the function is a not a compact interval. This can be generalised for the Lebesgue integral as follows.

2.9.9 Definition (Conditionally Lebesgue integrable functions on a general interval) Let $I \subseteq \mathbb{R}$ be an interval and let $f\colon I \to \overline{\mathbb{R}}$ be a function whose restriction to every compact subinterval of I is Lebesgue integrable.

(i) If $I = [a, b]$ then define
$$\oint_I f \, d\lambda_I = \int_I f \, d\lambda.$$

(ii) If $I = (a, b]$ then define
$$\oint_I f \, d\lambda_I = \lim_{r_a \downarrow a} \int_{[r_a,b]} f \, d\lambda_{[r_a,b]}$$
if the limit exists.

(iii) If $I = [a, b)$ then define
$$\oint_I f \, d\lambda_I = \lim_{r_b \uparrow b} \int_{[a,r_b]} f \, d\lambda_{[a,r_b]}$$
if the limit exists.

(iv) If $I = (a, b)$ then define
$$\oint_I f \, d\lambda_I = \lim_{r_a \downarrow a} \int_{[r_a,c]} f \, d\lambda_{[r_a,c]} + \lim_{r_b \uparrow b} \int_{[c,r_b]} f \, d\lambda_{[c,r_b]}$$
for some $c \in (a, b)$, if the limit exists.

(v) If $I = (-\infty, b]$ then define
$$\oint_I f \, d\lambda_I = \lim_{R \to \infty} \int_{[-R,b]} f \, d\lambda_{[-R,b]}$$
if the limit exists.

(vi) If $I = (-\infty, b)$ then define
$$\oint_I f \, d\lambda_I = \lim_{R \to \infty} \int_{[-R,c]} f \, d\lambda_{[-R,c]} + \lim_{r_b \uparrow b} \int_{[c,r_b]} f \, d\lambda_{[c,r_b]}$$
for some $c \in (-\infty, b)$, if the limit exists.

(vii) If $I = [a, \infty)$ then define
$$\oint_I f \, d\lambda_I = \lim_{R \to \infty} \int_{[a,R]} f \, d\lambda_{[a,R]}$$
if the limit exists.

(viii) If $I = (a, \infty)$ then define

$$\oint_I f \, \mathrm{d}\lambda_I = \lim_{r_a \downarrow a} \int_{[r_a, c]} f \, \mathrm{d}\lambda_{[r_a, c]} + \lim_{R \to \infty} \int_{[c, R]} f \, \mathrm{d}\lambda_{[c, R]}$$

for some $c \in (a, \infty)$, if the limit exists.

(ix) If $I = \mathbb{R}$ then define

$$\oint_{\mathbb{R}} f \, \mathrm{d}\lambda = \lim_{R \to \infty} \int_{[-R, c]} f \, \mathrm{d}\lambda_{[-R, c]} + \lim_{R \to \infty} \int_{[c, R]} f \, \mathrm{d}\lambda_{[c, R]}$$

for some $c \in \mathbb{R}$, if the limit exists.

If, for a given I and f, the appropriate of the above limits exists, then f is ***conditionally Lebesgue integrable*** on I, and the ***conditional Lebesgue integral*** is the value of the limit. •

It is not usual to define the conditional Lebesgue integral, but we do so in order to make our analogies with the Riemann integral, explored in Section 2.9.3, more clear. Thus a few comments are relevant at this point.

2.9.10 Remarks (On the conditional Lebesgue integral)

1. Since the Lebesgue integral is so general, it is not really natural to restrict the definition of the Lebesgue integral to functions defined on intervals. Indeed, a somewhat more natural construction would be as follows. Let $A \in \mathscr{L}(\mathbb{R})$ and let $f \colon A \to \overline{\mathbb{R}}$ be measurable. By Theorem 2.4.19 let $(K_j)_{j \in \mathbb{Z}_{>0}}$ be a family of compact sets such that $K_j \subseteq A$, $K_j \subseteq K_{j+1}$, $j \in \mathbb{Z}_{>0}$, and $\lambda(A) = \lim_{j \to \infty} \lambda(K_j)$. Then we can define the conditional Lebesgue integral of f by

$$\oint_A f \, \mathrm{d}\lambda_A = \lim_{j \to \infty} \int_{K_j} (f|K_j) \, \mathrm{d}\lambda_{K_j}.$$

This construction generalises the more complicated, but more direct construction of Definition 2.9.9. Since we will not use this level of generality for the conditional Lebesgue integral, we shall stick to the more concrete Definition 2.9.9 as our definition of the conditional Lebesgue integral. It also make more clear the comparison with the Riemann integral.

2. The conditional Lebesgue integral shares with the Lebesgue integral the usual properties with respect to operations on functions, i.e., those properties given in Section 2.7.2 for the general integral. The verification of this is a matter of using the results of Section 2.7.2, the fact that the conditional Lebesgue integral is defined as a limit, and the fact that limits commute with natural operations as shown in Section I-2.3.6. We leave the details of proving this statement to a sufficiently bored reader. However, we shall make free use of these facts ourselves.

3. In Theorem 2.9.11 we shall show that the (conditional) Lebesgue integral generalises the (conditional) Riemann integral. For this reason, to give an example of a function that is conditionally Lebesgue integrable but not Lebesgue integrable, it suffices to give an example of a function that is conditionally Riemann integrable but not Riemann integrable. Such a function is given in Example I-3.4.20. •

2.9.3 Properties of the Lebesgue integral

In this section we shall give some useful properties of the Lebesgue integral and the conditional Lebesgue integral. In the preceding section we constructed two versions of the Lebesgue integral for functions of a single variable. As was pointed out in the course of these constructions, these two integral mirror in spirit the development in Section I-3.4 for the Riemann integral. We begin this section by showing that the Riemann integral is generalised by the Lebesgue integral.

One of the intentions of Section 2.1 was to show that the Riemann integral suffers a few theoretical defects. If the Lebesgue integral is to redress these problems, it would be helpful it applied in all cases when the Riemann integral applies. This is indeed the case.

2.9.11 Theorem (The (conditional) Lebesgue integral generalises the (conditional) Riemann integral) *If* $\mathrm{I} \subseteq \mathbb{R}$ *is an interval and if* $\mathrm{f}\colon \mathrm{I} \to \mathbb{R}$ *is (conditionally) Riemann integrable, then* f *is (conditionally) Lebesgue integrable, and*

$$(\mathrm{C}) \int_{\mathrm{I}} \mathrm{f}(\mathrm{x})\, \mathrm{dx} = (\mathrm{C}) \int_{\mathrm{I}} \mathrm{f}\, \mathrm{d}\lambda.$$

Proof First let us consider the case where $I = [a, b]$ is compact. Suppose that $f\colon [a, b] \to \mathbb{R}$ is Riemann integrable. For $k \in \mathbb{Z}_{>0}$ let P_k be a partition with the property that $A_+(f, P_k) - A_-(f, P_k) < \frac{1}{k}$. By redefining partitions if necessary we can assume that the endpoints of the intervals for P_{k+1} contain those for P_k, cf. Lemma I-1 from the proof of Theorem I-3.4.9. Upon doing this, the sequences $(s_+(f, P_k)(x))_{k\in\mathbb{Z}_{>0}}$ and $(s_-(f, P_k)(x))_{k\in\mathbb{Z}_{>0}}$ are increasing and decreasing, respectively, for each $x \in [a, b]$. Moreover, since the functions in these sequences are step functions, they are simple functions and so are measurable. It is also clear that the Riemann integral of a step function is equal to the Lebesgue integral of the same function, by definition of the Riemann integral of a step function and the Lebesgue integral of a simple function. Thus

$$\int_{[a,b]} s_+(f, P_k)\, \mathrm{d}\lambda = A_+(f, P_k), \qquad \int_{[a,b]} s_-(f, P_k)\, \mathrm{d}\lambda = A_-(f, P_k).$$

Denote

$$f_+(x) = \lim_{k\to\infty} s_+(f, P_k)(x), \quad f_-(x) = \lim_{k\to\infty} s_-(f, P_k)(x),$$

for $x \in [a, b]$. Proposition 2.6.18 implies that f_+ and f_- are measurable. Note that f, and therefore f_+ and f_-, are bounded. Thus f_+ and f_- are bounded in absolute value by a constant function. Such a function is obviously in $\mathscr{L}^{(1)}([a,b];\mathbb{R})$, and so the Dominated Convergence Theorem implies that

$$\int_{[a,b]} f_+ \,\mathrm{d}\lambda = \lim_{k\to\infty} A_+(f,P_k) = \int_a^b f(x)\,\mathrm{d}x$$

and

$$\int_{[a,b]} f_- \,\mathrm{d}\lambda = \lim_{k\to\infty} A_-(f,P_k) = \int_a^b f(x)\,\mathrm{d}x,$$

where we have used the characterisation of the Riemann integral in Theorem I-3.4.9. From this we conclude that

$$\int_{[a,b]} (f_+ - f_-)\,\mathrm{d}\lambda = 0,$$

which implies that $f_+(x) = f_-(x)$ for almost every $x \in [a,b]$ by Proposition 2.7.14. Since $f_-(x) \le f(x) \le f_+(x)$ for every $x \in []1,b]$, it, therefore, follows that f is itself measurable (being almost everywhere equal to the measurable functions f_+ and f_-) and Lebesgue integrable (again, being almost everywhere equal to the Lebesgue integrable functions f_+ and f_-). Moreover, by Proposition 2.7.11 it follows that

$$\int_{[a,b]} f_+ \,\mathrm{d}\lambda = \int_{[a,b]} f_- \,\mathrm{d}\lambda = \int_{[a,b]} f \,\mathrm{d}\lambda = \int_a^b f(x)\,\mathrm{d}x,$$

as desired.

Now we consider an arbitrary interval $I \subseteq \mathbb{R}$ and suppose that f is Riemann integrable. Here, we first take f to be nonnegative-valued. In this case, the definition of the Riemann integral from Definition I-3.4.14 implies that there exists a sequence $(I_k)_{k\in\mathbb{Z}_{>0}}$ of compact intervals such that $I_k \subseteq I_{k+1}$, $k \in \mathbb{Z}_{>0}$, such that $I = \cup_{k\in\mathbb{Z}_{>0}} I_k$, and such that

$$\int_I f(x)\,\mathrm{d}x = \lim_{k\to\infty} \int_{I_k} f(x)\,\mathrm{d}x.$$

From the Monotone Convergence Theorem, Theorem 2.7.24, and the first part of the proof it then follows that

$$\int_I f\,\mathrm{d}\lambda = \lim_{k\to\infty} \int_{I_k} f(x)\,\mathrm{d}x = \int_I f(x)\,\mathrm{d}x.$$

For general $\mathbb{R}$-valued f, the result follows from writing $f = f_+ - f_-$, and using linearity of the Riemann and Lebesgue integrals, Propositions I-3.4.22 and 2.7.17.

Finally, we consider an arbitrary interval I and suppose that f is conditionally Riemann integrable. According to Definition 2.9.9 there exists a sequence $(K_j = [a_j,b_j])_{j\in\mathbb{Z}_{>0}}$ of compact intervals such that $K_j \subseteq K_{j+1}$, $j \in \mathbb{Z}_{>0}$, and such that $I \cup_{j\in\mathbb{Z}_{>0}} K_j$. By our arguments above we have

$$\int_{K_j} (f|K_j)\,\mathrm{d}\lambda_{K_j} = \int_{a_j}^{b_j} f(x)\,\mathrm{d}x, \qquad j \in \mathbb{Z}_{>0}.$$

Therefore,

$$\lim_{j\to\infty}\int_{K_j}(f|K_j)\,\mathrm{d}\lambda_{K_j}=\lim_{j\to\infty}\int_{a_j}^{b_j}f(x)\,\mathrm{d}x,$$

and the result follows by the definitions of the conditional Riemann and Lebesgue integrals. ■

We must, of course, also show that there are Lebesgue integrable functions that are not Riemann integrable.

2.9.12 Example (A Lebesgue integrable, but not Riemann integrable, function) Let $I=[0,1]$ and let $A=\mathbb{Q}\cap[0,1]$. Then define $f\colon[0,1]\to\mathbb{R}$ by $f=\chi_A$. Note that f is not Riemann integrable; see Example I-3.4.10. However, f is Lebesgue integrable, as can be seen in many ways. Most directly, f is the characteristic function of the Lebesgue measurable set A, and so is Lebesgue integrable simply by definition. If one wishes, one can also "derive" the Lebesgue integrability of f. For example, if we let $(q_k)_{k\in\mathbb{Z}_{>0}}$ be an enumeration of the set A, we can define $g_k\colon[0,1]\to\mathbb{R}$ by

$$g_k(x)=\begin{cases}1, & x\in\{q_1,\dots,q_k\},\\ 0, & \text{otherwise.}\end{cases}$$

The functions g_k, $k\in\mathbb{Z}_{>0}$, are Lebesgue integrable, indeed Riemann integrable, cf. Example 2.1.11. Moreover, $f(x)=\lim_{k\to\infty}g_k(x)$ for all $x\in[0,1]$. By the Dominated Convergence Theorem (verify its hypotheses!), we then have

$$\int_{[0,1]}f\,\mathrm{d}\lambda=\lim_{k\to\infty}\int_{[0,1]}g_k\,\mathrm{d}\lambda=\lim_{k\to\infty}\int_0^1 g_k(x)\,\mathrm{d}x=0.$$

Thus f is indeed Lebesgue integrable, with Lebesgue integral zero. •

It is rather important not to overstate the importance of this example. It is not interesting, but it does serve to easily verify that the Lebesgue integral generalises the Riemann integral.

2.9.13 Notation and Remarks (Riemann integral versus Lebesgue integral) Having now established the relationship between the Riemann and Lebesgue integrals, we shall often use the sometimes more convenient notation for the Riemann integral when we actually are using the Lebesgue integral. Thus, for example, we may well write

$$\int_a^b f(x)\,\mathrm{d}x,\quad \int_{-\infty}^b f(x)\,\mathrm{d}x,\quad \oint_a^\infty f(x)\,\mathrm{d}x$$

where we really mean

$$\int_{[a,b]}f\,\mathrm{d}\lambda_{[a,b]},\quad \int_{(-\infty,b]}f\,\mathrm{d}\lambda_{(-\infty,b]},\quad \oint_{[a,\infty)}f\,\mathrm{d}\lambda_{[a,\infty)},$$

respectively.

This confounding of notation for the Lebesgue and Riemann integrals suggests that the additional generality of the Lebesgue integral is not of great importance. This is both true and not true. It *is* true that we shall not encounter specific examples of Lebesgue integrable functions that are not Riemann integrable. That is to say, we shall not often care to compute the Lebesgue integral in cases where the Riemann integral will not suffice. However, it *is* the case that the Riemann integral has certain undesirable features, as we discussed in Section 2.1.2. These undesirable features come in two basic flavours.

1. The Riemann and Lebesgue integrals both possess a Dominated Convergence Theorem, Theorems II-1.7.8 and 2.7.28, respectively. However, the two theorems differ in a crucial way. Specifically, in the Dominated Convergence Theorem for the Riemann integral, the Riemann integrability of the limit function is an hypothesis, while in the Dominated Convergence Theorem for the Lebesgue integral, the integrability of the limit function is a conclusion. This inability of the Dominated Convergence Theorem for the Riemann integral to predict the integrability of the limit function is a crucial defect. We shall discuss this further in Section 2.9.11.
2. It is interesting to consider not just individual Riemann or Lebesgue integrable functions, but the *set* of all Riemann or Lebesgue integrable functions. We have already denoted by $\mathsf{L}^{(1)}(I;\mathbb{R})$ the set of $\mathbb{R}$-valued Lebesgue integrable functions on the interval I. Let us denote by $\mathsf{R}^{(1)}(I;\mathbb{R})$ the set of $\mathbb{R}$-valued Riemann integrable functions on I, cf. the discussion preceding Proposition 2.1.12. Both $\mathsf{L}^{(1)}(I;\mathbb{R})$ and $\mathsf{R}^{(1)}(I;\mathbb{R})$ are $\mathbb{R}$-vector spaces by the standard linearity properties of the integral. In Chapter 3 we shall discuss the notion of a normed vector space and the important related notion of completeness. We shall show in Theorem 3.8.59 (essentially) that the set of Lebesgue integrable functions form a complete normed vector space. This is not the case for Riemann integrable functions, as we show in Proposition 2.1.12. It may not be clear at this point why this is important, but this is, in fact, *extremely* important. As we go along, and we use the Lebesgue integral at various points in these volumes, we shall point out instances where the particular properties of the Lebesgue integral are crucial. •

Now that we have established the close relationship between the Lebesgue and Riemann integrals, let us explore some of the properties of Lebesgue integrable functions. In Section 2.9.1 we explored the manner in which Lebesgue measurable functions can be pointwise approximated by "nice" functions like step functions or continuous functions. Lebesgue integrable functions, being Lebesgue measurable, are subject to the same approximations. However, for Lebesgue integrable functions we have another sort of approximation that is possible by virtue of the integral.

2.9.14 Theorem (Lebesgue integrable functions are approximated by step func-

tions) *If* $\mathrm{I} = [\mathrm{a}, \mathrm{b}]$ *is a compact interval, if* $\mathrm{f} \in \mathsf{L}^{(1)}(\mathrm{I}; \overline{\mathbb{R}})$*, and if* $\epsilon \in \mathbb{R}_{>0}$*, then there exists a step function* $\mathrm{g}\colon \mathrm{I} \to \mathbb{R}$ *such that*

$$\int_{\mathrm{I}} |\mathrm{f} - \mathrm{g}| \,\mathrm{d}\lambda_{\mathrm{I}} < \epsilon.$$

Proof Let us first consider the case when f is bounded. Let $M \in \mathbb{R}_{>0}$ be such that $f(x) \leq M$ for all $x \in I$. Let $\epsilon \in \mathbb{R}_{>0}$. By Theorem 2.9.2 there exists a continuous function $g\colon I \to \mathbb{R}_{\geq 0}$ such that

$$\lambda\Big(\Big\{x \in I \;\Big|\; |f(x) - g(x)| < \tfrac{\epsilon}{(2(b-a))}\Big\}\Big) < \frac{\epsilon}{2M}.$$

Then

$$\int_I |f(x) - g(x)| \,\mathrm{d}\lambda_I < \frac{\epsilon}{2(b-a)}(b-a) + \frac{\epsilon}{2M}M < \epsilon,$$

giving the result in this case.

Next we consider the case when f is possibly unbounded and takes values in $\overline{\mathbb{R}}_{\geq 0}$. Let $\epsilon \in \mathbb{R}_{>0}$. For $M \in \mathbb{R}_{>0}$ define

$$f_M(x) = \begin{cases} f(x), & f(x) \leq M, \\ M, & f(x) > M. \end{cases}$$

Since $f \in \mathsf{L}^{(1)}(I; \overline{\mathbb{R}})$ we have $f(x) = \lim_{M\to\infty} f_M(x)$ for almost every $x \in I$. By the Dominated Convergence Theorem,

$$\lim_{M\to\infty} \int_I (f - f_M) \,\mathrm{d}\lambda_I = 0.$$

Thus there exists M sufficiently large that

$$\int_I |f(x) - f_M(x)| \,\mathrm{d}\lambda_I < \frac{\epsilon}{2}.$$

By the argument in the previous paragraph there exists a step function $g\colon I \to \mathbb{R}_{\geq 0}$ such that

$$\int_I |f_M - g| \,\mathrm{d}\lambda_I < \frac{\epsilon}{2}.$$

Then, using the triangle inequality and monotonicity of the integral, Proposition 2.7.19,

$$\int_I |f - g| \,\mathrm{d}\lambda_I \leq \int_I |f - f_M| \,\mathrm{d}\lambda_I + \int_I |f_M - g| \,\mathrm{d}\lambda_I < \epsilon,$$

giving the result in this case.

Finally, if f is $\overline{\mathbb{R}}$-valued, we write $f = f_+ - f_-$ for f_+ and f_- taking values in $\overline{\mathbb{R}}_{\geq 0}$. Let $\epsilon \in \mathbb{R}_{>0}$. By our arguments above there exists step functions $g_+, g_-\colon I \to \mathbb{R}_{\geq 0}$ such that

$$\int_I |f_+ - g_+| \,\mathrm{d}\lambda_I < \frac{\epsilon}{2}, \quad \int_I |f_- - g_-| \,\mathrm{d}\lambda_I < \frac{\epsilon}{2}.$$

Taking $g = g_+ - g_-$, the triangle inequality and Proposition 2.7.19 then give

$$\int_I |f - g| \, d\lambda_I \leq \int_I |f_+ - g_+| \, d\lambda_I + \int_I |f_- - g_-| \, d\lambda_I < \epsilon,$$

as desired. ■

A similar result as the previous holds for approximations of integrable functions by continuous functions. However, in this case it is possible to even be more general in terms of the domain of definition of the functions involved. The notion of support is used in the title of this theorem, but will only be introduced in Definition 3.8.28.

2.9.15 Theorem (Lebesgue integrable functions are approximated by compactly supported continuous functions) *If* $I \subseteq \mathbb{R}$ *is an interval, if* $f \in \mathsf{L}^{(1)}(I; \overline{\mathbb{R}})$*, and if* $\epsilon \in \mathbb{R}_{>0}$*, then there exists a continuous function* $g \colon I \to \mathbb{R}$ *such that*

$$\int_I |f - g| \, d\lambda_I < \epsilon$$

and such that the support of f*, i.e., the set*

$$\mathrm{cl}_I(\{x \in I \mid f(x) \neq 0\}),$$

is compact.

Proof If I is compact, then the result follows just like Theorem 2.9.14, but using Theorem 2.9.3 rather than Theorem 2.9.2. Thus the result holds when I is compact.

Thus we need only consider the case when I is not compact. Let $\epsilon \in \mathbb{R}_{>0}$. We let $(I_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence of compact intervals such that $I_j \subseteq I_{j+1}$ for each $j \in \mathbb{Z}_{>0}$ and such that $\cup_{j \in \mathbb{Z}_{>0}} I_j = I$. Define a sequence $(f_j)_{j \in \mathbb{Z}_{>0}}$ in $\mathsf{L}^{(1)}(I; \mathbb{R})$ by

$$f_j(x) = \begin{cases} f(x), & x \in I_j, \\ 0, & \text{otherwise.} \end{cases}$$

By the Monotone Convergence Theorem we have

$$\lim_{j \to \infty} \int_I |f - f_j| \, d\lambda_I = \int_I \lim_{j \to \infty} |f - f_j| \, d\lambda_I = 0.$$

Thus $(f_j)_{j \in \mathbb{Z}_{>0}}$ converges to f in $\mathsf{L}^{(1)}(I; \mathbb{F})$. Now, for each $j \in \mathbb{Z}_{>0}$, the fact that the theorem holds for compact intervals ensures the existence of a continuous function $h_j \colon I_j \to \mathbb{R}_{\geq 0}$ such that

$$\int_{I_j} |f_j|I_j - h_j| \, d\lambda_{I_j} < \frac{\epsilon}{4}.$$

Note that if we extend h_j to I by asking that it be zero on $I \setminus I_j$ then this extension may not be continuous. However, we can linearly taper h_j to zero on $I \setminus I_j$ to arrive at a continuous function $g_j \colon I \to \mathbb{R}_{\geq 0}$ with compact support satisfying

$$\int_{I \setminus I_j} |g_j| \, d\lambda_{I \setminus I_j} < \frac{\epsilon}{4}.$$

Then

$$\int_I |f_j - g_j| \, d\lambda_I = \int_{I_j} |f_j - h_j| \, d\lambda_{I_j} + \int_{I \setminus I_j} |g_j(x)| \, d\lambda_{I \setminus I_j} < \frac{\epsilon}{4} + \frac{\epsilon}{4} < \frac{\epsilon}{2}.$$

Now choose $N \in \mathbb{Z}_{>0}$ sufficiently large that

$$\int_I |f - f_j| \, d\lambda_I < \frac{\epsilon}{2}.$$

Then, by the triangle inequality,

$$\int_I |f - g_j| \, d\lambda_I \leq \int_I |f - f_j| \, d\lambda_I + \int_I |f_j - g_j| \, d\lambda_I < \epsilon,$$

as desired. ■

2.9.4 Swapping operations with the Lebesgue integral

It is useful to have at hand results that tell us the nature of an integral as a function of a parameter. Thus we let $A \in \mathscr{L}(\mathbb{R})$ and let (a, b) be an open interval. We suppose that $f\colon (a, b) \times A \to \mathbb{R}$ has the property that, for $p \in (a, b)$, the function $x \mapsto f(p, x)$ is integrable. We denote $f^p(x) = f(p, x)$ and $f_x(p) = f(p, x)$. We then define

$$I_f(p) = \int_A f^p(x) \, dx.$$

The next result indicates when such a function is continuous or differentiable.

2.9.16 Theorem (Continuous and differentiable dependence of integral on a parameter) *Let* $(a, b) \subseteq \mathbb{R}$, *let* $A \in \mathscr{L}(\mathbb{R})$, *and let* $f\colon (a, b) \times A \to \mathbb{R}$ *have the property that* $f^p \in \mathsf{L}^{(1)}(A; \mathbb{R}))$ *for every* $p \in (a, b)$. *Let* $p_0 \in (a, b)$.

(i) If f_x *is continuous at* p_0 *for almost every* $x \in A$ *and if there exists* $g \in \mathsf{L}^{(1)}(A; \mathbb{R})$ *and a neighbourhood* U *of* p_0 *in* (a, b) *for which* $|f^p(x)| \leq g(x)$ *for all* $p \in U$, *then* I_f *is continuous at* p_0.

(ii) If there exists $\epsilon \in \mathbb{R}_{>0}$ *so that*

(a) $(p_0 - \epsilon, p_0 + \epsilon) \subseteq (a, b)$,

(b) f^p *is differentiable on* $(p_0 - \epsilon, p_0 + \epsilon)$, *and*

(c) there exists $g \in \mathsf{L}^{(1)}(A; \mathbb{R})$ *so that* $\left|\frac{\partial f}{\partial p}(p, x)\right| \leq g(x)$ *for* $p \in (p_0 - \epsilon, p_0 + \epsilon)$ *and for almost every* $x \in A$,

then I_f *is differentiable at* p_0 *and*

$$I_f'(p_0) = \int_A \frac{\partial f}{\partial p}(p_0, x) \, dx.$$

Proof (i) Let $(p_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in U this neighbourhood converging to p_0. By the Dominated Convergence Theorem we have

$$\lim_{j \to \infty} \int_A f(p_j, x) \, dx = \int_A \lim_{j \to \infty} f(p_j, x) \, dx = \int_A f(p_0, x) \, dx,$$

the final equality by continuity of f_x for almost every $x \in A$ and by Theorem I-3.1.3. This shows that $\lim_{j\to\infty} I_f(p_j) = I_f(p_0)$, giving the result by another application of Theorem I-3.1.3.

(ii) We again let $(p_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence approaching p_0. By the Mean Value Theorem, for each $j \in \mathbb{Z}_{>0}$, there exists q_j between p_j and p_0 such that

$$\frac{f(p_j,x) - f(p_0,x)}{p_j - p_0} = \frac{\partial f}{\partial p}(q_j,x).$$

Note that we necessarily have $\lim_{j\to\infty} q_j = p_0$. Then we compute

$$\begin{aligned}\lim_{j\to\infty}\frac{I_f(p_j) - I_f(p_0)}{p_j - p_0} &= \lim_{j\to\infty}\int_A \frac{f(p_j,x) - f(p_0,x)}{p_j - p_0}\,\mathrm{d}x = \int_A \lim_{j\to\infty}\frac{f(p_j,x) - f(p_0,x)}{p_j - p_0}\,\mathrm{d}x\\ &= \int_A \lim_{j\to\infty}\frac{\partial f}{\partial p}(q_j,x)\,\mathrm{d}x = \int_A \frac{\partial f}{\partial p}(p_0,x)\,\mathrm{d}x.\end{aligned}$$

Here the interchanging of the limit and the integral is valid by the Dominated Convergence Theorem. ■

The above theorem is proved using tools that we presently have at our disposal. It suffices for many purposes. However, it is possible to weaken the hypotheses significantly while retaining the same conclusions, but at a price of using the notion of absolute continuity we introduce in Section 2.9.6 and the formalism of distributions we introduce in Chapter IV-3.

2.9.17 Theorem (A strong theorem on differential dependence of integral on a parameter) *Let* $\mathrm{A} \in \mathscr{L}(\mathbb{R})$ *and let* $\mathrm{f}\colon \mathbb{R}\times\mathrm{A} \to \mathbb{R}$ *have the properties*

(i) that $\mathrm{f_x}$ *is locally absolutely continuous for almost every* $\mathrm{x} \in \mathrm{A}$ *and*

(ii) that, for every compact subset $\mathrm{K} \subseteq \mathbb{R}$*, the functions*

$$(\mathrm{p},\mathrm{x}) \mapsto \mathrm{f}(\mathrm{p},\mathrm{x}), \quad (\mathrm{p},\mathrm{x}) \mapsto \mathbf{D}_1\mathrm{f}(\mathrm{p},\mathrm{x}),$$

when restricted to $\mathrm{K}\times\mathrm{A}$*, are integrable.*

Then, if $\mathrm{I_f}\colon \mathbb{R}\to\mathbb{R}$ *is as above,* $\mathrm{I_f}$ *is locally absolutely continuous and*

$$\mathrm{I}'_{\mathrm{f}}(\mathrm{p}) = \int_{\mathrm{A}} \mathbf{D}_1\mathrm{f}(\mathrm{p},\mathrm{x})\,\mathrm{dx}$$

for almost every $\mathrm{p} \in \mathrm{I}$*.*

Proof For $x \in A$ let $\theta_f(x) \in \mathscr{D}'(\mathbb{R};\mathbb{R})$ be the regular distribution associated with f_x. Adopting and slightly modifying the notation used in Proposition IV-3.2.42, let us define $F_f\colon A\times\mathscr{D}(\mathbb{R};\mathbb{R}) \to \mathbb{R}$ by

$$F_f(x,\phi) = \langle\theta_f(x);\phi\rangle = \int_{\mathbb{R}} f(p,x)\phi(p)\,\mathrm{d}p,$$

for $\phi \in \mathscr{D}(\mathbb{R};\mathbb{R})$ define $F_{f,\phi}\colon A \to \mathbb{R}$ by

$$F_{f,\phi}(x) = F_f(x,\phi),$$

and then define $\Theta_f\colon \mathscr{D}(\mathbb{R};\mathbb{R}) \to \mathbb{R}$ by

$$\Theta_f(\phi) = \int_A F_{f,\phi}\,\mathrm{d}\lambda = \int_A\Big(\int_{\mathbb{R}} f(p,x)\phi(p)\,\mathrm{d}p\Big)\,\mathrm{d}x.$$

We first claim that $\Theta_f \in \mathscr{D}'(\mathbb{R};\mathbb{R})$. We prove this by verifying that the hypotheses of Proposition IV-3.2.42 are satisfied. Let $(\phi_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence converging to zero in $\mathscr{D}(\mathbb{R};\mathbb{R})$. Let $K \subseteq \mathbb{R}$ be a compact interval for which $\operatorname{supp}(\phi_j) \subseteq K$ for every $j \in \mathbb{Z}_{>0}$. Let

$$M = \sup\{|\phi_j(p)| \mid j \in \mathbb{Z}_{>0},\ p \in \mathbb{R}\},$$

noting that $M < \infty$ since the sequence $(\phi_j)_{j\in\mathbb{Z}_{>0}}$ converges uniformly to zero. Then we have

$$\begin{aligned}\int_A |\sup\{F_{f,\phi_j}(x) \mid j \in \mathbb{Z}_{>0}\}|\,\mathrm{d}x &\le \int_A\Big(\int_{\mathbb{R}} \sup\{|f(p,x)\phi_j(p)| \mid j \in \mathbb{Z}_{>0}\}\,\mathrm{d}p\Big)\,\mathrm{d}x\\ &\le M\int_A\Big(\int_K |f(p,x)|\,\mathrm{d}p\Big)\,\mathrm{d}x < \infty,\end{aligned}$$

by hypothesis.

Now, for $\phi \in \mathscr{D}(\mathbb{R};\mathbb{R})$ we compute

$$\begin{aligned}\Theta_f'(\phi) = -\Theta_f(\phi') &= -\int_A\Big(\int_{\mathbb{R}} f(p,x)\phi'(p)\,\mathrm{d}p\Big)\,\mathrm{d}x\\ &= \int_A\Big(\int_{\mathbb{R}} \boldsymbol{D}_1 f(p,x)\phi(p)\,\mathrm{d}p\Big)\,\mathrm{d}x = \int_{\mathbb{R}}\Big(\int_A \boldsymbol{D}_1 f(p,x)\,\mathrm{d}x\Big)\phi(p)\,\mathrm{d}p\end{aligned}$$

using Proposition 2.9.36, the fact that ϕ has compact support, and Fubini's Theorem. This shows that Θ_f' is equal to the regular distribution associated with the function

$$p \mapsto \int_A \boldsymbol{D}_1 f(p,x)\,\mathrm{d}x.$$

By Proposition IV-3.2.31 it follows that this function is locally absolutely continuous and that it is equal almost everywhere to the derivative of the function

$$p \mapsto \int_A f(p,x)\,\mathrm{d}x,$$

which is the desired result. ■

It is also useful to have at hand a result which indicates when holomorphicity of the integrand implies holomorphicity of the integral.

2.9.18 Theorem (Holomorphic dependence on a parameter) *Suppose we have the following data:*

(i) *a measurable subset* $A \subseteq \mathbb{R}$*;*

(ii) *an open subset* $D \subseteq \mathbb{C}$*;*

(iii) *a function* $G\colon A \times D \to \mathbb{C}$ *such that*

(a) *the function* $x \mapsto G(x, z)$ *is in* $\mathsf{L}^{(1)}(A; \mathbb{C})$ *for each* $z \in D$*,*

(b) *the function* $z \mapsto G(x, z)$ *is in* $\mathsf{H}(D, \mathbb{C})$ *for each* $x \in A$*, and*

(c) *for each* $z_0 \in D$ *there exists a neighbourhood* U *of* z_0 *in* D *and* $h \in \mathsf{L}^{(1)}(A; \mathbb{R}_{\geq 0})$ *such that* $|G(x, z)| \leq h(x)$ *for each* $z \in U$*.*

Then the function $F\colon D \to \mathbb{C}$ *defined by*

$$F(z) = \int_A G(x, z)\,\mathrm{d}x$$

is in $\mathsf{H}(D; \mathbb{C})$*.*

Proof By Theorem 2.9.16 we know that F is continuous in D. Now let Γ be a closed contour in D. Parameterise Γ with a map $\gamma\colon [0, L] \to D$ so that make sure this is right

$$\int_\Gamma F(z)\,\mathrm{d}z = \int_0^L \Big(\int_A G(x, \gamma(s))\,\mathrm{d}x\Big)\,\mathrm{d}s.$$

Then the function $x \mapsto G(x, \gamma(s))$ is in $\mathsf{L}^{(1)}(A; \mathbb{C})$ for every $s \in [0, L]$. Also, the function $s \mapsto G(x, \gamma(s))$ is in $\mathsf{L}^{(1)}([0, L]; \mathbb{C})$ for every $x \in A$ since it is a continuous function defined on a compact interval. Therefore, Fubini's Theorem gives

$$\int_\Gamma F(z)\,\mathrm{d}z = \int_A \Big(\int_0^L G(x, \gamma(s))\,\mathrm{d}s\Big)\,\mathrm{d}x = \int_A \Big(\int_\Gamma G(x, z)\,\mathrm{d}z\Big) = 0,$$

using Cauchy's Theorem and holomorphicity of $z \mapsto G(x, z)$. Since this holds for every closed contour in D, Morera's Theorem allows us to conclude that F is holomorphic in D. ■ make sure this is right

2.9.5 Locally Lebesgue integrable functions

Very often on wants to speak of functions that are integrable about every point, but which may not be integrable on their entire domain. This is another instance of the concept of "locality" that we have encountered many times before.

2.9.19 Definition (Locally Lebesgue integrable function) If $A \in \mathscr{L}(\mathbb{R})$ then $f \in \mathsf{L}^{(0)}(A; \overline{\mathbb{R}})$ is ***locally Lebesgue integrable***, or merely ***locally integrable***, if, for every compact set $K \subseteq A$, $f|K \in \mathsf{L}^{(1)}(K; \overline{\mathbb{R}})$. The set of locally Lebesgue integrable functions on A is denoted by $\mathsf{L}^{(1)}_{\mathrm{loc}}(A; \overline{\mathbb{R}})$. •

Note that if $f \in \mathsf{L}^{(0)}(A; \overline{\mathbb{R}})$ then one can define $\bar{f}\colon \mathbb{R} \to \overline{\mathbb{R}}$ to be defined to be equal to f on A and zero elsewhere. Moreover, $f \in \mathsf{L}^{(1)}_{\mathrm{loc}}(A; \overline{\mathbb{R}})$ if and only if $\bar{f} \in \mathsf{L}^{(1)}_{\mathrm{loc}}(\mathbb{R}; \overline{\mathbb{R}})$. Therefore, when talking about locally integrable functions one can, without loss of generality, think about functions whose domain is $\mathbb{R}$. When it is convenient to do this, we shall.

It is obvious that if f is integrable then it is locally integrable. Let us give some examples which clarify the meaning of local integrability as opposed to integrability.

2.9.20 Examples (Local integrability)

1. The function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is locally integrable (its restriction to every compact set is continuous and bounded) but not integrable.
2. The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by
$$f(x) = \begin{cases} x^{-1/2}, & x \in \mathbb{R}_{>0}, \\ 0, & \text{otherwise}, \end{cases}$$
is locally integrable but not integrable.
3. The function $f\colon [0,1] \to \mathbb{R}$ defined by
$$f(x) = \begin{cases} x^{-1/2}, & x \in (0,1], \\ 0, & x = 0, \end{cases}$$
is both locally integrable and integrable.
4. The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by
$$f(x) = \begin{cases} x^{-1}, & x \in (0,1], \\ 0, & \text{otherwise}, \end{cases}$$
is both locally integrable and integrable. •

The following characterisation of locally integrable functions is sometimes useful.

2.9.21 Proposition (Characterisation of locally Lebesgue integrable functions) *For a function* $f\colon \mathbb{R} \to \overline{\mathbb{R}}$ *the following statements are equivalent:*

(i) f *is locally Lebesgue integrable;*

(ii) for each $x \in \mathbb{R}$ *there exists a neighbourhood* U *of* x *such that* $f|U \in \mathsf{L}^{(1)}(U; \overline{\mathbb{R}})$*;*

(iii) for every continuous function $g\colon \mathbb{R} \to \mathbb{R}$ *such that* $\mathrm{supp}(g)$ *is compact, it holds that* $fg \in \mathsf{L}^{(1)}(\mathbb{R}; \overline{\mathbb{R}})$.

Proof (i) $\implies$ (ii) Let $x \in \mathbb{R}$ and let $K = [x-1, x+1]$. By hypothesis, $f|K \in \mathsf{L}^{(1)}(K; \overline{\mathbb{R}})$ and so $f \in \mathsf{L}^{(1)}((x-1, x+1); \overline{\mathbb{R}})$ by Proposition 2.7.22. This gives the result with $U = (x-1, x+1)$.

(ii) $\implies$ (iii) Let $g\colon \mathbb{R} \to \mathbb{R}$ be continuous with compact support. For $x \in \operatorname{supp}(g)$ there exists a neighbourhood U_x of x such that $f \in \mathsf{L}^{(1)}(U_x, \overline{\mathbb{R}})$. Since $(U_x)_{x \in \operatorname{supp}(g)}$ covers the compact set $\operatorname{supp}(g)$ there exists $x_1, \dots, x_k \in \operatorname{supp}(g)$ such that $\operatorname{supp}(g) \subseteq \cup_{j=1}^k U_{x_j}$. Since g is continuous with compact support there exists $M \in \mathbb{R}_{>0}$ such that $|g(x)| \leq M$ for every $x \in \mathbb{R}$ by Theorem I-3.1.22. Then

$$\int_{\mathbb{R}} |fg| \, d\lambda \leq M \int_{\operatorname{supp}(g)} f \, d\lambda_{\operatorname{supp}(g)} \leq M \sum_{j=1}^k \int_{U_{x_j}} f \, d\lambda_{U_{x_j}} < \infty,$$

since $\operatorname{supp}(g) \subseteq \cup_{j=1}^k U_{x_j}$. This gives the desired conclusion.

(iii) $\implies$ (i) Let $K \subseteq \mathbb{R}$ be compact and let $a, b \in \mathbb{R}$ be such that $K \subseteq [a, b]$. Now take $g\colon \mathbb{R} \to \mathbb{R}$ defined by

$$g(x) = \begin{cases} 1, & x \in [a,b], \\ x - (a-1), & x \in [a-1, a), \\ -x + (b+1), & x \in (b, b+1], \\ 0, & \text{otherwise}. \end{cases}$$

Note that g is positive, continuous with compact support, and $g(x) = 1$ for all $x \in [a, b]$. Then

$$\int_K |f| \, d\lambda_K \leq \int_{[a,b]} |f| \, d\lambda_{[a,b]} \leq \int_{\mathbb{R}} |fg| \, d\lambda < \infty,$$

giving the result. ■

Using the preceding characterisation of locally integrable functions, one can easily prove that the set of locally integrable functions is a subspace of the measurable functions.

2.9.22 Proposition (Algebraic operations on locally integrable functions) *If* $A \in \mathscr{L}(\mathbb{R})$*, if* $f, g \in \mathsf{L}^{(1)}_{\mathrm{loc}}(A; \overline{\mathbb{R}})$*, and if* $a \in \mathbb{R}$*, then*

(i) $f + g \in \mathsf{L}^{(1)}_{\mathrm{loc}}(A; \overline{\mathbb{R}})$ *and*

(ii) $af \in \mathsf{L}^{(1)}_{\mathrm{loc}}(A; \overline{\mathbb{R}})$.

Proof This follows from the definition of local integrability, along with Proposition 2.7.17. ■

Local integrability is not preserved by products and quotients, cf. Example 2.7.18.

2.9.6 Absolute continuity

In this section we introduce a special class of continuous functions that are almost everywhere differentiable. With this class of functions one can prove a stronger form of the Fundamental Theorem of Calculus than was possible when we initially discussed this in Section I-3.4.6.

The definition of absolute continuity shares with the definition of bounded variation the feature of being unbearably cryptic at first sight. However, we shall see as we go along that absolute continuity is a notion that arises naturally from the Lebesgue integral.

2.9.23 Definition ((Locally) absolutely continuous function) Let $[a,b]$ be a compact interval. A function $f\colon [a,b] \to \mathbb{R}$ is ***absolutely continuous*** if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $((a_j,b_j))_{j\in\{1,\dots,k\}}$ is a finite family of disjoint open intervals for which

$$\sum_{j=1}^{k} |b_j - a_j| < \delta,$$

then

$$\sum_{j=1}^{k} |f(b_j) - f(a_j)| < \epsilon.$$

For a general interval $I \subseteq \mathbb{R}$, a function $f\colon I \to \mathbb{R}$ is ***locally absolutely continuous*** if $f|J$ is absolutely continuous for every compact interval $J \subseteq I$. We denote by $\mathsf{AC}(I;\mathbb{R})$ (resp. $\mathsf{AC}_{\mathrm{loc}}(I;\mathbb{R})$) the set of absolutely continuous (resp. locally absolutely continuous) functions on the interval I. •

We can make the same sort of comments concerning "absolute continuity" versus "local absolute continuity" as were made in Notation I-3.3.8 concerning the relationship between "bounded variation" and "locally bounded variation."

The following result gives the most basic properties of absolutely functions.

2.9.24 Proposition (Locally absolutely continuous functions are continuous and of locally bounded variation) *If* I $\subseteq \mathbb{R}$ *is an interval and if* f: I $\to \mathbb{R}$ *is a locally absolutely continuous function, then* f *is continuous and has locally bounded variation.*

Proof We first consider the case where $I = [a,b]$. Let $x \in [a,b]$ and let $\epsilon \in \mathbb{R}_{>0}$. Then, by definition of absolute continuity, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $[c,d] \subseteq [a,b]$ is an interval for which $d - c < \delta$, then $|f(d) - f(c)| < \epsilon$. In particular, if $y \in \mathsf{B}(\delta,x) \cap I$, then $|f(y) - f(x)| < \epsilon$, giving continuity of f at x. Now let $\epsilon \in \mathbb{R}_{>0}$ and let $\delta \in \mathbb{R}_{>0}$ have the property that for any family $((a_j,b_j))_{j\in\{1,\dots,k\}}$ of disjoint intervals for which

$$\sum_{j=1}^{k} |b_j - a_j| < \delta,$$

we have

$$\sum_{j=1}^{k} |f(b_j) - f(a_j)| < \epsilon.$$

Now let P be a partition of $[a, b]$ for which $|P| < \delta$, and let $\mathrm{EP}(P) = (x_0, x_1, \dots, x_k)$. Noting that $((x_{j-1}, x_j))_{j\in\{1,\dots,k\}}$ is a finite family of disjoint intervals, we have

$$\sum_{j=1}^{k} |f(x_j) - f(x_{j-1})| < k\epsilon.$$

Since this holds for any partition P for which $|P| < \delta$, and since the expression

$$\sum_{j=1}^{k} |f(x_j) - f(x_{j-1})|$$

is monotonically increasing as a function of $|P|$, it follows that

$$\mathrm{TV}(f) = \sup\Big\{ \sum_{j=1}^{l} |f(x_j) - f(x_{j-1})| \;\Big|\; (x_0, x_1, \dots, x_l) = \mathrm{EP}(P),\ P \in \mathrm{Part}([a, b])\Big\} \le k\epsilon,$$

showing that f has bounded variation.

The result for general intervals follows directly from the result for compact intervals, along with the definition of local absolute continuity. ■

The converse of the preceding result is generally not true, as the following example illustrates.

2.9.25 Example (A continuous function of bounded variation that is not absolutely continuous) We consider the Cantor function $f_C\colon [0, 1] \to \mathbb{R}$ of Example I-3.2.27. We have shown that f_C is continuous, and since it is monotonically increasing, it is necessarily of bounded variation by Theorem I-3.3.3. We claim, nonetheless, that f_C is not locally absolutely continuous. To see this, let $\delta \in \mathbb{R}_{>0}$. Recall from Example I-2.5.39 that C is the intersection of a family $(C_k)_{k\in\mathbb{Z}_{>0}}$ of sets for which each of the sets C_k is a collection of 2^k disjoint closed intervals of length 3^{-k}. Therefore, since the total lengths of the intervals comprising C_k (i.e., $\lim_{k\to\infty} 2^k 3^{-k}$) goes to zero as k goes to infinity, there exists $N \in \mathbb{Z}_{>0}$ such that we can cover C_N with a finite family, say $((a_j, b_j))_{j\in\{1,\dots,2^N\}}$, of disjoint open intervals for which

$$\sum_{j=1}^{2^N} |b_j - a_j| < \delta.$$

Now note that since C is closed, $[0, 1] \setminus C$ is open, and so, by Proposition I-2.5.6, is a countable union of open intervals. By construction, f_C is constant on each of these open intervals. Since $C \subseteq C_N$, it follows that $[0, 1] \setminus C_N \subseteq [0, 1] \setminus C$ and so

$[0,1]\setminus C_N$ is itself a countable (in fact, finite) collection of open intervals, each having the property that f_C is constant when restricted to it. Since f_C is monotonically increasing and continuous, it then follows that

$$\sum_{j=1}^{2^N} |f_C(b_j) - f_C(a_j)| = f(1) - f(0) = 1.$$

Since this conclusion is independent of $\delta \in \mathbb{R}_{>0}$, we therefore are forced to deduce that f_C is not absolutely continuous. •

This example illustrates that there is a "gap" between the notion of absolute continuity and the notion of continuous and bounded variation. It is perhaps not immediately clear why we should care about this. In order to clarify this, we have the following definition.

2.9.26 Definition (Singular function) Let $I \subseteq \mathbb{R}$ be an interval. A function $f\colon I \to \mathbb{R}$ is ***singular*** if

(i) it is continuous,

(ii) it is if of locally bounded variation, and

(iii) $f'(x) = 0$ for almost every $x \in I$. •

As we can see from our discussion in Example I-3.2.27, the Cantor function is singular. The following result explains the importance of singular functions.

2.9.27 Theorem (Lebesgue decomposition of a function of bounded variation) *If* $\mathrm{I} \subseteq \mathbb{R}$ *is an interval and if* $\mathrm{f}\colon \mathrm{I} \to \mathbb{R}$ *has locally bounded variation, then* $\mathrm{f} = \mathrm{f}_{\mathrm{ac}} + \mathrm{f}_{\mathrm{sing}} + \mathrm{f}_{\mathrm{jump}}$*, where* $\mathrm{f}_{\mathrm{abs}}$ *is locally absolutely continuous,* $\mathrm{f}_{\mathrm{sing}}$ *is a singular function, and* $\mathrm{f}_{\mathrm{jump}}$ *is a saltus function.*

Proof It is sufficient to take $I = [a, b]$ (by definition of locally bounded variation) and assume that f is monotonically increasing (by Theorem I-3.3.3(ii)). By Proposition I-3.3.22, we can write $f = f_{\mathrm{cont}} + f_{\mathrm{jump}}$ for a monotonically increasing continuous function f_{cont} (necessarily of bounded variation by Theorem I-3.3.3(ii)) and for a saltus function f_{jump}. Define

$$f_{\mathrm{abs}}(x) = \int_a^x f'_{\mathrm{cont}}(x)\,\mathrm{d}x, \quad f_{\mathrm{sing}}(x) = f_{\mathrm{cont}}(x) - f_{\mathrm{abs}}(x)$$

and note that f_{abs} is absolutely continuous by Theorem 2.9.33(i) below. Also, for almost every $x \in [a, b]$,

$$f'_{\mathrm{sing}}(x) = f'_{\mathrm{cont}}(x) - \frac{\mathrm{d}}{\mathrm{d}x}\int_a^x f'_{\mathrm{cont}}(x)\,\mathrm{d}x = 0$$

by Theorem 2.9.33(ii) below. ■

Locally absolutely continuous functions, by virtue of also being of locally bounded variation, are almost everywhere differentiable. The next result we state provides us with a large collection of locally absolutely continuous functions based on their differentiability. The result also strengthens Proposition I-3.3.14 where the hypotheses are the same, but here we draw the sharper conclusion of absolute continuity, not just bounded variation.

2.9.28 Proposition (Nice differentiable functions are locally absolutely continuous) *If* $\mathrm{I} \subseteq \mathbb{R}$ *is an interval and if* $\mathrm{f}\colon \mathrm{I} \to \mathbb{R}$ *is a differentiable function having the property that* f' *is locally bounded, then* f *is locally absolutely continuous.*

Proof Clearly it suffices to consider the case where $I = [a, b]$. Let $M \in \mathbb{R}_{>0}$ have the property that $|f'(x)| < M$ for each $x \in [a, b]$. Then, for $\epsilon \in \mathbb{R}_{>0}$ take $\delta = \frac{\epsilon}{M}$ and note that, if $((a_j, b_j))_{j\in\{1,\dots,k\}}$ is a finite family of disjoint open intervals satisfying

$$\sum_{j=1}^{k} |b_j - a_j| < \epsilon,$$

then

$$\sum_{j=1}^{k} |f(b_j) - f(a_j)| = \sum_{j=1}^{k} |f'(c_j)(b_j - a_j)| < \epsilon,$$

where $c_j \in (a_j, b_j)$, $j \in \{1, \dots, k\}$, are as asserted by the Mean Value Theorem. ■

The boundedness of the derivative in the preceding result is essential, as Example I-3.3.15 shows.

Let us next consider how absolutely continuous functions behave under the standard algebraic operations on functions. First we consider the standard algebraic operations.

2.9.29 Proposition (Addition and multiplication, and local absolute continuity) *Let* $\mathrm{I} \subseteq \mathbb{R}$ *be an interval and let* $\mathrm{f}, \mathrm{g}\colon \mathrm{I} \to \mathbb{R}$ *be locally absolutely continuous. Then the following statements hold:*

(i) $\mathrm{f} + \mathrm{g}$ *is locally absolutely continuous;*

(ii) fg *is locally absolutely continuous;*

(iii) if additionally there exists $\alpha \in \mathbb{R}_{>0}$ *such that* $|\mathrm{g}(\mathrm{x})| \geq \alpha$ *for all* $\mathrm{x} \in \mathrm{I}$*, then* $\frac{\mathrm{f}}{\mathrm{g}}$ *is locally absolutely continuous.*

Proof Throughout the proof we suppose, without loss of generality, that $I = [a, b]$ is a compact interval.

(i) For $\epsilon \in \mathbb{R}_{>0}$ let $\delta \in \mathbb{R}_{>0}$ have the property that, if $((a_j, b_j))_{j\in\{1,\dots,k\}}$ is a finite family of disjoint open intervals satisfying

$$\sum_{j=1}^{k} |b_j - a_j| < \delta,$$

then

$$\sum_{j=1}^{k} |f(b_j) - f(a_j)| < \tfrac{\epsilon}{2}, \quad \sum_{j=1}^{k} |g(b_j) - g(a_j)| < \tfrac{\epsilon}{2}.$$

Then, again for any finite collection $((a_j, b_j))_{j\in\{1,\dots,k\}}$ of disjoint open intervals satisfying

$$\sum_{j=1}^{k} |b_j - a_j| < \delta,$$

we have

$$\sum_{j=1}^{k} |(f+g)(b_j) - (f+g)(a_j)| \leq \sum_{j=1}^{k} |f(b_j) - f(a_j)| + \sum_{j=1}^{k} |g(b_j) - g(a_j)| < \epsilon,$$

using the triangle inequality.

(ii) Let

$$M_f = \sup\{|f(x)| \mid x \in [a,b]\}, \quad M_g = \sup\{|g(x)| \mid x \in [a,b]\}.$$

Let $\epsilon \in \mathbb{R}_{>0}$ and let $\delta \in \mathbb{R}_{>0}$ have the property that, if $((a_j, b_j))_{j\in\{1,\dots,k\}}$ is a finite family of disjoint open intervals satisfying

$$\sum_{j=1}^{k} |b_j - a_j| < \delta,$$

then

$$\sum_{j=1}^{k} |f(b_j) - f(a_j)| < \frac{\epsilon}{2M_f}, \quad \sum_{j=1}^{k} |g(b_j) - g(a_j)| < \frac{\epsilon}{2M_g}.$$

Then, for any finite collection $((a_j, b_j))_{j\in\{1,\dots,k\}}$ of disjoint open intervals satisfying

$$\sum_{j=1}^{k} |b_j - a_j| < \delta,$$

we compute

$$\begin{aligned}\sum_{j=1}^{k} |f(b_j)g(b_j) - f(a_j)g(a_j)| &\leq \sum_{j=1}^{k} |f(b_j)g(b_j) - f(a_j)g(b_j)| \\ &\quad + \sum_{j=1}^{k} |f(a_j)g(b_j) - f(a_j)g(a_j)| \\ &\leq \sum_{j=1}^{k} M_g |f(b_j) - f(a_j)| + \sum_{j=1}^{k} M_f |g(b_j) - g(a_j)| \\ &< \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon,\end{aligned}$$

giving the result.

(iii) Let $\epsilon \in \mathbb{R}_{>0}$ and let $\delta \in \mathbb{R}_{>0}$ have the property that, if $((a_j, b_j))_{j\in\{1,\dots,k\}}$ is a finite collection of disjoint open intervals satisfying

$$\sum_{j=1}^{k} |b_j - a_j| < \delta,$$

then

$$\sum_{j=1}^{k} |f(b_j) - f(a_j)| < \alpha^2 \epsilon.$$

Then, for any finite collection $((a_j, b_j))_{j\in\{1,\dots,k\}}$ of disjoint open intervals satisfying

$$\sum_{j=1}^{k} |b_j - a_j| < \delta,$$

we compute

$$\sum_{j=1}^{k} \left|\frac{1}{g(b_j)} - \frac{1}{g(a_j)}\right| = \sum_{j=1}^{k} \left|\frac{g(a_j) - g(b_j)}{g(b_j)g(a_j)}\right| \le \sum_{j=1}^{k} \left|\frac{g(b_j) - g(a_j)}{\alpha^2}\right| < \epsilon.$$

Thus $\frac{1}{g}$ is locally absolutely continuous, and this part of the result follows from part (ii). ■

Next let us show that local absolute continuity for a function on an interval can be determined by breaking the interval into parts, and determining local absolute continuity on each.

2.9.30 Proposition (Local absolute continuity on disjoint subintervals) *Let* $\mathrm{I} \subseteq \mathbb{R}$ *be an interval and let* $\mathrm{I} = \mathrm{I}_1 \cup \mathrm{I}_2$*, where* $\mathrm{I}_1 \cap \mathrm{I}_2 = \{\mathrm{c}\}$*, where* c *is the right endpoint of* I_1 *and the left endpoint of* I_2*. Then* f *is locally absolutely continuous if and only if* $\mathrm{f}|\mathrm{I}_1$ *and* $\mathrm{f}|\mathrm{I}_2$ *are locally absolutely continuous.*

Proof It suffices to consider the case where $I = [a, c]$, $I_1 = [a, c]$, and $I_2 = [c, b]$.

First suppose that f is absolutely continuous and, for $\epsilon \in \mathbb{R}_{>0}$, choose $\delta \in \mathbb{R}_{>0}$ such that, if $((a_j, b_j))_{j\in\{1,\dots,k\}}$ is a finite family of disjoint open intervals satisfying

$$\sum_{j=1}^{k} |b_j - a_j| < 2\delta.$$

then

$$\sum_{j=1}^{k} |f(b_j) - f(a_j)| < \epsilon.$$

Then let $((a_j, c_j))_{j\in\{1,\dots,k_1\}}$ and $((d_j, b_j))_{j\in\{1,\dots,k_2\}}$ be finite families of disjoint open subintervals of $[a, c]$ and $[a, c]$, respectively, satisfying

$$\sum_{j=1}^{k_1} |c_j - a_j| < \delta, \quad \sum_{j=1}^{k_2} |b_j - d_j| < \delta.$$

Then $((a_j, c_j))_{j\in\{1,\dots,k_1\}} \cup ((d_j, b_j))_{j\in\{1,\dots,k_2\}}$ is a finite collection of disjoint open subintervals of $[a, b]$ satisfying

$$\sum_{j=1}^{k_1} |c_j - a_j| + \sum_{j=1}^{k_2} |b_j - d_j| < 2\delta.$$

Therefore,

$$\sum_{j=1}^{k_1} |f(c_j) - f(a_j)| + \sum_{j=1}^{k_2} |f(b_j) - f(d_j)| < \epsilon,$$

implying that

$$\sum_{j=1}^{k_1} |f(c_j) - f(a_j)| < \epsilon, \quad \sum_{j=1}^{k_2} |f(b_j) - f(d_j)| < \epsilon.$$

Thus $f|[a, c]$ and $f|[c, b]$ are absolutely continuous.

Now suppose that $f|[a, c]$ and $f|[c, b]$ are absolutely continuous. Let $\epsilon \in \mathbb{R}_{>0}$ and let $\delta \in \mathbb{R}_{>0}$ be chosen such that, if $((a_j, c_j))_{j\in\{1,\dots,k_1\}}$ and $((d_j, b_j))_{j\in\{1,\dots,k_2\}}$ are finite collections of disjoint open subintervals of $[a, c]$ and $[c, b]$, respectively, satisfying

$$\sum_{j=1}^{k_1} |c_j - a_j| < \delta, \quad \sum_{j=1}^{k_2} |b_j - d_j| < \delta,$$

then

$$\sum_{j=1}^{k_1} |f(c_j) - f(a_j)| < \tfrac{\epsilon}{2}, \quad \sum_{j=1}^{k_2} |f(b_j) - f(d_j)| < \tfrac{\epsilon}{2}.$$

Now let $((a_j, b_j))_{j\in\{1,\dots,k\}}$ be a finite collection of disjoint subintervals of $[a, b]$ satisfying

$$\sum_{j=1}^{k} |b_j - a_j| < \delta.$$

If $c \in (a_{j_0}, b_{j_0})$ for some $j_0 \in \{1, \dots, k\}$, then define the collection of disjoint open intervals

$$(((a_j, b_j))_{j\in\{1,\dots,k\}} \setminus ((a_{j_0}, b_{j_0}))) \cup ((a_{j_0}, c), (c, b_{j_0})),$$

i.e., split the interval containing c into two intervals. Denote this collection of disjoint open intervals by $((\tilde{a}_j, \tilde{b}_j))_{j\in\{1,\dots,\tilde{k}\}}$. If c is not contained in any of the intervals $((a_j, b_j))_{j\in\{1,\dots,k\}}$, then denote $((\tilde{a}_j, \tilde{b}_j))_{j\in\{1,\dots,\tilde{k}\}} = ((a_j, b_j))_{j\in\{1,\dots,k\}}$. Note that

$$\sum_{j=1}^{\tilde{k}} |\tilde{b}_j - \tilde{a}_j| < \delta.$$

This new collection of disjoint open intervals is then the union of two collections of disjoint open intervals, $((\tilde{a}_j, \tilde{c}_j))_{j\in\{1,\dots,k_1\}}$ and $((\tilde{d}_j, \tilde{b}_j))_{j\in\{1,\dots,k_2\}}$, the first being subintervals of $[a, c]$ and the second being subintervals of $[c, b]$. These collections satisfy

$$\sum_{j=1}^{k_1} |\tilde{c}_j - \tilde{a}_j| < \delta, \quad \sum_{j=1}^{k_2} |\tilde{b}_j - \tilde{d}_j| < \delta,$$

and so we have

$$\sum_{j=1}^{k}|f(b_j)-f(a_j)| \le \sum_{j=1}^{\tilde{k}}|f(\tilde{b}_j)-f(\tilde{a}_j)| = \sum_{j=1}^{k_1}|f(\tilde{c}_j)-f(\tilde{a}_j)| + \sum_{j=1}^{k_2}|f(\tilde{b}_j)-f(\tilde{d}_j)| < \epsilon,$$

which shows that f is absolutely continuous. ■

Next we show that one of the standard operations on functions does *not* respect absolute continuity.

2.9.31 Example (Compositions of locally absolutely continuous functions need not be locally absolutely continuous) In Example I-3.3.16 we gave two functions of bounded variation whose composition was not a function of bounded variation. In fact, the functions we used were not only of bounded variation, but absolutely continuous. These functions, therefore, show that the composition of absolutely continuous functions may not be an absolutely continuous function.

Let us show that the functions in question are, in fact, absolutely continuous. Recall that the functions $f, g\colon [-1,1] \to \mathbb{R}$ were given by $f(x) = x^{1/3}$ and by

$$g(x) = \begin{cases} x^3(\sin\frac{1}{x})^3, & x \neq 0, \\ 0, & x = 0. \end{cases}$$

That g is absolutely continuous follows from Proposition 2.9.28 since we showed in Example I-3.3.16 that g was of class C^1. It then only remains to show that f is absolutely continuous. Let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta = \frac{\epsilon^3}{4}$. Now let $((a_j, b_j))_{j\in\{1,\dots,k\}}$ be a finite collection of open intervals satisfying

$$\sum_{j=1}^{k}|b_j - a_j| < \delta.$$

Let $\ell \le 2$ and let $[a,b] \subseteq [-1,1]$ be an interval of length ℓ. One can easily see that, if one fixes the length of the interval at ℓ, then the quantity $|f(b)-f(a)|$ is maximum when one takes $a = -\frac{\ell}{2}$ and $b = \frac{\ell}{2}$. From this it follows that

$$\sum_{j=1}^{k}|f(b_j)-f(a_j)| < |f(\tfrac{\delta}{2}) - f(-\tfrac{\delta}{2})| = \epsilon.$$

Thus f is absolutely continuous, as desired. •

2.9.7 The Fundamental Theorem of Calculus for the Lebesgue integral

In this section we explore the Fundamental Theorem of Calculus that is associated with the Lebesgue integral. As we shall see, it is here that the notion of absolute continuity comes up in a natural way.

Before we state the main result,

2.9.32 Lemma (Locally absolutely continuous functions with a.e. zero derivative) *Let* $\mathrm{I} \subseteq \mathbb{R}$ *be an interval and let* $\mathrm{f}\colon \mathrm{I} \to \mathbb{R}$ *be locally absolutely continuous and having the property that the set*

$$\{\mathrm{x} \in \mathrm{I} \mid \mathrm{f} \textit{ is not differentiable at } \mathrm{x}\} \cap \{\mathrm{x} \in \mathrm{I} \mid \mathrm{f}'(\mathrm{x}) \neq 0\}$$

has measure zero. Then there exists $\mathrm{c} \in \mathbb{R}$ *such that* $\mathrm{f}(\mathrm{x}) = \mathrm{c}$ *for all* $\mathrm{x} \in \mathrm{I}$.

Proof Consider an interval $[a, b] \subseteq I$. Let

$$E = \{x \in I \mid f \text{ is not differentiable at } x\} \cap \{x \in I \mid f'(x) \neq 0\}.$$

For $\epsilon \in \mathbb{R}_{>0}$ choose $\eta \in \mathbb{R}_{>0}$ such that, if $((a_j, b_j))_{j\in\{1,\dots,k\}}$ is a finite collection of disjoint intervals having the property that

$$\sum_{j=1}^{k} |b_j - a_j| < \eta,$$

then

$$\sum_{j=1}^{k} |f(b_j) - f(a_j)| < \epsilon.$$

Let $((c_\alpha, d_\alpha))_{\alpha\in A}$ be a countable collection of open intervals satisfying

$$E \subseteq \bigcup_{\alpha\in A} (c_\alpha, d_\alpha)$$

and

$$\sum_{\alpha\in A} |d_\alpha - c_\alpha| < \eta.$$

Now define $\delta\colon [a, b] \to \mathbb{R}_{>0}$ according to the following:

1. if $x \in E$ take $\delta(x)$ such that $\mathsf{B}(\delta(x), x) \cap [a, b] \subseteq (c_\alpha, d_\alpha)$ for some $\alpha \in A$;
2. if $x \notin E$ take $\delta(x)$ such that $|f(y) - f(x)| < \epsilon|y - x|$ for $y \in \mathsf{B}(\delta(x), x) \cap [a, b]$.

Now let $((c_1, I_1), \dots, (c_k, I_k))$ be a δ-fine tagged partition and write $\{1, \dots, k\} = K_1 \mathring{\cup} K_2$ where

$$K_1 = \{j \in \{1, \dots, k\} \mid c_j \in E\}, \quad K_2 = \{j \in \{1, \dots, k\} \mid c_j \notin E\}.$$

We then compute, denoting $\mathrm{EP}(P) = (x_0, x_1, \dots, x_k)$,

$$\begin{aligned} |f(b) - f(a)| &= \Big|\sum_{j=1}^{k} (f(x_j) - f(x_{j-1}))\Big| \\ &\leq \sum_{j\in K_1} |f(x_j) - f(x_{j-1})| + \sum_{j\in K_2} |f(x_j) - f(x_{j-1})| \\ &\leq \epsilon + \sum_{j\in K_2} \epsilon(x_j - x_{j-1}) \leq \epsilon(1 + b - a). \end{aligned}$$

This shows that $|f(b) - f(a)|$ can be made arbitrarily small, and so gives the result since a and b are arbitrary. ■

The main result in this section is the following.

2.9.33 Theorem (The Fundamental Theorem of Calculus for the Lebesgue integral) *For an interval $I \subseteq \mathbb{R}$ the following statements hold:*

(i) a function $F\colon I \to \mathbb{R}$ defined on an interval I is locally absolutely continuous if and only if there exists $f \in L^{(1)}_{\mathrm{loc}}(I;\mathbb{R})$ and $x_0 \in I$ such that

$$F(x) = F(x_0) + \int_{x_0}^{x} f(\xi)\,d\xi,$$

where we adopt the convention that if $x < x_0$ we have

$$\int_{x_0}^{x} g(\xi)\,d\xi = -\int_{x}^{x_0} g(\xi)\,d\xi;$$

(ii) if $x_0 \in I$, if $f \in L^{(1)}_{\mathrm{loc}}(I;\mathbb{R})$, and define $F\colon I \to \mathbb{R}$ by

$$F(x) = \int_{x_0}^{x} f(\xi)\,d\xi,$$

then F is differentiable for almost every $x \in I$ and $F'(x) = f(x)$ for almost every $x \in I$.

Proof (i) We first consider the case when I is compact: $I = [a,b]$.

First suppose that

$$F(x) = F(x_0) + \int_{x_0}^{x} f(\xi)\,d\xi$$

for $x_0 \in [a,b]$ and $f \in L^{(1)}([a,b];\mathbb{R})$. Note that, by Proposition 2.7.22, we have

$$\begin{aligned} F(x) &= F(x_0) + \int_a^x f(\xi)\,d\xi - \int_a^{x_0} f(\xi)\,d\xi \\ &= F(x_0) + \int_a^x f(\xi)\,d\xi + (F(a) - F(x_0) = F(a) + \int_a^x f(\xi)\,d\xi. \end{aligned}$$

Thus we can take $x_0 = a$ without loss of generality. First assume that f is nonnegative-valued. For $k \in \mathbb{Z}_{>0}$ define

$$f_k(x) = \begin{cases} f(x), & f(x) \le k \\ k, & \text{otherwise.} \end{cases}$$

Note that f_k is bounded and that for each $x \in [a,b]$ we have $\lim_{k\to\infty} f_k(t) = f(t)$. Therefore, the Monotone Convergence Theorem asserts that

$$\lim_{k\to\infty} \int_a^b (f(x) - f_k(x))\,dx = 0.$$

Now let $\epsilon \in \mathbb{R}_{>0}$. Choose $N \in \mathbb{Z}_{>0}$ such that

$$\int_a^b (f(x) - f_k(x))\,dx < \frac{\epsilon}{2}, \qquad k \ge N.$$

Letting $\delta = \frac{\epsilon}{2N}$ and letting $((a_j, b_j))_{j\in\{1,\dots,n\}}$ be any finite family of nonoverlapping intervals in $[a, b]$ satisfying

$$\sum_{j=1}^{n} |b_j - a_j| < \delta,$$

we have

$$\int_A f(x)\,\mathrm{d}x = \int_A (f(x) - f_N(x))\,\mathrm{d}x + \int_A f_N(x)\,\mathrm{d}x \le \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

where A denotes the union of the intervals $((a_j, b_j))_{j\in\{1,\dots,n\}}$. Note that since f is nonnegative, it follows that F is monotonically increasing. Thus

$$\sum_{j=1}^{n} |F(b_j) - F(a_j)| = \sum_{j=1}^{n} (F(b_j) - F(a_j)).$$

Given the definition of F we thus have

$$\sum_{j=1}^{n} |F(b_j) - F(a_j)| = \int_A f(x)\,\mathrm{d}x < \epsilon,$$

and we conclude that F is absolutely continuous.

If f is not nonnegative-valued, then we write $f = f_+ - f_-$ where f_+ and f_- are nonnegative. Our arguments above show that the functions

$$x \mapsto \int_a^x f_+(\xi)\,\mathrm{d}\xi, \quad x \mapsto \int_a^x f_-(\xi)\,\mathrm{d}\xi$$

are absolutely continuous. Therefore, since

$$F(x) = F(a) + \int_a^x f_+(\xi)\,\mathrm{d}\xi - \int_a^x f_-(\xi)\,\mathrm{d}\xi,$$

it follows that F is the sum of three absolutely continuous functions (a constant function is trivially absolutely continuous) and so F is itself absolutely continuous by Proposition 2.9.29.

Now suppose that F is absolutely continuous, and so of bounded variation by Proposition 2.9.24. Now, by part (I-ii) of Theorem I-3.3.3, write $F = F_+ - F_-$ for monotonic functions F_+ and F_-. By part (I-vi) of Theorem I-3.3.3 the derivative of F exists almost everywhere and we then have

$$F'(x) = F'_+(x) - F'_-(x) \quad \Longrightarrow \quad |F'(x)| \le |F'_+(x)| + |F'_-(x)|$$

for almost every $x \in [a, b]$. Therefore we have

$$\int_a^b |F'(x)|\,\mathrm{d}x \le F_+(b) + F_-(b) - F_+(a) - F_-(a),$$

implying that $F' \in \mathsf{L}^{(1)}([a, b]; \mathbb{R})$. Note that the function

$$x \mapsto \int_a^x F'(\xi)\,\mathrm{d}\xi$$

is now absolutely continuous by our arguments from the first part of the proof, so that the function

$$x \mapsto F(x) - \int_a^x F'(\xi)\,\mathrm{d}\xi$$

is also absolutely continuous by Proposition 2.9.29. This function also has derivative zero, and the result now follows by Lemma 2.9.32.

Now suppose that I is an arbitrary interval. We first suppose that

$$F(x) = F(x_0) + \int_{x_0}^x f(\xi)\,\mathrm{d}\xi$$

for some $x_0 \in I$ and some locally integrable function f. Let $[a, b] \subseteq I$ be a compact subinterval. As we determined in the first part of the proof, we have

$$F(x) = F(a) + \int_a^x f(\xi)\,\mathrm{d}\xi,$$

from which we conclude that $F|[a, b]$ is absolutely continuous, since we proved have already proved the theorem for compact intervals. It then follows that F is locally absolutely continuous since this can be done for any compact subinterval. Conversely, suppose that F is locally absolutely continuous and let $x_0 \in I$. Let $x \in I$, supposing that $x > x_0$. Note that, since $F|[x_0, x]$ is absolutely continuous, the first part of the proof allows us to conclude that

$$F(x) = F(x_0) + \int_{x_0}^x f(\xi)\,\mathrm{d}\xi.$$

If $x < x_0$ we have that $F|[x, x_0]$ is absolutely continuous and so we can write

$$F(x_0) = F(x) + \int_x^{x_0} f(\xi)\,\mathrm{d}\xi,$$

and the theorem follows by a rearrangement of this equation, using the stated convention for integrals whose lower limit exceeds the upper limit.

(ii) We first prove a technical lemma from which this part of the theorem will follow.

1 **Lemma** *If* $\mathrm{A} \subseteq \mathbb{R}$ *then*

$$\lim_{\beta \downarrow 0} \frac{\lambda(\mathrm{A} \cap (\mathrm{x}, \mathrm{x} + \beta))}{\beta} = \lim_{\alpha \downarrow 0} \frac{\lambda(\mathrm{A} \cap (\mathrm{x} - \alpha, \mathrm{x}))}{\alpha} = \lim_{\alpha, \beta \downarrow 0} \frac{\lambda(\mathrm{A} \cap (\mathrm{x} - \alpha, \mathrm{x} + \beta))}{\alpha + \beta} = 1$$

for almost every $\mathrm{x} \in \mathrm{A}$. *If we additionally have* $\mathrm{A} \in \mathscr{L}(\mathbb{R})$ *then the above limits are equal to zero for almost every* $\mathrm{x} \in \mathbb{R} \setminus \mathrm{A}$.

Proof First suppose that A is bounded so that $\lambda^*(A) < \infty$. By definition of Lebesgue outer measure, for $k \in \mathbb{Z}_{>0}$ there exists a countable collection $((a_{k,j}, b_{k,j}))_{j \in \mathbb{Z}_{>0}}$ of open intervals such that

$$\sum_{j=1}^{\infty} |b_{k,j} - a_{k,j}| - 2^{-k} < \lambda^*(A).$$

If we define $U''_k = \cup_{j=1}^\infty (a_{k,j}, b_{k,j})$ then we have

$$\lambda(U''_k) - 2^{-k} \leq \sum_{j=1}^\infty |b_{k,j} - a_{k,j}| - 2^{-k} < \lambda^*(A).$$

Then define $U'_m = \cap_{k=1}^m U''_k$ so that $U'_{m+1} \subseteq U'_m$, $m \in \mathbb{Z}_{>0}$, and

$$\lambda(U'_m) - 2^{-m} \leq \lambda(U''_m) - 2^{-m} < \lambda^*(A).$$

Finally, let (a, b) be such that $A \subseteq (a, b)$ and define $U_k = U'_k \cap (a, b)$, $k \in \mathbb{Z}_{>0}$. Then $A \subseteq \cap_{k \in \mathbb{Z}_{>0}} U_k$, $U_{k+1} \subseteq U_k$, $k \in \mathbb{Z}_{>0}$, and $\lambda(U_k) - 2^{-k} < \lambda^*(A)$, $k \in \mathbb{Z}_{>0}$.

Now define $f_k \colon \mathbb{R} \to \mathbb{R}$, $k \in \mathbb{Z}_{>0}$, and $f \colon \mathbb{R} \to \mathbb{R}$ by

$$f_k(x) = \lambda(U_k \cap (a, x)), \quad f(x) = \lambda(A \cap (a, x)).$$

Since U_k is open for $k \in \mathbb{Z}_{>0}$, if $x \in U_k$ and if $\epsilon \in \mathbb{R}_{>0}$ is such that $(x - \epsilon, x + \epsilon) \subseteq U_k$ we have

$$\frac{f_k(x + \epsilon) - f_k(x)}{\epsilon} = \frac{f_k(x) - f_k(x - \epsilon)}{\epsilon} = 1.$$

Thus $f_k|U_k$ is differentiable with derivative 1.

Let $x_1, x_2 \in \mathbb{R}$ with $a \leq x_1 < x_2$. Then, for each $k \in \mathbb{Z}_{>0}$,

$$\begin{aligned} f_k(x_2) - f(x_2) - (f_k(x_1) - f(x_1)) &= \lambda(U_k \cap [x_1, x_2)) - \lambda(A \cap (a, x_2)) + \lambda(A \cap (a, x_1)) \\ &= \lambda(U_k \cap [x_1, x_2)) - \lambda(A \cap [x_1, x_2)) \geq 0 \end{aligned}$$

by monotonicity of Lebesgue measure. This shows that the function $f_k - f$ is monotonically increasing for each $k \in \mathbb{Z}_{>0}$. We also have

$$f_k(b) - f(b) = \lambda(U_k) - \lambda(A) < 2^{-k}$$

which gives

$$\sum_{k=1}^\infty (f_k(x) - f(x)) \leq \sum_{k=1}^\infty (f_k(b) - f(b)) \leq \sum_{k=1}^\infty 2^{-k} < \infty,$$

using Example I-2.4.2–1. If we define

$$g(x) = \sum_{k=1}^\infty (f_k(x) - f(x)),$$

then, by Theorems I-3.2.26 and I-3.6.25, g is almost everywhere differentiable and

$$g'(x) = \sum_{k=1}^\infty (f'_k(x) - f'(x))$$

for almost every $x \in (a, b)$. Since g is monotonically increasing, $g'(x)$ is finite for almost every $x \in [a, b]$. This gives

$$\sum_{k=1}^\infty (f'_k(x) - f'(x)) < \infty$$

for almost every $x \in (a, b)$. Since $f_k'(x) - f'(x) \geq 0$ for almost every $x \in (a, b)$ we must have

$$\lim_{k\to\infty} f_k'(x) - f'(x) = 0$$

for almost every $x \in (a, b)$. Let $N \subseteq (a, b)$ be the set of points on which the above limit does not hold, so $\lambda^*(N) = 0$. Let $x \in (\cap_{k\in\mathbb{Z}_{>0}} U_k) - N$. Then $f_k'(x) = 1$ for every $k \in \mathbb{Z}_{>0}$ and so $f'(x) = 1$. Thus $f'(x) = 1$ for $x \in A - N$, giving the first assertion of the lemma in the case when A is bounded. If A is not bounded then we can write A as a countable union of bounded sets: $A = \cup_{j\in\mathbb{Z}_{>0}} A_j$. Let $N_j \subseteq A_j$ be the subset of A_j where the limits in the first assertion of the theorem do not have the value 1. Then the limits in the first assertion of the theorem hold for all $x \in A \setminus \cup_{j\in\mathbb{Z}_{>0}} N_j$. Since $\cup_{j\in\mathbb{Z}_{>0}} N_j$ has measure zero by Exercise I-2.5.11, the first part of the theorem is proved.

For the second assertion, if A is measurable then we have

$$\alpha + \beta = \lambda((x - \alpha, x + \beta)) = \lambda(A \cap (x - \alpha, x + \beta)) + \lambda((\mathbb{R} \setminus A)(x - \alpha, x + \beta)).$$

Thus

$$1 = \frac{\lambda(A \cap (x - \alpha, x + \beta))}{\alpha + \beta} + \frac{\lambda((\mathbb{R} \setminus A)(x - \alpha, x + \beta))}{\alpha + \beta},$$

and taking the limit as α and β decrease to zero gives

$$\lim_{\alpha,\beta\downarrow 0} \frac{\lambda((\mathbb{R} \setminus A)(x - \alpha, x + \beta))}{\alpha + \beta},$$

using the fact that the first part of the proof has been proved. ▼

Proceeding with the proof of the theorem, first consider the case when $I = [a, b]$ and

$$F(x) = \int_a^x f(\xi)\,\mathrm{d}\xi;$$

the lower limit can be taken to be a as we saw in the first part of the proof. We first consider the case where f is a finite nonnegative simple function,

$$f(x) = \sum_{j=1}^{k} a_j \chi_{A_j}(x).$$

Then, by linearity of the integral,

$$F(x) = \int_a^x f(\xi)\,\mathrm{d}\xi = \sum_{j=1}^{k} a_j \lambda(A_j \cap (a, x)).$$

By the lemma it follows that F is differentiable for almost every $x \in (a, b)$ and that $F'(x) = f(x)$ for almost every $x \in (a, b)$.

Now suppose that f is a nonnegative simple function and let $(g_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of nonnegative simple functions as in part (ii) of Proposition 2.6.39. For $j \in \mathbb{Z}_{>0}$ define

$$G_j(x) = \int_a^x g_j(\xi)\,\mathrm{d}\xi.$$

By the Monotone Convergence Theorem,

$$\begin{aligned} F(x) &= \int_a^x f(\xi)\,d\xi = \lim_{j\to\infty}\int_a^x g_j(\xi)\,d\xi = \lim_{j\to\infty} G_j(x) \\ &= G_1(x) + \sum_{j=1}^{\infty}(G_{j+1}(x) - G_j(x)) \end{aligned}$$

for every $x \in [a, b]$. Note that for each $j \in \mathbb{Z}_{>0}$ the functions G_j and $G_{j+1} - G_j$ are monotonically increasing, being the indefinite integrals of nonnegative functions. Therefore, we can apply Theorem I-3.6.25 to arrive at the equality

$$F'(x) = G_1'(x) + \sum_{j=1}^{\infty}(G_{j+1}'(x) - G_j'(x)) = \lim_{j\to\infty} G_j'(x)$$

for almost every $x \in (a, b)$. Since the theorem has been proved for nonnegative simple functions, we have

$$\lim_{j\to\infty} G_j'(x) = \lim_{j\to\infty} g_j(x) = f(x)$$

for almost every $x \in (a, b)$. Therefore, $F'(x) = f(x)$ for almost every $x \in (a, b)$.

Now let I be an arbitrary interval with $x_0 \in I$ and

$$F(x) = \int_{x_0}^x f(\xi)\,d\xi.$$

Let $(I_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of bounded intervals all containing x_0 such that $I_j \subseteq I_{j+1}$, $j \in \mathbb{Z}_{>0}$, and such that $\cup_{j\in\mathbb{Z}_{>0}} I_j = I$ (make sure you understand why this is possible). By the arguments above, $F'(x) = f(x)$ for almost every $x \in I_j$ and for every $j \in \mathbb{Z}_{>0}$. Thus, if $N_j \subseteq I_j$ is the set of measure zero for which F' does not exist or, if it exists is not equal to $f(x)$, then $F'(x) = f(x)$ for all $x \in I \setminus \cup_{j\in\mathbb{Z}_{>0}} N_j$. Since $\lambda(\cup_{j\in\mathbb{Z}_{>0}} N_j) = 0$ by Exercise I-2.5.11, the theorem follows. ■

In Example I-3.4.31 we considered a collection of examples illustrating the Fundamental Theorem of Calculus for the Riemann integral. The examples where this version of the Fundamental Theorem applies still apply for the Lebesgue integral by virtue of Theorem 2.9.11. However, in Example I-3.4.31 we saw an instance of a differentiable function on $[0, 1]$ that is everywhere differentiable and with bounded derivative, but the derivative is not Riemann integrable. This example is more satisfactory with the Lebesgue integral.

2.9.34 Example (The Fundamental Theorem of Calculus for the Lebesgue integral) The reader should go back and carefully read the construction of Example I-3.4.31. The reader will see that the example is of a function $F\colon [0, 1] \to \mathbb{R}$ with the property that F is everywhere differentiable with a bounded derivative. However, F' is not Riemann integrable. By Proposition 2.9.28, however, F is absolutely continuous, and so F' is Lebesgue integrable by Theorem 2.9.33. •

One of the conclusions of the Fundamental Theorem of Calculus for the Lebesgue integral is that an absolutely continuous function is almost everywhere differentiable. As we saw in Proposition 2.9.24, absolutely continuous functions are continuous. One might speculate, then, that a characterisation of absolute continuity using continuity and the derivative might be possible. For example, here are some guesses, along with counterexamples.

1. *An absolutely continuous function is one that is continuous and differentiable almost everywhere*. This is false as seen by Example I-3.3.15.
2. *An absolutely continuous function is one that is continuous, differentiable almost everywhere, and with integrable derivative*. This is false by virtue of Example 2.9.25.
3. *An absolutely continuous function is one that is differentiable almost everywhere*. This is false by virtue of Example I-3.3.15.

However, there is the following result, which is sometimes enough to understand absolute continuity.

2.9.35 Theorem (A class of absolutely continuous functions) *If* $F\colon [a,b] \to \mathbb{R}$ *is*

(i) continuous,

(ii) differentiable at all but countable many points in $[a,b]$*, and*

(iii) the function

$$f(x) = \begin{cases} F'(x), & \textit{the derivative exists}, \\ 0, & \textit{otherwise}, \end{cases}$$

is in $\mathsf{L}^{(1)}([a,b];\mathbb{R})$*,*

then F *is absolutely continuous.*

Proof Our proof relies on the definition in Section II-1.10.2 of lower semicontinuous functions. Note that, by Proposition II-1.10.14, lower semicontinuous functions are Borel measurable. With this notion recalled, we have the following lemma.

1 Lemma *If* $f \in \mathsf{L}^{(1)}([a,b];\overline{\mathbb{R}})$ *then, for each* $\epsilon \in \mathbb{R}_{>0}$*, there exists a lower semicontinuous* $g \in \mathsf{L}^{(1)}([a,b];(-\infty,\infty])$ *such that* $f(x) \le g(x)$ *for every* $x \in [a,b]$ *and*

$$\int_a^b g(x)\,dx < \int_a^b f(x)\,dx + \epsilon.$$

Proof Let $\epsilon \in \mathbb{R}_{>0}$. We first consider the case when f is nonnegative-valued. We let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of simple functions as in part (ii) of Proposition 2.6.39. Then

$$f(x) = \lim_{j\to\infty} f_j(x) = f_1(x) + \sum_{j=1}^{\infty} (f_{j+1}(x) - f_j(x)). \tag{2.21}$$

Let $j \in \mathbb{Z}_{>0}$ and write

$$f_j = \sum_{k=1}^{m} a_k \chi_{A_k}, \quad f_{j+1} = \sum_{l=1}^{n} b_l \chi_{B_l}.$$

For each $l \in \{1, \dots, n\}$ write
$$B_l = \cup_{k=1}^n (A_k \cap B_l).$$
If $A_k \cap B_l \neq \emptyset$ then, on $A_k \cap B_l$ the value of $f_{j+1} - f_j$ is $b_l - a_k \in \mathbb{R}_{>0}$. Thus $(f_{j+1} - f_j)|B_l$ is a nonnegative simple function. Since this is true for every l it follows that $f_{j+1} - f_j$ is a nonnegative simple function. Thus, by (2.21), f is an infinite sum of nonnegative simple functions. Thus we write

$$f = \sum_{k=1}^{\infty} a_k \chi_{A_k}$$

where the numbers $a_k \in \mathbb{R}_{>0}$ and the sets A_k, $k \in \mathbb{Z}_{>0}$, are not related to those above. For $k \in \mathbb{Z}_{>0}$ let U_k be an open set such that $A_k \subseteq U_k$ and such that

$$\lambda(U_k) < \lambda(A_k) + \tfrac{\epsilon}{a_k 2^k}.$$

Then
$$\sum_{k=1}^{\infty} a_k \lambda(U_k) < \sum_{k=1}^{\infty} a_k \lambda(A_k) + \sum_{k=1}^{\infty} \frac{\epsilon}{2^k} = \sum_{k=1}^{\infty} a_k \lambda(A_k) + \epsilon,$$
where we use Example I-2.4.2–1. By Example II-1.10.16–3 each of the functions $a_k \chi_{U_k}$ is lower semicontinuous. Define

$$h_m(x) = \sum_{k=1}^{m} a_k \lambda(U_k)$$

and
$$h(x) = \sum_{k=1}^{\infty} a_k \lambda(U_k).$$

need result on addition of lower semicontinuous functions

Then h_m is lower semicontinuous by , and, since

$$h(x) = \sup\{h_m(x) \mid m \in \mathbb{Z}_{>0}\},$$

h is also lower semicontinuous by Proposition II-1.10.17. We then have

$$\int_a^b h(x)\,\mathrm{d}x < \int_a^b f(x)\,\mathrm{d}x + \epsilon$$

and $f(x) \leq h(x)$ for all $x \in [a, b]$.

Now suppose that $f \in \mathsf{L}^{(1)}([a, b]; \mathbb{R})$ and, for $k \in \mathbb{Z}_{>0}$, define

$$f_k(x) = \begin{cases} f(x), & f(x) > -k, \\ -k, & f(x) \leq -k. \end{cases}$$

By the Dominated Convergence Theorem,

$$\int_a^b f(x)\,\mathrm{d}x = \lim_{k\to\infty} \int_a^b f_k(x)\,\mathrm{d}x.$$

Now let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\int_a^b f_N(x)\,dx < \int_a^b f(x)\,dx - \frac{\epsilon}{2}.$$

Since $f_N + N\chi_{[a,b]}$ is nonnegative, from the first part of the proof there exists a lower semicontinuous function h such that $f_N(x) + N \leq h(x)$ for every $x \in [a,b]$ and such that

$$\int_a^b h(x)\,dx < \int_a^b (f_N(x) + N\chi_{[a,b]})\,dx + \frac{\epsilon}{2}.$$

Define $g = h - N\chi_{[a,b]}$. Then $f(x) \leq f_N(x) \leq g(x)$ for every $x \in [a,b]$ and

$$\int_a^b g(x)\,dx = \int_a^b (h(x) - N\chi_{[a,b]})\,dx < \int_a^b f_N(x)\,dx < \int_a^b f(x)\,dx - \frac{\epsilon}{2},$$

as desired. ▼

2 Lemma *Let* $h\colon [a,b] \to \mathbb{R}$ *be continuous and let* $C \subseteq [a,b]$ *be countable. If, for each* $x \in [a,b) - C$, *there exists* $r_x \in \mathbb{R}_{>0}$ *such that* $h(z) > h(x)$ *for each* $z \in (x, x + r_x)$, *then* h *is monotonically increasing.*

Proof Suppose that h is continuous and that $x_1, x_2 \in [a,b]$ satisfy $x_1 < x_2$ and $h(x_1) > h(x_2)$. For $y \in (h(x_2), h(x_1))$ define

$$x_y = \sup\{x \in [x_1, x_2] \mid h(x) > y\}.$$

Then there exists a sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ in $[x_1, x_2]$ such that $x_j \leq x_y$, $j \in \mathbb{Z}_{>0}$, and such that $\lim_{j\to\infty} x_j = x_y$. By continuity of h, $\lim_{j\to\infty} h(x_j) = y$, using Theorem I-3.1.3. We claim that, for any $r_y \in \mathbb{R}_{>0}$, there exists $z \in (x_y, x_y + r_y)$ such that $h(z) \leq h(x_y)$. Indeed, were this not so, then there would exist $z > x_y$ such that $h(z) > h(x_y) = y$, contradicting the definition of x_y. Since this construction can be made for every $y \in (h(x_2), h(x_1))$, this shows, therefore, that the complement to the set

$$\{x \in [a,b) \mid \text{ there exists } r_x \in \mathbb{R}_{>0} \text{ such that } h(z) > h(x) \text{ for each } z \in (x, x + r_x)\}$$

is not countable, which give the lemma. ▼

Proceeding with the proof, let $\epsilon \in \mathbb{R}_{>0}$. Denote by $C \subseteq [a,b]$ the countable subset at whose points F is not differentiable. By Lemma 1 let $h\colon [a,b] \to (-\infty, \infty]$ be lower semicontinuous and such that $f(t) \leq h(t)$ for $t \in [a,b] \setminus C$ and such that

$$\int_a^b h(x)\,dx < \int_a^b f(x)\,dx + \frac{\epsilon}{2}.$$

Then, if we define $g = h + \frac{\epsilon}{2(b-a)}$, then $f(t) < g(t)$ for $t \in [a,b] \setminus C$ and

$$\int_a^b g(x)\,dx < \int_a^b f(x)\,dx + \epsilon.$$

Let $G\colon [a,b] \to \mathbb{R}$ be defined by

$$G(x) = F(a) + \int_a^x g(\xi)\,\mathrm{d}\xi$$

Let $x \in [a,b)$. Since g is lower semicontinuous, for each $\eta \in \mathbb{R}_{>0}$ there exists $\delta \in \mathbb{R}_{>0}$ such that, if $x' \in [x, x+\delta]$, we have $g(x') > g(x) - \eta$. Then, for any $y \in [x, x+\delta]$ we have

$$\begin{aligned} G(y) - G(x) = F(a) + \int_a^y g(\xi)\,\mathrm{d}\xi - F(a) - \int_a^x g(\xi)\,\mathrm{d}\xi = \int_x^y g(\xi)\,\mathrm{d}\xi \\ > \int_x^y (g(x) - \eta)\,\mathrm{d}\xi = (g(x) - \eta)(y - x), \end{aligned}$$

or

$$\frac{G(y) - G(x)}{y - x} > g(x) - \eta.$$

This implies that

$$\liminf_{y \downarrow x} \frac{G(y) - G(x)}{y - x} \geq g(x)$$

for every $x \in [a,b)$. Therefore, if $x \in [a,b] \setminus C$ we have

$$\begin{aligned} &\liminf_{y \downarrow x} \frac{(G(y) - F(y)) - (G(x) - F(x))}{y - x} \\ &\qquad = \liminf_{y \downarrow x} \frac{G(y) - G(x)}{y - x} - \liminf_{y \downarrow x} \frac{F(y) - F(x)}{y - x} \geq g(x) - f(x) > 0. \end{aligned}$$

This implies that, if $x \in [a,b] \setminus C$, there exists $r_x \in \mathbb{R}_{>0}$ such that

$$\frac{(G(y) - F(y)) - (G(x) - F(x))}{y - x} > 0$$

for $y \in (x, x + r_x)$. Since $y - x > 0$ for $y \in (x, x + r_x)$ this implies that

$$(G(y) - F(y)) - (G(x) - F(x)) > 0, \qquad y \in (x, x + r_x).$$

By Lemma 2 this implies that $G - F$ is nondecreasing. Therefore, since $G(a) = F(a)$ it follows that $F(x) \leq G(x)$ for $x \in [a,b]$. Therefore,

$$\begin{aligned} F(x) \leq G(x) &= F(a) + \int_a^x g(\xi)\,\mathrm{d}\xi \\ &= F(a) + \int_a^x f(\xi)\,\mathrm{d}\xi + \int_a^x (g(\xi) - f(\xi))\,\mathrm{d}\xi \\ &\leq F(a) + \int_a^x f(\xi)\,\mathrm{d}\xi + \epsilon \end{aligned}$$

by the definition of g. Since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary, this shows that

$$F(x) \leq F(a) + \int_a^x f(\xi)\,\mathrm{d}\xi.$$

A similar argument to the above, applied to $-F$, gives

$$-F(x) \leq -F(a) - \int_a^x f(\xi)\,\mathrm{d}\xi \quad \Longrightarrow \quad F(x) \geq F(a) + \int_a^x f(\xi)\,\mathrm{d}\xi,$$

which gives the theorem. ■

Our definition of absolute continuity allows us to state a more powerful version of the integration by parts formula than was given as Proposition I-3.4.28 for the Riemann integral.

2.9.36 Proposition (Integration by parts) *If* $f, g\colon [a, b] \to \mathbb{R}$ *are absolutely continuous, then*

$$\int_a^b f(x)g'(x)\,\mathrm{d}x = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\,\mathrm{d}x.$$

Proof We have

$$\begin{aligned} & f(x) = f(a) + \int_a^x f'(\xi)\,\mathrm{d}\xi \\ \Longrightarrow \quad & \int_a^b f(x)g'(x)\,\mathrm{d}x = \int_a^b f(a)g'(x)\,\mathrm{d}x + \int_a^b g'(x)\Big(\int_a^x f'(\xi)\,\mathrm{d}\xi\Big)\,\mathrm{d}x \\ \Longrightarrow \quad & \int_a^b f(x)g'(x)\,\mathrm{d}x = f(a)(g(b) - g(a)) + \int_a^b g'(x)\Big(\int_a^b \chi_{[a,x]}(\xi)f'(\xi)\,\mathrm{d}\xi\Big)\,\mathrm{d}x. \end{aligned} \tag{2.22}$$

By Corollary 2.8.8 the function $F(x, \xi) = g(x)\chi_{[a,x]}(\xi)f'(\xi)$ is integrable with respect to $\lambda_{[a,b]} \times \lambda_{[a,b]}$. Thus we may apply Fubini's Theorem (the version in Theorem 2.8.4) to the last of the above integrals to get

$$\begin{aligned} \int_a^b g'(x)\Big(\int_a^b \chi_{[a,x]}(\xi)f'(\xi)\,\mathrm{d}\xi\Big)\,\mathrm{d}x &= \int_a^b f'(\xi)\Big(\int_a^b \chi_{[\xi,b]}(x)g'(x)\,\mathrm{d}x\Big)\,\mathrm{d}\xi \\ &= \int_a^b f'(\xi)\Big(\int_\xi^b g'(x)\,\mathrm{d}x\Big)\,\mathrm{d}\xi \\ &= \int_a^b f'(\xi)(g(b) - g(\xi))\,\mathrm{d}\xi \\ &= f(b)g(b) - f(a)g(b) - \int_a^b f'(\xi)g(\xi)\,\mathrm{d}\xi, \end{aligned}$$

using the fact that $\chi_{[a,x]}(\xi) = \chi_{[\xi,b]}(x)$. Combining this with (2.22) gives the result. ■

2.9.8 Lebesgue points

One might speculate that Lebesgue measurable functions are very nasty. However, in Theorems 2.9.2 and 2.9.3 we show that measurable functions can be approximated well by "nice" functions. In this section we show that if a function is additionally integrable, then we can make some further conclusions about how nice it is.

The main result we state relies on taking a limit over intervals where the length of the interval goes to zero. To make this precise we need to define a directed set for the limit to be well-defined. We refer to for this notion of convergence using directed sets and nets. We let $I \subseteq \mathbb{R}$ be an interval, let $x_0 \in I$, and let $\mathscr{C}(x_0, I)$ be the set of closed subintervals of I containing x_0. This set is partially ordered by saying that $J_1 \preceq J_2$ if $J_1 \supseteq J_2$. It is easily verified that $\mathscr{C}(x_0)$ is a directed set with this partial order. If $f \in \mathsf{L}^{(1)}_{\mathrm{loc}}(I;\mathbb{R})$ then we define $P_{f,x_0} : \mathscr{C}(x_0, I) \to \mathbb{R}$ by

what

$$P_{f,x_0}(J) = \frac{1}{\lambda(J)} \int_J |f(x) - f(x_0)| \,\mathrm{d}x,$$

which defines a $\mathscr{C}(x_0, I)$ net.

2.9.37 Theorem (Almost every point is a Lebesgue point for an integrable function) *If* $\mathrm{f} \in \mathsf{L}^{(1)}(\mathrm{I};\mathbb{R})$ *then* $\lim \mathrm{P}_{\mathrm{f},\mathrm{x}_0} = 0$ *for almost every* $\mathrm{x}_0 \in \mathbb{R}$.

Proof We begin with a technical lemma.

1 Lemma *If* $\mathrm{f} \in \mathsf{L}^{(1)}([\mathrm{a},\mathrm{b}];\mathbb{R})$ *then there exists* $\mathrm{A} \subseteq [\mathrm{a},\mathrm{b}]$ *such that*

(i) $\lambda([\mathrm{a},\mathrm{b}] \setminus \mathrm{A}) = 0$ *and such that*

(ii) for all $\alpha \in \mathbb{R}$ *and for all* $\mathrm{x} \in \mathrm{A}$,

$$\lim_{\delta\downarrow 0} \frac{1}{\delta} \int_{\mathrm{x}}^{\mathrm{x}+\delta} |\mathrm{f}(\xi) - \alpha| \,\mathrm{d}\xi = \lim_{\delta\downarrow 0} \frac{1}{\delta} \int_{\mathrm{x}-\delta}^{\mathrm{x}} |\mathrm{f}(\xi) - \alpha| \,\mathrm{d}\xi = |\mathrm{f}(\mathrm{x}) - \alpha|.$$

Proof Let $\alpha \in \mathbb{R}$, let $(q_j)_{j\in\mathbb{Z}_{>0}}$ be an enumeration of the rationals and, for $j \in \mathbb{Z}_{>0}$, define $f_j \in \mathsf{L}^{(1)}([a,b];\mathbb{R})$ by

$$f_j(x) = |f(x) - \alpha|.$$

By part (ii) of Theorem 2.9.33, for each $j \in \mathbb{Z}_{>0}$ there exists a set $A_j \subseteq [a,b]$ such that $\lambda([a,b] \setminus A_j) = 0$ and such that, for all $x \in A_j$,

$$\lim_{\delta\downarrow 0} \frac{1}{\delta} \int_x^{x+\delta} g_j(\xi) \,\mathrm{d}\xi = \lim_{\delta\downarrow 0} \frac{1}{\delta} \int_{x-\delta}^{x} g_j(\xi) \,\mathrm{d}\xi = g_j(x).$$

Take $A = \cap_{j\in\mathbb{Z}_{>0}} A_j$, and note that

$$\lambda([a,b] \setminus A) = \lambda(\cup_{j\in\mathbb{Z}_{>0}} [a,b] \setminus A_j) = 0,$$

where we have used De Morgan's Laws and Exercise I-2.5.11.

Let $\delta \in \mathbb{R}_{>0}$ and let $k \in \mathbb{Z}_{>0}$ be such that $|q_k - \alpha| < \frac{\delta}{3}$. By Exercise I-2.2.8 we have

$$||f(x) - \alpha| - |f(x) - q_k|| \le |q_j - \alpha| < \tfrac{\delta}{3}$$

for all $x \in [a,b]$. Therefore,

$$\left|\frac{1}{\delta} \int_x^{x+\delta} |f(\xi) - \alpha| \,\mathrm{d}\xi - \frac{1}{\delta} \int_x^{x+\delta} |g_k(\xi) - \alpha| \,\mathrm{d}\xi\right| \le frac1\delta \int_x^{x+\delta} \frac{\delta}{3} \,\mathrm{d}\xi = \frac{\delta}{3}$$

for every $\delta \in \mathbb{R}_{>0}$ such that the integrals are defined. Therefore, we let $x \in A$ and let $\delta_0 \in \mathbb{R}_{>0}$ be such that

$$\left|\frac{1}{\delta}\int_x^{x+\delta} g_k(\xi)\,\mathrm{d}\xi - g_k(x)\right| < \frac{\delta}{3}$$

for all $\delta \in (0, \delta_0)$. Then, provided that $\delta \in (0, \delta_0)$ we have

$$\begin{aligned}\left|\frac{1}{\delta}\int_x^{x+\delta}|f(\xi)-\alpha|\,\mathrm{d}\xi - f(x)\right| &\le \left|\frac{1}{\delta}\int_x^{x+\delta}|f(\xi)-\alpha|\,\mathrm{d}\xi - \frac{1}{\delta}\int_x^{x+\delta}|g_k(\xi)-\alpha|\,\mathrm{d}\xi\right| \\ &+ \left|\frac{1}{\delta}\int_x^{x+\delta} g_k(\xi)\,\mathrm{d}\xi - g_k(x)\right| + |q_k - \alpha| < \frac{\delta}{3}+\frac{\delta}{3}+\frac{\delta}{3} = \delta,\end{aligned}$$

using the triangle inequality. This gives the left limit equal to the right expression in the statement of the lemma. The proof that the middle limit is equal to the right expression follows along entirely similar lines. ▼

It is now somewhat easy to complete the proof of the theorem. First suppose that $I = [a, b]$ is compact. As per the preceding lemma, let $A \subseteq [a, b]$ be such that $\lambda([a, b] \setminus A) = 0$ and such that, for all $\alpha \in \mathbb{R}$ and $x \in A$,

$$\lim_{\delta\downarrow 0}\frac{1}{\delta}\int_x^{x+\delta}|f(\xi)-\alpha|\,\mathrm{d}\xi = \lim_{\delta\downarrow 0}\frac{1}{\delta}\int_{x-\delta}^{x}|f(\xi)-\alpha|\,\mathrm{d}\xi = |f(x)-\alpha|.$$

Now let $x_0 \in A \cap (a, b)$ and let $\epsilon \in \mathbb{R}_{>0}$. Then there exists $\delta_0 \in \mathbb{R}_{>0}$ such that

$$\left|\frac{1}{\delta}\int_{x_0}^{x_0+\delta}|f(x)-f(x_0)|\,\mathrm{d}x\right| < \frac{\epsilon}{2}$$

and

$$\left|\frac{1}{\delta}\int_{x_0-\delta}^{x_0}|f(x)-f(x_0)|\,\mathrm{d}x\right| < \frac{\epsilon}{2}$$

for $\delta \in (0, \delta_0)$. Define $J_0 = [x_0 - \delta_0, x_0 + \delta_0]$. We may suppose that δ_0 is sufficiently small that $J_0 \in [a, b]$. If $J_0 \preceq J$ then $J \subseteq J_0$ and so we have

$$\begin{aligned}&\left|\frac{1}{\lambda(J)}\int_J|f(x)-f(x_0)|\,\mathrm{d}x\right| \\ &\qquad\le \left|\frac{1}{\delta_0}\int_{x_0}^{x_0+\delta_0}|f(x)-f(x_0)|\,\mathrm{d}x + \frac{1}{\delta_0}\int_{x_0-\delta_0}^{x_0}|f(x)-f(x_0)|\,\mathrm{d}x\right| \le \frac{\epsilon}{2}+\frac{\epsilon}{2} = \epsilon.\end{aligned}$$

This shows that $\lim P_{f,x_0} = 0$, giving the theorem when I is compact. When I is not compact, then we can write I as a countable union of compact intervals $(I_j)_{j\in\mathbb{Z}_{>0}}$. For each $j \in \mathbb{Z}_{>0}$ let $N_j \subseteq I_j$ be the set of measure zero such that if $x \in N_j$ we have $\lim P_{f,x} \neq 0$. Since $\lambda(\cup_{j\in\mathbb{Z}_{>0}} N_j) = 0$ by Exercise I-2.5.11, and since if $x \in I \setminus \cup_{j\in\mathbb{Z}_{>0}} N_j$ we have $\lim P_{f,x} = 0$, the theorem follows. ■

2.9.9 Maximal functions

2.9.10 The change of variables formula

In this section we state and prove a simple version of the change of variables formula for the Lebesgue integral. This is one of the places in our development where the extension to the Lebesgue integral on $\mathbb{R}^n$ is not so easily accomplished. Indeed, the higher-dimensional versions are difficult to prove in any useful degree of generality, and normally require the Radon–Nikodym Theorem (see, for example, Rudin [1986]). We refer to [Varberg 1971] for a quite general statement of the multivariable change of variable formula. Fortunately, we shall only need the single-variable change of variable, and this can be proved more directly, even though, as the reader can see, the proof is not quite trivial.

2.9.38 Theorem (Change of variable) *Let* $\mathrm{I}, \mathrm{J} \subseteq \mathbb{R}$ *be intervals with* $\phi\colon \mathrm{I} \to \mathrm{J}$ *a map with the properties that*

(i) ϕ *is surjective,*

(ii) ϕ *is either monotonically decreasing or monotonically increasing, and*

absolute continuity

(iii) there exists an integrable function $\phi'\colon \mathrm{I} \to \mathbb{R}$ *and* $\mathrm{x}_0 \in \mathrm{I}$ *so that*

$$\phi(\mathrm{x}) = \phi(\mathrm{x}_0) + \int_{[\mathrm{x}_0,\mathrm{x}]} \phi' \,\mathrm{d}\lambda_{[\mathrm{x}_0,\mathrm{x}]}.$$

If $\mathrm{f}\colon \mathrm{J} \to \mathbb{R}$ *is integrable then* $\mathrm{f} \circ \phi$ *is measurable,* $\mathrm{f} \circ \phi|\phi'|$ *is integrable, and*

$$\int_{\mathrm{J}} \mathrm{f} \,\mathrm{d}\lambda_{\mathrm{J}} = \int_{\mathrm{I}} \mathrm{f} \circ \phi|\phi'| \,\mathrm{d}\lambda_{\mathrm{I}}.$$

Proof We first take the case where $I = [a, b]$ and $J = [c, d]$. We claim that the theorem is true for step functions in this case. Indeed, let $g\colon [c, d] \to \mathbb{R}$ be a step function and write

$$g = \sum_{j=1}^{k} \alpha_j \chi_{I_j}$$

where $I_j = (x_j, x_{j-1}]$, $j \in \{0, 1, \dots, k\}$, are the endpoints of a partition $(I_1, \dots, I_k)$ of $[c, d]$. Corresponding to this partition of $[c, d]$ we define a partition $(J_1, \dots, J_k)$ of $[a, b]$ endpoints $(\xi_0, \xi_1, \dots, \xi_k)$ such that $\phi(\xi_j) = x_j$, $j \in \{0, 1, \dots, k\}$. There may be ambiguity in this definition of ξ_j, $j \in \{0, 1, \dots, k\}$, but this does not matter. Assuming that $\phi'(x) \geq 0$ for all x we then compute

$$\begin{aligned}
\int_a^b g \circ \phi(\xi)\phi'(\xi) \,\mathrm{d}\xi &= \sum_{j=1}^{k} \int_{\xi_{j-1}}^{\xi_j} \alpha_j \phi'(\xi) \,\mathrm{d}\xi \\
&= \sum_{j=1}^{k} \alpha_j (\phi(\xi_j) - \phi(\xi_{j-1})) \\
&= \sum_{j=1}^{k} \alpha_j (x_j - x_{j-1}) = \int_c^d g(x) \,\mathrm{d}x.
\end{aligned}$$

A similarly styled computation shows that the result is also true if $\phi'(x) \leq 0$.

Now suppose that f takes values in $[0, \infty)$. Using Theorem 2.9.2, let $(g_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of step functions on $[c, d]$ with the property that for almost every $x \in [c, d]$ we have $\lim_{j\to\infty} g_j(x) = f(x)$. Let us denote by Z_1 the subset of measure zero where this limit does not hold. By examining the proofs of Proposition 2.6.39 and Theorem 2.9.2 we see that we can take the sequence $(g_j)_{j\in\mathbb{Z}_{>0}}$ so that for each $x \in [c, d] \setminus Z_1$ the sequence $(g_j(x))_{j\in\mathbb{Z}_{>0}}$ is nondecreasing. Therefore the sequence $(g_j \circ \phi(\xi)|\phi'(\xi)|)_{j\in\mathbb{Z}_{>0}}$ is also nondecreasing provided that $\phi(\xi) \notin Z_1$. Indeed, provided that either

1. $\phi(\xi) \notin Z_1$ or
2. $\phi'(x) = 0$

hold, then we have $\lim_{j\to\infty} g_j \circ \phi(\xi)|\phi'(\xi)| = f \circ \phi(\xi)|\phi'(\xi)|$. We claim that the set of points $Z_2 \subseteq [a, b]$ where both conditions 1 and 2 fail to hold has measure zero. We do this with the aid of a lemma.

1 **Lemma** $Z \subseteq \mathbb{R}$ *has Lebesgue measure zero if and only if there exists a sequence* $(I_j)_{j\in\mathbb{Z}_{>0}}$ *of nonempty open intervals such that*

(i) $\sum_{j=1}^{\infty} \lambda(I_j) < \infty$ *and*

(ii) for each $x \in Z$ *there exists a sequence* $(j_k)_{k\in\mathbb{Z}_{>0}}$ *of* $\mathbb{Z}_{>0}$ *so that* $x \in I_{j_k}$, $k \in \mathbb{Z}_{>0}$.

Proof By definition, Z has Lebesgue measure zero if for each $\epsilon \in \mathbb{R}_{>0}$ there exists a family $(\tilde{I}_\ell)_{\ell\in\mathbb{Z}_{>0}}$ of open intervals, some possibly empty, for which $Z \subseteq \cup_{\ell\in\mathbb{Z}_{>0}} \tilde{I}_\ell$ and $\sum_{\ell=1}^{\infty} \lambda(\tilde{I}_\ell) < \epsilon$.

Suppose that there exists a collection of intervals $(I_j)_{j\in\mathbb{Z}_{>0}}$ having properties (i) and (ii) and let $\epsilon \in \mathbb{R}_{>0}$. Choose a finite collection $I_{j_1}, \dots, I_{j_m}$ of intervals so that

$$\sum_{j=1}^{\infty} \lambda(I_j) - \sum_{k=1}^{m} \lambda(I_{j_k}) < \epsilon.$$

It then follows that the family $(I_j)_{j\in\mathbb{Z}_{>0}} \setminus (I_{j_k})_{k\in\{1,\dots,m\}}$ of open intervals has total length less than ϵ. Furthermore, since only a finite number of intervals are removed from $(I_j)_{j\in\mathbb{Z}_{>0}}$, the remaining intervals still cover Z. Thus Z has Lebesgue measure zero.

Now suppose that Z has measure zero. For $n \in \mathbb{Z}_{>0}$ let $(I_{n,j})_{j\in\mathbb{Z}_{>0}}$ have the property that $Z \subseteq \cup_{j\in\mathbb{Z}_{>0}} I_{n,j}$ and that

$$\sum_{j=1}^{\infty} \lambda(I_{n,j}) < \frac{1}{2^n}.$$

Then the collection $(I_{j,n})_{j,n\in\mathbb{Z}_{>0}}$ satisfies (i) and (ii). ▼

According to the lemma, choose a sequence $(I_j)_{j\in\mathbb{Z}_{>0}}$ of intervals covering Z_1 and whose total length is finite. Define a step function $g_n \colon [c, d] \to \mathbb{R}$ by

$$g_n = \sum_{j=1}^{n} \chi_{I_j},$$

and note that for each $x \in Z_1$ we have $\lim_{n\to\infty} g_n(x) = \infty$. If $\xi \in Z_2$ it follows that $\lim_{n\to\infty} g_n \circ \phi(\xi)|\phi'(\xi)| = \infty$. Now note that

$$\int_a^b g_n \circ \phi(\xi)|\phi'(\xi)|\, d\xi = \int_c^d g_n(x)\, dx < \sum_{j=1}^{\infty} \lambda(I_j) < \infty.$$

It follows from Proposition 2.7.12 that $\lambda(Z_2) = 0$.

Thus we have shown that, provided that $I = [a, b]$, that $J = [c, d]$, and that $f(J) \subseteq [0, \infty)$, for almost every $\xi \in [a, b]$ and almost every $x \in [c, d]$ we have

$$\lim_{j\to\infty} g_j(x) = f(x), \quad \lim_{j\to\infty} g_j \circ \phi(\xi)|\phi'(\xi)| = f \circ \phi(\xi)|\phi'(\xi)|$$

with both limits being monotonic, and that for each $j \in \mathbb{Z}_{>0}$ we have

$$\int_a^b g_j \circ \phi(\xi)\, \mathrm{d}\xi = \int_c^d g_j(x)\, \mathrm{d}x.$$

The result under the current assumptions now follows by the Monotone Convergence Theorem. For an arbitrary f with $I = [a, b]$ and $J = [c, d]$ the result follows from breaking f into its positive and negative parts.

It remains to prove the result for general intervals I and J. Let $(I_n = [a_n, b_n])_{j\in\mathbb{Z}_{>0}}$ be a sequence of intervals with the property that $\mathrm{int}(I) = \cup_{n\in\mathbb{Z}_{>0}} I_n$. Define $J_n = \phi(I_n)$, $n \in \mathbb{Z}_{>0}$, noting that J_n so defined is a closed interval by monotonicity of ϕ. We then have, by the Dominated Convergence Theorem,

$$\int_I f\, \mathrm{d}\lambda_I = \lim_{n\to\infty} \int_I \chi_{I_n} f\, \mathrm{d}\lambda_I, \quad \int_J f \circ \phi|\phi'|\, \mathrm{d}\lambda_J = \lim_{n\to\infty} \int_J \chi_{J_n} f \circ \phi|\phi'|\, \mathrm{d}\lambda_J.$$

From this the result follows since

$$\int_I \chi_{I_n} f\, \mathrm{d}\lambda_I = \int_J \chi_{J_n} f \circ \phi|\phi'|\, \mathrm{d}\lambda_J. \qquad \blacksquare$$

2.9.11 Topological characterisations of the deficiencies of the Riemann integral[16]

In Section 2.7.5 we saw that it was possible to give interesting topological characterisations of the Dominated Convergence Theorem for the general measure theoretic integral. These characterisations are, of course, inherited by the Lebesgue integral. That is to say, one can specialise Theorems 2.7.40 and 2.7.42 to the Lebesgue integral as follows.

2.9.39 Theorem (Topological "everywhere" Dominated Convergence Theorem for the Lebesgue integral) *If* $\mathrm{A} \in \mathscr{L}(\mathbb{R})$ *then* C_p*-bounded subsets of* $\mathsf{L}^{(1)}(\mathrm{A}; \mathbb{R})$ *are* C_p*-sequentially closed.*

2.9.40 Theorem (Limit structure "almost everywhere" Dominated Convergence Theorem for the Lebesgue integral) *If* $\mathrm{A} \in \mathscr{L}(\mathbb{R})$ *then* $\mathscr{L}_{\lambda_\mathrm{A}}$*-bounded subsets of* $\mathsf{L}^1(\mathrm{A}; \mathbb{R})$ *are* $\mathscr{L}_{\lambda_\mathrm{A}}$*-sequentially closed.*

In this section we give a couple of examples that show that these theorems do not hold for the Riemann integral. First we consider the "everywhere" version of the Dominated Convergence Theorem.

[16] The results in this section are not used in an essential way anywhere else in the text.

2.9.41 Example (The topological "everywhere" Dominated Convergence Theorem does not hold for the Riemann integral) By means of an example, we show that there are C_p-bounded subsets of the seminormed vector space $\mathsf{R}^{(1)}([0,1];\mathbb{R})$ that are not C_p-sequentially closed. Let us denote

$$B = \{f \in \mathsf{R}^{(1)}([0,1];\mathbb{R}) \mid |f(x)| \leq 1\},$$

noting by Proposition 2.7.39 that B is C_p-bounded. Let $(q_j)_{j\in\mathbb{Z}_{>0}}$ be an enumeration of the rational numbers in $[0,1]$ and define a sequence $(f_k)_{k\in\mathbb{Z}_{>0}}$ in $\mathsf{R}^{(1)}([0,1];\mathbb{R})$ by

$$f_k(x) = \begin{cases} 1, & x \in \{q_1,\dots,q_k\}, \\ 0, & \text{otherwise}. \end{cases}$$

The sequence converges in the C_p-topology to the characteristic function of $\mathbb{Q}\cap[0,1]$; let us denote this function by f. This limit function is not Riemann integrable and so not in $\mathsf{R}^{(1)}([0,1];\mathbb{R})$. Thus B is not C_p-sequentially closed. •

Next we turn to the "almost everywhere" version of the Dominated Convergence Theorem for the Riemann integral.

2.9.42 Example (The limit structure "almost everywhere" Dominated Convergence Theorem does not hold for the Riemann integral) Recall from Section 2.6.6 that $\mathsf{L}^0([0,1];\mathbb{R})$ denotes the set of equivalence classes of $\mathbb{R}$-valued measurable functions on $[0,1]$ under the equivalence relation of almost everywhere equality. We denote by $\mathsf{R}^1([0,1];\mathbb{R})$ the image of $\mathsf{R}^{(1)}([0,1];\mathbb{R})$ by the projection from $\mathsf{L}^{(0)}([0,1];\mathbb{R})$ to $\mathsf{L}^0([0,1];\mathbb{R})$. Thus elements of $\mathsf{R}^1([0,1];\mathbb{R})$ are equivalence classes of $\mathbb{R}$-valued Riemann integrable functions under the equivalence relation of almost everywhere equality. We denote elements of $\mathsf{R}^1([0,1];\mathbb{R})$ by $[f]$, reflecting the fact that they are equivalence classes of functions. For brevity we denote the Lebesgue measure on $[0,1]$ by λ.

We give an example that shows that $\mathscr{L}_\lambda$-bounded subsets of the normed vector space $\mathsf{R}^1([0,1];\mathbb{R})$ are not $\mathscr{L}_\lambda$-sequentially closed. We first remark that the construction of Example 2.9.41, projected to $\mathsf{R}^1([0,1];\mathbb{R})$, does not suffice because $[f]$ is equal to the equivalence class of the zero function which *is* Riemann integrable, even though f is not. The fact that $[f]$ contains functions that are Riemann integrable and functions that are not Riemann integrable is a reflection of the fact that the set

$$\mathsf{R}_0([0,1];\mathbb{R}) = \left\{f\colon [0,1]\to\mathbb{R} \;\middle|\; f \text{ Riemann integrable and } \int_0^1 f(x)\,\mathrm{d}x = 0\right\}$$

is not sequentially closed. This is a phenomenon of interest, but it is not what is of interest here.

We use the construction of the function f from the proof of Proposition 2.1.12. In that proof, the function f was shown to have the following properties:

1. f is the pointwise limit of a sequence $(f_k)_{k\in\mathbb{Z}_{>0}}$ of Riemann integrable functions;
2. any function almost everywhere equal to f is not Riemann integrable.

Therefore, by Theorem 2.6.51 it follows that $([f_k])_{k\in\mathbb{Z}_{>0}}$ is $\mathscr{L}_\lambda$-convergent to $[f]$. Moreover, $[f] \notin \mathsf{R}^1([0,1];\mathbb{R})$. To complete the example, we note that the sequence $([G_k])_{j\in\mathbb{Z}_{>0}}$ is in the set

$$B = \{[f] \in \mathsf{R}^1([0,1];\mathbb{R}) \mid |f(x)| \leq 1 \text{ for almost every } x \in [0,1]\},$$

which is $\mathscr{L}_\lambda$-bounded by Proposition 2.7.41. The example shows that this $\mathscr{L}_\lambda$-bounded subset of $\mathsf{R}^1([0,1];\mathbb{R})$ is not $\mathscr{L}_\lambda$-sequentially closed. •

Exercises

2.9.1 Use Lemma 2.9.32 to directly conclude that the Cantor function of Example 2.9.25 is not absolutely continuous.

2.9.2 Give an example of a function $f\colon \mathbb{R} \to \mathbb{R}$ such that $|f|$ is Lebesgue measurable, but f is not Lebesgue measurable.

2.9.3 Answer the following two questions.

(a) Why must a Riemann integrable function $f\colon [a,b] \to \mathbb{R}$ on a compact interval be bounded?

(b) Provide an unbounded function on $[a,b]$ that is continuous when restricted to (a,b), and that is Lebesgue integrable.

One of the differences between the Lebesgue and Riemann integral is that the Lebesgue integral is defined by first approximating a measurable function by a sequence of simple function only from below. In contrast, for the Riemann integral, one asks that the function be approximated from below *and* above by step functions. One might legitimately wonder whether this is asking too much of the approximation, and whether one can get away, as one does with the Lebesgue integral, by approximation from (say) below. The following exercise asks you to explore this.

2.9.4 Let $I = [0,1]$ and let

$$f = \chi_{I\cap\mathbb{Q}}, \quad g = \chi_{I\cap(\mathbb{R}\setminus\mathbb{Q})}.$$

Answer the following questions.

(a) Show that $I_-(f) = I_-(g) = 0$. Thus, when approximated just by step functions from below, both f and g have zero "integral."

(b) Show that $I_-(f+g) \neq I_-(f) + I_-(g)$. Thus the "integral" is not linear.

2.9.5 Let $A \subseteq I = [0,1]$ be the subset of irrational numbers, and let χ_A be the characteristic function. Show that $\int_I \chi_A \,\mathrm{d}\lambda = 1$.

2.9.6 Show that there is a function $f\colon [0,1] \to \mathbb{R}$ that is not Riemann integrable, but for which $|f|$ is Riemann integrable.

2.9.7 Let $I = [0, \infty)$ and define $f\colon I \to \mathbb{R}$ by $f(x) = x$. Use the Monotone Convergence Theorem to show that f is not integrable.

2.9.8 Let $I \subseteq \mathbb{R}$ be an interval, and let $f\colon I \to \mathbb{R}$ be continuous. Show that if

$$\lambda(\{x \in I \mid f(x) \neq 0\}) = 0$$

then $f(x) = 0$ for every $x \in I$.

Section 2.10

The multivariable Lebesgue integral

2.10.1 Lebesgue measurable functions

In this section we repeat for the n-dimensional Lebesgue measure the results of the preceding section. The general ideas of the proofs are the same, but there are a few points where additional technicalities arise.

We begin with the definitions.

2.10.1 Definition (Borel measurable, Lebesgue measurable) Let $A \subseteq \mathbb{R}^n$. A function $f\colon A \to \overline{\mathbb{R}}$ is

(i) ***Borel measurable*** if $A \in \mathscr{B}(\mathbb{R}^n)$ and if f is $\mathscr{B}(A)$-measurable and

(ii) ***Lebesgue measurable*** if $A \in \mathscr{L}(\mathbb{R}^n)$ and if f is $\mathscr{L}(A)$-measurable. •

We recall from Section II-1.6.1 the notion of a step function defined on a fat compact rectangle in $\mathbb{R}^n$.

2.10.2 Theorem (Lebesgue measurable functions are approximated by step functions) *If* $\mathrm{R} \subseteq \mathbb{R}^{\mathrm{n}}$ *is a fat compact rectangle, if* $\mathrm{f}\colon \mathrm{R} \to \overline{\mathbb{R}}$ *is measurable and satisfies*

$$\lambda_{\mathrm{n}}(\{\mathbf{x} \in \mathrm{R} \mid \mathrm{f}(\mathbf{x}) \in \{-\infty, \infty\}\} = 0,$$

and if $\epsilon_1, \epsilon_2 \in \mathbb{R}_{>0}$, *then there exists a step function* $\mathrm{g}\colon \mathrm{R} \to \mathbb{R}_{\geq 0}$ *such that*

$$\lambda_{\mathrm{n}}(\{\mathbf{x} \in \mathrm{R} \mid |\mathrm{f}(\mathbf{x}) - \mathrm{g}(\mathbf{x})| \geq \epsilon_1\}) < \epsilon_2.$$

Proof It suffices to prove the theorem when $\epsilon_1 = \epsilon_2 = \epsilon$. Thus we take $\epsilon \in \mathbb{R}_{>0}$.

For $k \in \mathbb{Z}_{>0}$ define

$$A_k = \{\boldsymbol{x} \in R \mid |f(\boldsymbol{x})| \geq k\},$$

and note that the sequence $(\lambda_n(R \setminus A_k))_{k \in \mathbb{Z}_{>0}}$ is monotonically increasing and bounded above by $\lambda_n(R)$. Thus it is convergent by Theorem I-2.3.8. Moreover, it converges to $\lambda_n(R)$. Indeed, if the sequence converges to $\nu < \lambda_n(R)$ then this would imply, by Proposition 2.3.3, that

$$\lim_{k\to\infty} \lambda_n(R \setminus A_k) = \lambda_n(R \setminus \cup_{k \in \mathbb{Z}_{>0}} A_k) < \lambda_n(R).$$

Thus there exists a set $B \subseteq R$ of positive measure such that $R = (\cup_{k \in \mathbb{Z}_{>0}} A_k \,\mathring{\cup}\, B)$. Note if $\boldsymbol{x} \in B$ then $|f(\boldsymbol{x})| = \infty$, contradicting our assumptions on f. Thus we indeed have $\lim_{k\to\infty} \lambda_n(R \setminus A_k) = \lambda_n(R)$. Thus there exists $M \in \mathbb{Z}_{>0}$ such that $\lambda_n(R \setminus A_M) < \lambda_n(R) - \frac{\epsilon}{2}$, i.e., $\lambda_n(A_M) < \frac{\epsilon}{2}$. Therefore,

$$\lambda_n(\{\boldsymbol{x} \in R \mid |f(\boldsymbol{x})| \geq M\}) < \tfrac{\epsilon}{2}.$$

Then define $f_M\colon R \to \mathbb{R}$ by

$$f_M(x) = \begin{cases} f(x), & |f(x)| < M, \\ M, & |f(x)| \geq M, \\ -M, & f(x) < -M. \end{cases}$$

Note that f_M is measurable by Proposition 2.6.16.

Now take $K \in \mathbb{Z}_{>0}$ such that $2^{-K} < \epsilon$ and such that $K \geq M$. If we follow the construction in the proof of Proposition 2.6.39 then we define

$$A_{+,K,j} = \{x \in R \mid 2^{-K}(j-1) \leq f_M(x) < 2^{-K}j\}$$

and

$$A_{-,K,j} = \{x \in R \mid -2^{-K}j \leq f_M(x) < -2^{-K}(j-1)\}$$

for $j \in \{1, \dots, K2^K\}$. Since $K \geq M$ we have

$$R = (\cup_{j=1}^{K2^K} A_{+,K,j}) \cup (\cup_{j=1}^{K2^K} A_{-,K,j}).$$

Moreover, if we define a simple function $h\colon R \to \mathbb{R}$ by

$$h(x) = \begin{cases} 2^{-K}(j-1), & x \in A_{+,K,j}, \\ -2^{-K}j, & x \in A_{-,K,j}, \end{cases}$$

then we have $|h(x) - f_M(x)| < \epsilon$ for every $x \in R$.

Now that we have a $\mathbb{R}$-valued simple function h that approximates f_M to within ϵ on R, let us dispense with the cumbersome notation above we introduced to define h, and instead write $h = \sum_{j=1}^k a_j \chi_{A_j}$ for $a_1, \dots, a_k \in \mathbb{R}$ and for a partition $(A_1, \dots, A_k)$ of R into Lebesgue measurable sets. Fix $j \in \{1, \dots, k\}$. Since A_j is measurable, by Corollary 2.5.19 we can write $A_j = U_j \cup B_j$ where U_j is open and where $B_j \subseteq U_j$ satisfies $\lambda_n(B_j) < \frac{\epsilon}{8k}$. Recall from Proposition II-1.2.21 that open sets are countable unions of open balls. This result holds also for cubes as well as balls, cf. the proof of Lemma 2.5.32. Thus, since U_j is open, it is a countable union of open rectangles. If U_j is in fact a finite union of open rectangles then recall from Proposition 2.1.2 that U_j is a finite disjoint union of (not necessarily open) rectangles. Let V_j denote the union of the interiors of these rectangles. If any of the disjoint open rectangles comprising V_j have common boundary, then these rectangles may be shrunk so that they their complement in A_j has measure at most $\frac{\epsilon}{2k}$. Next suppose that U_j is a countably infinite union of open rectangles $(R_{j,l})_{l \in \mathbb{Z}_{>0}}$. Since U_j is bounded we must have $\sum_{l=1}^\infty \lambda_n(R_{j,l}) < \infty$. Therefore, there exists $N_j \in \mathbb{Z}_{>0}$ such that $\sum_{j=N_j+1}^\infty \lambda_n(R_{l,j}) < \frac{\epsilon}{8k}$. As above, by Proposition 2.1.2 we write the union of $R_{1,j}, \dots, R_{N_j,j}$ as a finite disjoint union of (not necessarily open) rectangles. Let V_j denote the union of the interiors of these rectangles. If any of the disjoint open rectangles comprising V_j have common boundary, they can be shrunk while maintaining the fact that the measure of their complement in A_j is at most $\frac{\epsilon}{2k}$. Define $g\colon R \to \mathbb{R}$ on V_j by asking that $g(x) = a_j$ for $x \in V_j$. Doing this for each $j \in \{1, \dots, k\}$ defines $g\colon R \to \mathbb{R}$ on the set $\cup_{j=1}^k V_j$ which is a finite disjoint union of open rectangles whose complement has measure at most $\frac{\epsilon}{2}$. The complement to $\cup_{j=1}^k V_j$ is a union of rectangles by Proposition 2.1.8. On these rectangles define g to be, say, 0. Note that g as constructed is not quite a step function since the rectangles will not generally be those of a partition of R. However, one can define a partition which has all of the rectangles used to define g as unions of subrectangles. To do this, one merely takes as endpoints for the partitions of the n axes the union of the endpoints of all rectangles. Note that $g(\boldsymbol{x}) = h(\boldsymbol{x})$ for $\boldsymbol{x} \in \cup_{j=1}^{k} V_j$.

Note that if $\boldsymbol{x} \in (\cup_{j=1}^{k} V_j) \cup (R \setminus A_M)$ we have

$$|g(\boldsymbol{x}) - f(\boldsymbol{x})| = |h(\boldsymbol{x}) - f_M(\boldsymbol{x})| < \epsilon.$$

Therefore,

$$\lambda_n(\{\boldsymbol{x} \in R \mid f(\boldsymbol{x}) - g(\boldsymbol{x}) \geq \epsilon\}) \subseteq R \setminus ((\cup_{j=1}^{k} V_j) \cup (R \setminus A_M)),$$

and

$$\lambda_n(R \setminus ((\cup_{j=1}^{k} V_j) \cup (R \setminus A_M))) < \epsilon,$$

giving the result. ■

A similar sort of result holds for approximations of measurable functions by continuous functions.

2.10.3 Theorem (Lebesgue measurable functions are approximated by continuous functions) *If* $\mathrm{R} \subseteq \mathbb{R}^{\mathrm{n}}$ *is a fat compact rectangle, if* $\mathrm{f}\colon \mathrm{R} \to \overline{\mathbb{R}}$ *is measurable and satisfies*

$$\lambda_{\mathrm{n}}(\{\mathbf{x} \in \mathrm{R} \mid \mathrm{f}(\mathbf{x}) \in \{-\infty, \infty\}\} = 0,$$

and if $\epsilon_1, \epsilon_2 \in \mathbb{R}_{>0}$*, then there exists a step function* $\mathrm{g}\colon \mathrm{R} \to \mathbb{R}_{\geq 0}$ *such that*

$$\lambda_{\mathrm{n}}(\{\mathbf{x} \in \mathrm{R} \mid |\mathrm{f}(\mathbf{x}) - \mathrm{g}(\mathbf{x})| \geq \epsilon_1\}) < \epsilon_2.$$

Proof We shall merely outline how this works, since the details are straightforward once one has the idea at hand. We assume that $\epsilon_1 = \epsilon_2 = \epsilon$. By the procedure of Theorem 2.9.2, we approximate f with a step function g such that

$$\lambda_n(\{\boldsymbol{x} \in R \mid |f(\boldsymbol{x}) - g(\boldsymbol{x})| \geq \epsilon\}) < \tfrac{\epsilon}{2}.$$

Moreover, the subset of R where $|f(\boldsymbol{x}) - g(\boldsymbol{x})| < \epsilon$ is a finite union of rectangles whose closures are disjoint; let us denote these rectangles by $R_1, \dots, R_k$. On each of the rectangles, g is constant. On the complement of these rectangles the value of g can be taken to be zero without loss of generality. Let us fix $j \in \{1, \dots, k\}$ and let $R'_j \subseteq R_j$ be a closed rectangle such that $\lambda_n(R'_j) > \lambda_n(R_j) - \frac{\epsilon}{2k}$. Let us suppose that $R_j = I_{j1} \times \cdots \times I_{jn}$ and that $R'_j = I_{j1} \times \cdots \times I_{jn}$. Then define $h_{jl}\colon I_{jl} \to \mathbb{R}$ such that

1. h_{jl} is continuous,
2. $h_{jl}|I'_{jl}$ takes the constant value of $g|R_j$, and
3. $\lim_{x \to_{I_{jl}} a_{jl}} h_{jl}(x) = 0$ and $\lim_{x \to_{I_{jl}} b_{jl}} h_{jl}(x) = 0$, where a_{jl} and b_{jl} are the left and right endpoints of I_{jl}.

Then define h on R_j by asking that

$$h(\boldsymbol{x}) = h_{j1}(x_1)\cdots h_{jn}(x_n), \qquad \boldsymbol{x} \in R_j.$$

Doing this for each of the rectangles $R_1, \dots, R_k$, and defining h to be zero at points outside these rectangles yields a continuous function h agreeing with g on $R_1' \cup \cdots \cup R_k'$. Finally, note that

$$\lambda_n(R \setminus \cup_{j=1}^k R_j') < \epsilon,$$

which completes the proof. ■

The notion of support given in Definition 2.9.4 for functions of a single variable are trivially adapted to functions of multiple variables.

2.10.4 Definition (Support of a measurable function) Let $f \in \mathsf{L}^{(0)}(\mathbb{R}^n; \overline{\mathbb{R}})$ and define

$$\mathscr{O}_f = \{U \subseteq \mathbb{R}^n \mid U \text{ open and } f(\boldsymbol{x}) = 0 \text{ for almost every } \boldsymbol{x} \in U\}.$$

Then the ***support*** of f is $\mathrm{supp}(f) = \mathbb{R} \setminus (\cup_{U \in \mathscr{O}_f} U)$. •

The following two results are proved exactly in the same way as their single variable counterparts.

2.10.5 Proposition (Characterisation of support) *For* $f, g \in \mathsf{L}^{(0)}(\mathbb{R}^n; \overline{\mathbb{R}})$, *the following statements hold:*

(i) $f(\mathbf{x}) = 0$ *for almost every* $\mathbf{x} \in \mathbb{R} \setminus \mathrm{supp}(f)$*;*

(ii) if $f(\mathbf{x}) = g(\mathbf{x})$ *for almost every* $\mathbf{x} \in \mathbb{R}^n$ *then* $\mathrm{supp}(f) = \mathrm{supp}(g)$*.*

2.10.6 Proposition (The support of a continuous function) *If* $f\colon \mathbb{R}^n \to \mathbb{R}$ *is continuous then*

$$\mathrm{supp}(f) = \mathrm{cl}(\{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) \neq 0\}).$$

2.10.2 The (conditional) Lebesgue integral

2.10.3 Properties of the Lebesgue integral

2.10.4 Swapping operations with the Lebesgue integral

2.10.5 The change of variables formula

2.10.7 Theorem (The multivariable change of variable theorem)

2.10.6 Locally integrable functions

2.10.8 Proposition (Characterisation of locally Lebesgue integrable functions) *For a function* $\mathrm{f}\colon \mathbb{R}^{\mathrm{n}} \to \overline{\mathbb{R}}$ *the following statements are equivalent:*

(i) f *is locally Lebesgue integrable;*

(ii) for each $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$ *there exists a neighbourhood* U *of* $\mathbf{x}$ *such that* $\mathrm{f}|\mathrm{U} \in \mathsf{L}^{(1)}(\mathrm{U};\overline{\mathbb{R}})$*;*

(iii) for every continuous function $\mathrm{g}\colon \mathbb{R}^{\mathrm{n}} \to \mathbb{R}$ *such that* $\operatorname{supp}(\mathrm{g})$ *is compact, it holds that* $\mathrm{fg} \in \mathsf{L}^{(1)}(\mathbb{R}^{\mathrm{n}};\overline{\mathbb{R}})$*.*

Proof (i) $\implies$ (ii) Let $x \in \mathbb{R}^n$ and let K be the cube with sides of length 2 and centre x:

$$K = [x_1 - 1, x_1 + 1] \times \cdots \times [x_n - 1, x_n + 1].$$

By hypothesis, $f|K \in \mathsf{L}^{(1)}(K;\overline{\mathbb{R}})$ and so $f \in \mathsf{L}^{(1)}(\operatorname{int}(K);\overline{\mathbb{R}})$ by Proposition 2.7.22. This gives the result with $U = \operatorname{int}(K)$.

(ii) $\implies$ (iii) Let $g\colon \mathbb{R} \to \mathbb{R}$ be continuous with compact support. For $x \in \operatorname{supp}(g)$ there exists a neighbourhood U_x of x such that $f \in \mathsf{L}^{(1)}(U_x, \overline{\mathbb{R}})$. Since $(U_x)_{x \in \operatorname{supp}(g)}$ covers the compact set $\operatorname{supp}(g)$ there exists $x_1, \dots, x_k \in \operatorname{supp}(g)$ such that $\operatorname{supp}(g) \subseteq \cup_{j=1}^k U_{x_j}$. Since g is continuous with compact support there exists $M \in \mathbb{R}_{>0}$ such that $|g(x)| \le M$ for every $x \in \mathbb{R}$ by Theorem II-1.3.31. Then

$$\int_{\mathbb{R}^n} |fg| \, d\lambda_n \le M \int_{\operatorname{supp}(g)} f \, d\lambda_{\operatorname{supp}(g)} \le M \sum_{j=1}^k \int_{U_{x_j}} f \, d\lambda_{U_{x_j}} < \infty,$$

since $\operatorname{supp}(g) \subseteq \cup_{j=1}^k U_{x_j}$. This gives the desired conclusion.

(iii) $\implies$ (i) Let $K \subseteq \mathbb{R}$ be compact and let $a_j, b_j \in \mathbb{R}$, $j \in \{1, \dots, n\}$, be such that

$$K \subseteq [a_1, b_1] \times \cdots \times [a_n, b_n] \triangleq R.$$

For $j \in \{1, \dots, n\}$ define

$$g_j(x) = \begin{cases} 1, & x_j \in [a_j, b_j], \\ x_j - (a_j - 1), & x_j \in [a_j - 1, a_j), \\ -x_j + (b_j + 1), & x \in (b_j, b_j + 1], \\ 0, & \text{otherwise}. \end{cases}$$

Now take $g\colon \mathbb{R} \to \mathbb{R}$ defined by $g = g_1 \dots g_n$. Note that g is positive, continuous with compact support, and $g(x) = 1$ for all $x \in R$. Then

$$\int_K |f| \, d\lambda_K \le \int_R |f| \, d\lambda_R \le \int_{\mathbb{R}} |fg| \, d\lambda_n < \infty,$$

giving the result. ■

Section 2.11

Differentiation of measures

In this section we consider the matter of comparing measures on a measurable space. As we shall see, there are natural classifications of measures relative to a fixed measure. This classification is helpful in understanding the nature of measures in general. The constructions arising from this section are also useful in many more or less concrete and important measure theoretic developments. In this regard, we particularly mention the multivariable change of variables theorem (Theorem 2.10.7) and the topological duals of spaces of integrable functions (Section 3.10.1).

Do I need to read this section? This section should be read when a need to read it is encountered. •

2.11.1 Absolutely continuous measures

In Section 2.9.6 we studied absolutely continuous functions of a real variable. In this section we consider the notion of a measure being absolutely continuous to another measure. The connection between these two concepts is that an absolutely continuous function has associated with it a measure that is absolutely continuous with respect to the Lebesgue measure. We shall not understand this fully until . what?

We begin with the definition, recalling that we sometimes call a "measure" a "positive measure" when we wish to explicitly distinguish it from a signed, complex, or vector measure.

2.11.1 Definition (Absolutely continuous measure) Let $(X, \mathscr{A})$ and let μ be a positive measure on $\mathscr{A}$.

(i) A positive measure ν is ***absolutely continuous*** with respect to μ if, for every $A \in \mathscr{A}$ for which $\mu(A) = 0$, it follows that $\nu(A) = 0$.

(ii) A signed or complex measure ν is ***absolutely continuous*** with respect to μ if, for every $A \in \mathscr{A}$ for which $\mu(A) = 0$, it follows that $|\nu|(A) = 0$.

(iii) A vector measure $\boldsymbol{\nu}$ taking values in $\mathbb{R}^n$ is ***absolutely continuous*** with respect to μ if, for every $A \in \mathscr{A}$ for which $\mu(A) = 0$, it follows that $\|\boldsymbol{\nu}\|_{\mathbb{R}^n}(A) = 0$.

Sometimes one writes $\nu \ll \mu$ to signify that ν is absolutely continuous with respect to μ. •

In order that we have a few example to hang onto as we discuss the various topics in this section, let us give a few simple example.

2.11.2 Examples (Absolutely continuous measures)

1. Let $(X, \mathscr{A}, \mu)$ be a measure space and let $f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$. Then the measure $f \cdot \mu$, (cf. Proposition 2.7.65), defined by

$$(f \cdot \mu)(A) = \int_X f \chi_A \, \mathrm{d}\mu$$

is absolutely continuous, since, if $Z \in \mathscr{A}$ satisfies $\mu(Z) = 0$, then

$$(f \cdot \mu)(A) = \int_X f \chi_Z \, \mathrm{d}\mu = 0$$

by virtue of Proposition 2.7.11.

2. Let $(X, \mathscr{A}, \mu)$ be a measure space with μ continuous, i.e., $\mu(\{x\}) = 0$ for every $x \in X$. Then the point mass measure δ_x defined by $\delta_x \colon \mathscr{A} \to \overline{\mathbb{R}}_{\geq 0}$ by

$$\delta_x(A) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A \end{cases}$$

is not absolutely continuous with respect to μ since $\mu(\{x\}) = 0$ but $\delta_x(\{x\}) = 1$. •

Let us now characterise the absolutely continuous positive measures.

2.11.3 Theorem (Radon–Nikodym[17] Theorem for positive measures) *Let* $(\mathrm{X}, \mathscr{A})$ *be a measurable space and let* μ *and* ν *be positive measures on* $\mathscr{A}$ *with* μ *being* σ*-finite. Then* ν *is absolutely continuous with respect to* μ *if and only if there exists* $\mathrm{h}_\nu \in \mathsf{L}^{(0)}((\mathrm{X}, \mathscr{A}); \overline{\mathbb{R}}_{\geq 0})$ *such that*

$$\nu(\mathrm{A}) = \int_{\mathrm{X}} \mathrm{h}_\nu \chi_{\mathrm{A}} \, \mathrm{d}\mu$$

for every $\mathrm{A} \in \mathscr{A}$*. If* ν *is also* σ*-finite then* h_ν *can be taken to be* $\mathbb{R}_{\geq 0}$*-valued. Moreover, if* $\mathrm{h}'_\nu \in \mathsf{L}^{(0)}((\mathrm{X}, \mathscr{A}); \mathbb{R}_{\geq 0})$ *is another such function, then*

$$\mu(\{\mathrm{x} \in \mathrm{X} \mid \mathrm{h}_\nu(\mathrm{x}) \neq \mathrm{h}'_\nu(\mathrm{x})\}) = 0.$$

Proof Note that the "if" assertion of the theorem is obvious since if $\mu(A) = 0$ then

$$\int_X h_\nu \chi_A \, \mathrm{d}\mu = 0$$

by Proposition 2.7.11.

For the converse, we first suppose that μ and ν are finite.

In this case, let us denote

$$\mathscr{S}_\nu = \left\{ f \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R}_{\geq 0}) \;\middle|\; \int_X f \chi_A \, \mathrm{d}\mu \leq \nu(A),\ A \in \mathscr{A} \right\}.$$

[17] Johann Radon (1887–1956) was born in what is now the Czech Republic and made mathematical contributions to analysis and differential geometry. Otton Marcin Nikodym (1887–1974) was born in what is now the Ukraine, but in 1887 was part of the Austro–Hungarian Empire. His mathematical work was in the area of analysis.

Note that $\mathscr{F}_\nu \neq \emptyset$ since the zero function is in $\mathscr{F}_\nu$. Let $(g_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathscr{F}_\nu$ for which

$$\lim_{j\to\infty}\int_X g_j\,\mathrm{d}\mu = \sup\Big\{\int_X f\,\mathrm{d}\mu \;\Big|\; f\in\mathscr{F}_\nu\Big\},$$

the expression on the right being finite since μ is finite. For $k\in\mathbb{Z}_{>0}$ and $x\in X$ define

$$f_k(x) = \max\{g_1(x),\dots,g_k(x)\}.$$

Note that $f_j \in \mathsf{L}^{(0)}((X,\mathscr{A});\mathbb{R}_{\geq 0})$ by Proposition 2.6.16 (along with an induction for $k > 2$). Moreover, if

$$A_j = \{x\in X \mid f_k(x) = g_j(x)\},\qquad j\in\{1,\dots,k\},$$

then we have, for $A\in\mathscr{A}$,

$$\begin{aligned}\int_X f_j\chi_A\,\mathrm{d}\mu &= \int_X g_1\chi_{A\cap A_1}\,\mathrm{d}\mu + \dots + \int_X g_k\chi_{A\cap A_k}\,\mathrm{d}\mu\\ &\leq \nu(A\cap A_1)+\dots+\nu(A\cap A_k) = \nu(A).\end{aligned}$$

Thus $f_k\in\mathscr{F}_\nu$ for each $k\in\mathbb{Z}_{>0}$. Moreover, $f_{k+1}(x)\geq f_k(x)$ for every $k\in\mathbb{Z}_{>0}$ and $x\in X$. Therefore, if we define $f_\nu(x) = \lim_{k\to\infty} f_k(x)$, $f_\nu \in \mathsf{L}^{(0)}((X,\mathscr{A});\overline{\mathbb{R}}_{\geq 0})$ by Proposition 2.6.18. Moreover, by the Monotone Convergence Theorem,

$$\int_X f_\nu\chi_A\,\mathrm{d}\mu = \lim_{j\to\infty}\int_X f_j\chi_A\,\mathrm{d}\mu = \Big\{\int_X f\,\mathrm{d}\mu \;\Big|\; f\in\mathscr{F}_\nu\Big\} \leq \nu(A)$$

for $A\in\mathscr{A}$. Thus $f_\nu\in\mathscr{F}_\nu$.

Note that since $f_\nu\in\mathscr{F}_\nu$ it follows that

$$\overline{\nu}(A) \triangleq \nu(A) - \int_X f_\nu\chi_A\,\mathrm{d}\mu \geq 0$$

for every $A\in\mathscr{A}$. Moreover, one can readily verify that $\overline{\nu}(\emptyset) = 0$ and that $\overline{\nu}$ is countably-additive. Thus $\overline{\nu}$ is a positive measure. We claim that it is the zero measure. Suppose not. Then $\overline{\nu}(X),\mu(X)\in\mathbb{R}_{>0}$, and so there exists $\epsilon\in\mathbb{R}_{>0}$ such that $\overline{\nu}(X) > \epsilon\mu(X)$. Note that $\overline{\nu}-\epsilon\mu$ is a finite signed measure. Let (P,N) be a Hahn decomposition for this signed measure. If $A\in\mathscr{A}$ then $A\cap P\subseteq P$ and so, positivity of P gives

$$(\overline{\nu}-\epsilon\mu)(A\cap P)\geq 0.$$

Thus

$$\begin{aligned}\nu(A) &= \int_X f_\nu\chi_A\,\mathrm{d}\mu + \overline{\nu}(A) \geq \int_X f_\nu\chi_A\,\mathrm{d}\mu + \overline{\nu}(A\cap P)\\ &\geq \int_X f_\nu\chi_A\,\mathrm{d}\mu + \epsilon\mu(A\cap P) = \int_X (f_\nu+\epsilon\chi_P)\chi_A\,\mathrm{d}\mu\end{aligned}$$

and so $f_\nu+\epsilon\chi_P\in\mathscr{F}_\nu$. Note that $\mu(P)\in\mathbb{R}_{>0}$. Indeed, suppose that $\mu(P) = 0$. Then $\nu(P) = 0$ by absolute continuity of ν with respect to μ, and

$$\overline{\nu}(P) = \nu(P) - \int_X g_\nu\chi_P\,\mathrm{d}\mu = 0$$

by Proposition 2.7.11. We would then have

$$\overline{\nu}(X) - \epsilon\mu(X) = \overline{\nu}(N) - \epsilon\mu(N) \leq 0,$$

contradicting our definition of ϵ. Thus we must indeed have $\mu(P) \in \mathbb{R}_{>0}$. Therefore,

$$\int_X (f_\nu + \epsilon\chi_P)\,\mathrm{d}\mu > \int_X f_\nu\,\mathrm{d}\mu.$$

Since we have shown above that $f_\nu + \epsilon\chi_P \in \mathscr{F}_\nu$, this contradicts the definition of f_ν. Therefore, we conclude that $\overline{\nu}$ is the zero measure, and so that

$$\nu(A) = \int_X f_\nu\chi_A\,\mathrm{d}\mu$$

for every $A \in \mathscr{A}$. Now we define

$$h_\nu(x) = \begin{cases} f_\nu(x), & f_\nu(x) \in \mathbb{R}_{>0}, \\ 0, & f_\nu(x) = \infty. \end{cases}$$

By Proposition 2.7.11 we have

$$\nu(A) = \int_X h_\nu\chi_A\,\mathrm{d}\mu$$

for every $A \in \mathscr{A}$, and so this gives the existence assertion of the theorem in the case when μ and ν are finite.

Now suppose that μ and ν are σ-finite. We can use Lemma 1 from the proof of Proposition 2.3.2 to assert that there exists families $(A_j)_{j\in\mathbb{Z}_{>0}}$ and $(B_j)_{j\in\mathbb{Z}_{>0}}$ of pairwise disjoint $\mathscr{A}$-measurable sets such that

$$X = \bigcup_{j\in\mathbb{Z}_{>0}} A_j = \bigcup_{j\in\mathbb{Z}_{>0}} B_j$$

and such that $\mu(A_j), \nu(B_j) < \infty$ for each $j \in \mathbb{Z}_{>0}$. For $j \in \mathbb{Z}_{>0}$ define

$$K_j = \{k \in \mathbb{Z}_{>0} \mid B_k \cap A_j \neq \emptyset\}.$$

Thus K_j is an empty or countable set. Let us define $\mathscr{C}_j = (B_k \cap A_j)_{k\in K_j}$. Then the family $\cup_{j\in\mathbb{Z}_{>0}}\mathscr{C}_j$ is a countable (by Proposition I-1.7.16) family of pairwise disjoint measurable sets, all of which have finite measure for both μ and ν, and whose union is X. Let us write this countable family of sets as $(C_j)_{j\in\mathbb{Z}_{>0}}$, after making a choice for indexing the sets. From the proof above, for each $j \in \mathbb{Z}_{>0}$ there exists $h_{\nu,j} \in \mathsf{L}^{(0)}((X,\mathscr{A});\mathbb{R}_{\geq 0})$ such that $h_{\nu,j}(x) = 0$ for $x \in X \setminus C_j$ and such that

$$\nu(A \cap C_j) = \int_X h_{\nu,j}\chi_{A\cap C_j}\,\mathrm{d}\mu$$

for every $A \in \mathscr{A}$. Let us then define $h_\nu\colon X \to \mathbb{R}_{\geq 0}$ by asking that $h_\nu|C_j = h_{\nu,j}|C_j$. Note that h_ν is $\mathscr{A}$-measurable by Proposition 2.6.18, after noting that $h_\nu(x) = \lim_{k\to\infty}\sum_{j=1}^k h_{\nu,j}(x)$. Moreover, by the Monotone Convergence Theorem,

$$\int_X h_\nu \chi_A \,\mathrm{d}\mu = \sum_{j=1}^\infty \int_X h_{\nu,j}\chi_{A\cap C_j}\,\mathrm{d}\mu = \sum_{j=1}^\infty \nu(A\cap C_j) = \nu(A),$$

giving the existence assertion of the theorem when μ and ν are σ-finite.

For uniqueness, suppose that $h_\nu, h'_\nu \in \mathsf{L}^{(0)}((X,\mathscr{A});\mathbb{R}_{\geq 0})$ satisfy

$$\nu(A) = \int_X h_\nu\chi_A\,\mathrm{d}\mu = \int_X h'_\nu\chi_A\,\mathrm{d}\mu.$$

First we suppose that ν is finite. Then $h_\nu, h'_\nu \in \mathsf{L}^{(1)}((X,\mathscr{A},\mu);\mathbb{R}_{\geq 0})$. Let us define

$$A_{h_\nu-h'_\nu} = \{x\in X \mid h_\nu(x)-h'_\nu(x)>0\},\quad A_{h'_\nu-h_\nu} = \{x\in X \mid h'_\nu(x)-h_\nu(x)>0\}.$$

Then

$$\nu(A_{h_\nu-h'_\nu}) - \nu(A_{h_\nu-h'_\nu}) = \int_X (h_\nu - h'_\nu)\chi_{A_{h_\nu-h'_\nu}}\,\mathrm{d}\mu = 0,$$

and we conclude from Proposition 2.7.14 that $\mu(A_{h_\nu-h'_\nu}) = 0$. Similarly, $\mu(A_{h'_\nu-h_\nu}) = 0$. Since

$$X = A_{h_\nu-h'_\nu}\cup A_{h'_\nu-h_\nu}\cup\{x\in X \mid h_\nu(x) = h'_\nu(x)\},$$

we conclude that $h_\nu(x) = h'_\nu(x)$ for μ-almost every $x\in X$.

If $\nu(X)$ is not finite, but ν is σ-finite, let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a family of $\mathscr{A}$-measurable sets such that $\nu(A_j)<\infty$ for $j\in\mathbb{Z}_{>0}$ and such that $X = \cup_{j\in\mathbb{Z}_{>0}}A_j$. If

$$Z = \{x\in X \mid h_\nu(x)\neq h'_\nu(x)\}$$

then $Z = \cup_{j\in\mathbb{Z}_{>0}} Z\cap A_j$. Since $\mu(Z\cap A_j) = 0$ it follows that $\mu(Z) = 0$ by Exercise 2.3.4.

Finally, we consider the case where ν may not be σ-finite. Here the following lemma is helpful.

1 **Lemma** *If $(\mathrm{X},\mathscr{A},\mu)$ is a σ-finite measure space and if ν is a positive measure that is absolutely continuous with respect to μ, then there exist sets $\mathrm{X}_\infty, \mathrm{X}_0\in\mathscr{A}$ such that*

(i) $\mathrm{X}_\infty\cap\mathrm{X}_0 = \varnothing$,

(ii) $\mathrm{X} = \mathrm{X}_\infty\cup\mathrm{X}_0$,

(iii) $\nu|\mathscr{A}_{\mathrm{X}_0}$ *is σ-finite, and*

(iv) *if* $\mathrm{A}\in\mathscr{A}$ *satisfies* $\mu(\mathrm{A}\cap\mathrm{X}_\infty)\in\mathbb{R}_{>0}$ *then* $\nu(\mathrm{A}\cap\mathrm{X}_\infty) = \infty$.

Proof Let us first suppose that μ is finite. Let (P_j, N_j) be a Hahn decomposition for the signed measure $\nu - j\mu$, noting that this measure is consistent by finiteness of μ. Let us take $X_\infty = \cap_{j\in\mathbb{Z}_{>0}}P_j$ and $X_0 = X\setminus X_\infty$. The first two conditions of the lemma are trivially satisfied. Let $A\subseteq X_\infty$ satisfy $\mu(A)\in\mathbb{R}_{>0}$. Then, since $A\subseteq P_j$ for every $j\in\mathbb{Z}_{>0}$, $\nu(A) - j\mu(A)\geq 0$ for every $j\in\mathbb{Z}_{>0}$. Therefore, $\nu(A) = \infty$, giving the fourth condition in the lemma. Now note that

$$X_0 = X\setminus X_\infty = X\setminus(\cap_{j\in\mathbb{Z}_{>0}}P_j) = \cup_{j\in\mathbb{Z}_{>0}}X\setminus P_j = \cup_{j\in\mathbb{Z}_{>0}}N_j,$$

by De Morgan's Laws. Note that $\nu(N_j) - j\mu(N_j) \leq 0$ which implies that $\nu(N_j) < \infty$ since μ is finite. Thus X_0 is a countable union of sets whose ν-measure is finite, and so $\nu|\mathscr{A}_{X_0}$ is σ-finite. This gives the third condition of the lemma.

Now suppose that μ is σ-finite, being the union (without loss of generality disjoint by Lemma 1 from the proof of Proposition 2.3.2) of μ-finite sets $(A_j)_{j\in\mathbb{Z}_{>0}}$. By the proof above for the finite case, for each $j \in \mathbb{Z}_{>0}$ we have $\mathscr{A}$-measurable disjoint sets $X_{j,\infty} X_{j,0} \subseteq A_j$ such that

1. $A_j = X_{j,\infty} \cup X_{j,0}$,
2. $\nu|\mathscr{A}_{X_{j,0}}$ is σ-finite, and
3. $\nu(B \cap X_{j,\infty}) = \infty$ for any set $B \subseteq A_j$ for which $\mu(B \cap X_{j,\infty}) \in \mathbb{R}_{>0}$.

We can then take $X_\infty = \cup_{j\in\mathbb{Z}_{>0}} X_{j,\infty}$ and $X_0 = \cup_{j\in\mathbb{Z}_{>0}} X_{j,0}$. It is clear that X_∞ has the first two properties and the fourth property is the statement of the lemma. To verify the third, note that $X_{j,0} = \cup_{k\in\mathbb{Z}_{>0}} B_{j,k}$ for $\mathscr{A}$-measurable sets $B_{j,k} \subseteq A_j$ for which $\nu(B_{k,j} < \infty$ for $k \in \mathbb{Z}_{>0}$. Then the (countably many by Proposition I-1.7.16) sets $B_{j,k}$, $j, k \in \mathbb{Z}_{>0}$ each have finite ν-measure and their union is X_0, as desired. ▼

Now we proceed with the proof, letting X_∞ and X_0 be as in the lemma. Since $\nu|\mathscr{A}_{X_0}$ is σ-finite, our proof in the σ-finite case gives the existence of $f_\nu\colon X_0 \to \overline{\mathbb{R}}$ such that

$$\nu(A) = \int_X f_\nu \chi_A \,\mathrm{d}\mu$$

for $A \subseteq X_0$. Let us also define $g_\nu\colon X_\infty \to \overline{\mathbb{R}}_{\geq 0}$ by $g_\nu(x) = \infty$ for all $x \in X_\infty$. We then take

$$h_\nu(x) = \begin{cases} f_\nu(x), & x \in X_0, \\ g_\nu(x), & x \in X_\infty, \end{cases}$$

and one can readily check that

$$\nu(A) = \int_X h_\nu \chi_A \,\mathrm{d}\mu \tag{2.23}$$

for every $A \in \mathscr{A}$. This gives the existence assertion of the theorem in the general case. The uniqueness assertion follows from that in the case where ν is σ-finite, along with the fact that the relation (2.23) implies that h_ν must be almost everywhere equal to ∞ on X_∞. ■

Now we turn to the case of signed, complex, or vector measures that are absolutely continuous with respect to a positive measure. First let us see how absolute continuity is reflected in the Jordan decompositions

$$\nu = \nu_+ - \nu_-$$

for a signed measure,

$$\nu = (\mathrm{Re}(\nu)_+ - \mathrm{Re}(\nu)_-) + \mathrm{i}(\mathrm{Im}(nu)_+ - \mathrm{Im}(\nu)_-)$$

for a complex measure, and

$$\boldsymbol{\nu} = \sum_{j=1}^{n} (\nu_{j,+} - \nu_{j,-}) \boldsymbol{e}_j$$

for a vector measure.

2.11.4 Proposition (Characterisation of absolute continuity of signed, complex, and vector measures) *For a measure space $(\mathsf{X}, \mathscr{A}, \mu)$ the following statements hold:*

(i) if ν is a signed measure, then ν is absolutely continuous with respect to μ if and only if ν_+ and ν_- are absolutely continuous with respect to μ;

(ii) if ν is a complex measure, then ν is absolutely continuous with respect to μ if and only if $\mathrm{Re}(\nu)_+$, $\mathrm{Re}(\nu)_-$, $\mathrm{Im}(\nu)_+$, and $\mathrm{Im}(\nu)_-$ are absolutely continuous with respect to μ;

(iii) if $\boldsymbol{\nu}$ is a vector measure taking values in $\mathbb{R}^{\mathrm{n}}$, then $\boldsymbol{\nu}$ is absolutely continuous with respect to μ if and only if $\nu_{\mathrm{j},+}$ and $\nu_{\mathrm{j},-}$, $\mathrm{j} \in \{1, \dots, \mathrm{n}\}$, are absolutely continuous with respect to μ.

Proof First let μ be a signed measure. Suppose that ν is absolutely continuous with respect to μ. Let $A \in \mathscr{A}$ be such that $\mu(A) = 0$ and so $|\nu|(A) = 0$. Then, since $|\nu| = \nu_+ + \nu_-$, it immediately follows that $\mu_+(A) = \mu_-(A) = 0$. Also, if μ_+ and μ_- are absolutely continuous with respect to μ, then, if $\mu(A) = 0$ we have

$$|\nu|(A) = \nu_+(A) + \nu_-(A) = 0,$$

giving absolute continuity of ν with respect to μ.

Next we consider the case where $\boldsymbol{\nu}$ is a vector measure; the case of a complex measure is a special case of this one. Let $A \in \mathscr{A}$ satisfy $\mu(A) = 0$. First suppose that $\boldsymbol{\nu}$ is absolutely continuous with respect to μ. Let $(A_1, \dots, A_k)$ be a partition of A and note that monotonicity of μ implies that $\mu(A_j) = 0$, $j \in \{1, \dots, k\}$. Then, for $l \in \{1, \dots, n\}$ and $j \in \{1, \dots, k\}$,

$$|\nu_l(A_j)| \le \sum_{l=1}^{n} |\nu_l(A_j)| \le \sqrt{n} \|\boldsymbol{\nu}(A_j)\|_{\mathbb{R}^n} \le \sqrt{n} \|\boldsymbol{\nu}\|_{\mathbb{R}^n}(A_j) = 0,$$

where we have used Proposition II-1.1.11 and the definition of $\|\boldsymbol{\nu}\|_{\mathbb{R}^n}$. Therefore,

$$\sum_{j=1}^{k} |\nu_l(A_j)| = 0$$

for every partition $(A_1, \dots, A_k)$ of A. By Proposition 2.3.47 it follows that $|\nu_j|(A) = 0$ and so ν_l is absolutely continuous with respect to μ for each $l \in \{1, \dots, n\}$.

Finally, suppose that $\nu_1, \dots, \nu_n$ are absolutely continuous with respect to μ. As above, let $(A_1, \dots, A_k)$ be a partition of A. Since $\|\nu_l\|_{\mathbb{R}^n}(A_j) = 0$ we have $\nu_l(A_j) = 0$ for each $j \in \{1, \dots, k\}$ and $l \in \{1, \dots, n\}$. We have, for $j \in \{1, \dots, k\}$,

$$\|\boldsymbol{\nu}(A_j)\|_{\mathbb{R}^n} \le \sum_{l=1}^{n} |\nu_l(A_j)| = 0$$

using Proposition II-1.1.11. Thus

$$\sum_{j=1}^{k} \|\boldsymbol{\nu}(A_j)\|_{\mathbb{R}^n} = 0$$

for every partition $(A_1, \dots, A_k)$ of A. The definition of the variation of $\boldsymbol{\nu}$ then gives $\|\boldsymbol{\nu}\|_{\mathbb{R}^n}(A) = 0$, and so gives absolute continuity of $\boldsymbol{\nu}$ with respect to μ. ■

We can now characterise the absolutely continuous signed, complex, and vector measures.

2.11.5 Theorem (Radon–Nikodym Theorem for signed, complex, and vector measures) *If* $(\mathsf{A}, \mathscr{A}, \mu)$ *is a σ-finite measure space then the following statements hold:*

(i) for $\nu \in \mathsf{M}((\mathsf{X}, \mathscr{A}); \mathbb{R})$, ν *is absolutely continuous with respect to* μ *if and only if there exists* $\mathrm{h}_\nu \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R})$ *such that*

$$\nu(\mathsf{A}) = \int_{\mathsf{X}} \mathrm{h}_\nu \chi_{\mathsf{A}} \, \mathrm{d}\mu$$

for every $\mathsf{A} \in \mathscr{A}$;

(ii) for $\nu \in \mathsf{M}((\mathsf{X}, \mathscr{A}); \mathbb{C})$, ν *is absolutely continuous with respect to* μ *if and only if there exists* $\mathrm{h}_\nu \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{C})$ *such that*

$$\nu(\mathsf{A}) = \int_{\mathsf{X}} \mathrm{h}_\nu \chi_{\mathsf{A}} \, \mathrm{d}\mu$$

for every $\mathsf{A} \in \mathscr{A}$;

(iii) for $\boldsymbol{\nu} \in \mathsf{M}((\mathsf{X}, \mathscr{A}); \mathbb{R}^{\mathrm{n}})$, $\boldsymbol{\nu}$ *is absolutely continuous with respect to* μ *if and only if there exists* $\mathbf{h}_{\boldsymbol{\nu}} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \mu); \mathbb{R}^{\mathrm{n}})$ *such that*

$$\boldsymbol{\nu}(\mathsf{A}) = \int_{\mathsf{X}} \mathbf{h}_{\boldsymbol{\nu}} \chi_{\mathsf{A}} \, \mathrm{d}\mu$$

for every $\mathsf{A} \in \mathscr{A}$.

Moreover, the functions h_ν *and* $\mathbf{h}_{\boldsymbol{\nu}}$ *whose existence is asserted above are unique in that any other functions with the asserted properties agree almost everywhere with* h_ν *and* $\mathbf{h}_{\boldsymbol{\nu}}$.

Proof As with Theorem 2.11.3, the "if" assertions of the theorem follow from Proposition 2.7.11.

For the converse, first suppose that the finite signed measure ν is absolutely continuous with respect to μ. By Proposition 2.11.4 it follows that ν_+ and ν_- are absolutely continuous with respect to ν. By Theorem 2.11.3 there exists $h_{\nu_+}, h_{\nu_-} \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R}_{\geq 0})$ such that

$$\nu_+(A) = \int_X h_{\nu_+} \chi_A \, \mathrm{d}\mu, \quad \nu_-(A) = \int_X h_{\nu_-} \chi_A \, \mathrm{d}\mu$$

for every $A \in \mathscr{A}$. Since ν is finite, h_{ν_+} and h_{ν_-} are μ-integrable. Moreover, if $h_\nu = h_{\nu_+} - h_{\nu_-}$,

$$\int_X h_\nu \chi_A \,\mathrm{d}\mu = \nu_+(A) - \nu_-(A) = \nu(A).$$

Moreover, if $h'_{\nu_+}, h'_{\nu_-} \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R}_{\geq 0})$ also satisfy

$$\nu_+(A) = \int_X h'_{\nu_+} \chi_A \,\mathrm{d}\mu, \quad \nu_-(A) = \int_X h'_{\nu_-} \chi_A \,\mathrm{d}\mu$$

for every $A \in \mathscr{A}$, then, by Theorem 2.11.3 the sets

$$A_+ = \{x \in X \mid h_{\nu_+}(x) \neq h'_{\nu_+}(x)\}, \quad A_- = \{x \in X \mid h_{\nu_-}(x) \neq h'_{\nu_-}(x)\}$$

have μ-measure zero. Since h_ν and $h'_\nu = h'_{\nu_+} - h'_{\nu_-}$ agree on $X \setminus (A_+ \cup A_-)$, they agree μ almost everywhere, giving the theorem for finite signed measures.

To prove the theorem for a vector measure $\boldsymbol{\nu}$, note that Proposition 2.11.4 implies that the components $\nu_1, \dots, \nu_n$ are absolutely continuous with respect to μ. As in the first part of the proof, this implies the existence of functions $h_{\nu_{j,+}}, h_{\nu_{j,-}} \in \mathsf{L}^{(0)}((X, \mathscr{A}); \mathbb{R}_{\geq 0})$ such that

$$\nu_{j,+}(A) = \int_X h_{\nu_{j,+}} \chi_A \,\mathrm{d}\mu, \quad \nu_{j,-}(A) = \int_X h_{\nu_{j,-}} \chi_A \,\mathrm{d}\mu$$

for every $A \in \mathscr{A}$. If we define $\boldsymbol{h}_{\boldsymbol{\nu}}$ by

$$\boldsymbol{h}_{\boldsymbol{\nu}}(A) = (h_{\nu_{1,+}}(A) - h_{\nu_{1,-}}(A), \dots, h_{\nu_{n,+}}(A) - h_{\nu_{n,-}}(A)),$$

it immediately follows that

$$\boldsymbol{\nu}(A) = \int_X \boldsymbol{h}_{\boldsymbol{\nu}} \chi_A \,\mathrm{d}\mu.$$

The uniqueness of $\boldsymbol{h}_{\boldsymbol{\nu}}$ follows from that for $h_{\nu_{j,+}}$ and $h_{\nu_{j,-}}$, $j \in \{1, \dots, n\}$. ■

We can now give some language to the two versions of the Radon–Nikodym Theorem.

2.11.6 Definition (Radon–Nikodym derivative) Let $(X, \mathscr{A}, \mu)$ be a σ-finite measure space.

(i) If ν is a positive, finite signed, or complex measure, absolutely continuous with respect to μ, then the function h_ν from Theorem 2.11.3, Theorem 2.11.5(i), or Theorem 2.11.5(ii) is a ***Radon–Nikodym derivative*** of ν with respect to μ. This function is denoted by $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$, understanding that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ is only defined up to its value on a set of measure zero.

(ii) If $\boldsymbol{\nu}$ is a vector measure, absolutely continuous with respect to μ, then the function $\boldsymbol{h}_{\boldsymbol{\nu}}$ from Theorem 2.11.5(iii) is a ***Radon–Nikodym derivative*** of $\boldsymbol{\nu}$ with respect to μ. This function is denoted by $\frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\mu}$, understanding that $\frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\mu}$ is only defined up to its value on a set of measure zero. •

It is possible to use the Radon–Nikodym derivative to characterise the variation of a measure that is absolutely continuous with respect to another measure.

2.11.7 Proposition (Variation of absolutely continuous measures) *For a measure space* $(\mathsf{X}, \mathscr{A}, \mu)$ *the following statements hold:*

(i) *if* ν *is a finite signed or complex measure that is absolutely continuous with respect to* μ*, then*

$$|\nu|(\mathsf{A}) = \int_{\mathsf{X}} \left|\frac{\mathrm{d}\nu}{\mathrm{d}\mu}\right| \chi_{\mathsf{A}}\,\mathrm{d}\mu$$

for every $\mathsf{A} \in \mathscr{A}$*;*

(ii) *if* $\boldsymbol{\nu}$ *is a* $\mathbb{R}^{\mathrm{n}}$*-valued vector measure that is absolutely continuous with respect to* μ*, then*

$$\|\boldsymbol{\nu}\|_{\mathbb{R}^{\mathrm{n}}}(\mathsf{A}) = \int_{\mathsf{X}} \left\|\frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\mu}\right\|_{\mathbb{R}^{\mathrm{n}}} \chi_{\mathsf{A}}\,\mathrm{d}\mu$$

for every $\mathsf{A} \in \mathscr{A}$*.*

Proof It suffices to prove the theorem for vector-valued measures, since the finite signed and complex cases follow from this. First let $A \in \mathscr{A}$ and let $(A_1, \dots, A_k)$ be a partition of A. Then

$$\sum_{j=1}^{k} \|\boldsymbol{\nu}(A_j)\|_{\mathbb{R}^n} = \sum_{j=1}^{k} \left\| \int_X \frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\mu} \chi_{A_j}\,\mathrm{d}\mu \right\|_{\mathbb{R}^n} \le \sum_{j=1}^{k} \int_X \left\|\frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\mu}\right\|_{\mathbb{R}^n} \chi_{A_j}\,\mathrm{d}\mu = \int_X \left\|\frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\mu}\right\|_{\mathbb{R}^n} \chi_A\,\mathrm{d}\mu.$$

Taking the supremum of the left-hand side over all partitions gives the inequality

$$\|\boldsymbol{\nu}(A)\|_{\mathbb{R}^n} \le \int_X \left\|\frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\mu}\right\|_{\mathbb{R}^n} \chi_A\,\mathrm{d}\mu.$$

To establish the opposite inequality, we use a lemma.

1 Lemma *For a measure space* $(\mathsf{X}, \mathscr{A})$ *and for* $\mathbf{f} \in \mathsf{L}^{(0)}((\mathsf{X}, \mathscr{A}); \mathbb{R}^{\mathrm{n}})$ *there exists a sequence* $(\mathbf{f}_{\mathrm{k}})_{\mathrm{k} \in \mathbb{Z}_{>0}}$ *of* $\mathbb{R}^{\mathrm{n}}$*-valued simple functions on* X *such that*

(i) $\|\mathbf{f}_{\mathrm{k}}(\mathrm{x})\|_{\mathbb{R}^{\mathrm{n}}} \le 1$ *for each* $\mathrm{x} \in \mathsf{X}$ *and* $\mathrm{k} \in \mathbb{Z}_{>0}$ *and*

(ii) $\lim_{\mathrm{k}\to\infty} \langle \mathbf{f}_{\mathrm{k}}(\mathrm{x}), \mathbf{f}(\mathrm{x}) \rangle_{\mathbb{R}^{\mathrm{n}}} = \|\mathbf{f}(\mathrm{x})\|_{\mathbb{R}^{\mathrm{n}}}$ *for every* $\mathrm{x} \in \mathsf{X}$*.*

Proof Let us denote

$$X_0 = \{x \in X \mid f(x) = 0\}, \quad X_1 = \{x \in X \mid f(x) \neq 0\}.$$

Note that, by Proposition 2.6.11 and Corollary 2.6.14, the function $x \mapsto \frac{f(x)}{\|f(x)\|_{\mathbb{R}^n}}$ is
what measurable since the norm is continuous by . By Proposition 2.6.44, choose a sequence $(f_k)_{k \in \mathbb{Z}_{>0}}$ of simple functions with domain X_1 with the following properties:

1. $\lim_{k\to\infty} f_k(x) = \frac{f(x)}{\|f(x)\|_{\mathbb{R}^n}}$ for every $x \in X$;
2. $\|f_k(x)\|_{\mathbb{R}^n} \le \frac{f(x)}{\|f(x)\|_{\mathbb{R}^n}} = 1$ for every $x \in X$ and $k \in \mathbb{Z}_{>0}$.

For $x \in X_1$ we have

$$\lim_{k\to\infty} \langle f_k(x), f(x) \rangle_{\mathbb{R}^n} = \left\langle \frac{f(x)}{\|f(x)\|_{\mathbb{R}^n}}, f(x) \right\rangle_{\mathbb{R}^n} = \|f(x)\|_{\mathbb{R}^n},$$

what where we swap the limit and the inner product by continuity of the inner product (). We then extend each of the functions f_k, $k \in \mathbb{Z}_{>0}$, to X by taking them to be zero on X_0, so the resulting function is still a simple function. This proves the lemma. ▼

By the lemma, let $(f_k)_{k\in\mathbb{Z}_{>0}}$ be a sequence of simple functions such that

1. $\lim_{k\to\infty}\langle f_k(x), \frac{d\nu}{d\mu}(x)\rangle_{\mathbb{R}^n} = \|\frac{d\nu}{d\mu}(x)\|_{\mathbb{R}^n}$ and
2. $\|f_k(x)\|_{\mathbb{R}^n} \le 1$ for each $x \in X$ and $k \in \mathbb{Z}_{>0}$.

Let $k \in \mathbb{Z}_{>0}$ and write

$$f_k = \sum_{j=1}^{r_k} a_{k,j}\chi_{A_{k,j}}$$

for $a_{k,j} \in \mathbb{R}^n$ and pairwise disjoint sets $A_{k,j} \in \mathscr{A}$, $j \in \{1, \dots, r_k\}$. Note that $\|a_{k,j}\|_{\mathbb{R}^n} \le 1$, $j \in \{1, \dots, r_k\}$, by construction. For $A \in \mathscr{A}$ we then compute

$$\begin{aligned}\Big|\int_X \Big\langle f_k, \frac{d\nu}{d\mu}\Big\rangle_{\mathbb{R}^n}\chi_A \, d\mu\Big| &= \Big|\sum_{j=1}^{r_k}\int_X \Big\langle a_{k,j}, \frac{d\nu}{d\mu}\Big\rangle\chi_{A\cap A_{k,j}}\, d\mu\Big| = \Big|\sum_{j=1}^{r_k}\langle a_{k,j}, \nu(A\cap A_j)\rangle_{\mathbb{R}^n}\Big| \\ &\le \sum_{j=1}^{r_k}|\langle a_{k,j}, \nu(A\cap A_j)\rangle_{\mathbb{R}^n}| \le \sum_{j=1}^{r_k}\|a_{k,j}\|_{\mathbb{R}^n}\|\nu(A\cap A_{k,j})\|_{\mathbb{R}^n} \\ &\le sum_{j=1}^{r_k}\|\nu(A\cap A_{k,j})\|_{\mathbb{R}^n} \le \|\nu\|_{\mathbb{R}^n}(A),\end{aligned}$$

using the Cauchy–Bunyakovsky–Schwarz inequality and the fact that $(A\cap A_1, \dots, A\cap A_{r_k})$ is a partition of A. Thus we have, by the Dominated Convergence Theorem,

$$\int_X \Big\|\frac{d\nu}{d\mu}\Big\|_{\mathbb{R}^n}\chi_A \, d\mu = \lim_{k\to\infty}\int_X \Big\langle f_k, \frac{d\nu}{d\mu}\Big\rangle_{\mathbb{R}^n}\chi_A\, d\mu \le \|\nu\|_{\mathbb{R}^n}(A),$$

giving the result. ■

This characterisation of the variation for a signed, complex, or vector measure allows one to neatly characterise integration with respect to such measures in a convenient way. This is recorded in the following two corollaries. For the first, we comment that a signed, complex, or vector measure is obviously absolutely continuous with respect to its own variation.

2.11.8 Corollary (Radon–Nikodym derivative with respect to variation) *If* $(X, \mathscr{A})$ *is a measurable space, the following statements hold:*

(i) if ν *is a finite signed or complex measure on* $\mathscr{A}$ *then* $|\frac{d\nu}{d|\nu|}(x)| = 1$ *for almost every* $x \in X$*;*

(ii) if ν *is a* $\mathbb{R}^n$*-valued vector measure on* $\mathscr{A}$*, then* $\|\frac{d\nu}{d\|\nu\|_{\mathbb{R}^n}}\|_{\mathbb{R}^n} = 1$ *for almost every* $x \in X$.

Proof We prove this for the case of vector measures, the other two cases following from this. For $A \in \mathscr{A}$ we note that

$$\|\nu\|_{\mathbb{R}^n}(A) = \int_X \Big\|\frac{d\nu}{d\|\nu\|_{\mathbb{R}^n}}\Big\|_{\mathbb{R}^n}\chi_A \, d\|\nu\|_{\mathbb{R}^n},$$

by Proposition 2.11.7. This shows that $x \mapsto \|\frac{d\nu}{d\|\nu\|_{\mathbb{R}^n}}(x)\|_{\mathbb{R}^n}$ is a Radon–Nikodym derivative of $\|\nu\|_{\mathbb{R}^n}$ with respect to $\|\nu\|_{\mathbb{R}^n}$. However, $x \mapsto 1$ is also such a Radon–Nikodym derivative, and so by the uniqueness assertion of Theorem 2.11.5, the result follows. ■

The next result then shows that integration with respect to a signed, complex, or vector measure amounts to integration of a function, $\mathbb{C}$-valued function, or vector-valued function, respectively, with respect to a positive measure.

2.11.9 Corollary (Integration by absolutely continuous measures) *For a measurable space* $(\mathsf{X}, \mathscr{A})$*, the following statements hold:*

(i) *for a finite signed or complex measure* ν *on* $\mathscr{A}$ *and for* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \nu); \mathbb{R})$ *we have*

$$\int_{\mathsf{X}} \mathsf{f}\,\mathrm{d}\nu = \int_{\mathsf{X}} \mathsf{f}\frac{\mathrm{d}\nu}{\mathrm{d}|\nu|}\,\mathrm{d}|\nu|;$$

(ii) *for a* $\mathbb{R}^{\mathrm{n}}$*-valued vector measure* $\boldsymbol{\nu}$ *on* $\mathscr{A}$ *and for* $\mathsf{f} \in \mathsf{L}^{(1)}((\mathsf{X}, \mathscr{A}, \boldsymbol{\nu}); \mathbb{R})$ *we have*

$$\int_{\mathsf{X}} \mathsf{f}\,\mathrm{d}\boldsymbol{\nu} = \int_{\mathsf{X}} \mathsf{f}\frac{\mathrm{d}\nu}{\mathrm{d}\|\boldsymbol{\nu}\|_{\mathbb{R}^{\mathrm{n}}}}\,\mathrm{d}\|\boldsymbol{\nu}\|_{\mathbb{R}^{\mathrm{n}}}.$$

Proof It suffices to prove the result for vector measures. For $A \in \mathscr{A}$ we have

$$\boldsymbol{\nu}(A) = \int_{\mathsf{X}} \frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\|\boldsymbol{\nu}\|_{\mathbb{R}^n}}\chi_A\,\mathrm{d}\|\boldsymbol{\nu}\|_{\mathbb{R}^n},$$

by the properties of the Radon–Nikodym derivative. Thus, in the language of Proposition 2.7.65, $\boldsymbol{\nu} = (\frac{\mathrm{d}\boldsymbol{\nu}}{\mathrm{d}\|\boldsymbol{\nu}\|_{\mathbb{R}^n}}) \cdot \|\boldsymbol{\nu}\|_{\mathbb{R}^n}$. The result then follows from Proposition 2.7.66. ■

It is illustrative to see how the preceding constructions, relying as they do on the rather complicated concept of the Radon–Nikodym derivative, are used every day in introductory calculus.

2.11.10 Example (Integration by absolutely continuous measures) We consider the measure space $([0,1], \mathscr{L}([0,1]), \lambda_{[0,1]})$. Let us consider the (okay, not so difficult) problem of integrating the function $f\colon [0,1] \to \mathbb{R}$ defined by $f(x) = \sqrt{1-x^2}$. One way to perform such an integral is with a substitution $x = \sin(2\pi y)$. One then writes

$$\int_0^1 \sqrt{1-x^2}\,\mathrm{d}x = \int \sqrt{1-\sin^2(2\pi y)}\frac{\mathrm{d}x}{\mathrm{d}y}\,\mathrm{d}y$$

2.11.2 Singular measures

2.11.3 Notes

The idea of the Radon–Nikodym Theorem, Theorem 2.11.3, we give is based on that of [Wilansky 1989], with the extension to the case where ν is not σ finite following the lemma of [Schep 2003].

Section 2.12

Measures on locally compact topological spaces

In this section we prove an important theorem about the *set* of measures on certain sorts of sets, locally compact topological spaces which we discussed in Section 1.11. The characterisation of measures we provide is often very useful when dealing with these objects, and moreover easily permits the extension to these spaces, and measures on these spaces, of concepts of Fourier theory such as will be discussed in Chapters IV-5, IV-6, and IV-7. We shall see specific instances of this in Sections IV-5.7 and IV-6.6. is this all?

The presentation in this section proceeds, like much in this chapter, from the general to the specific. While it is true that it is easy to carry out the program on, for example, the locally compact topological space $\mathbb{R}$, the fact is that the story is complicated in all cases, so it is best to treat specific cases as instances of a general one to avoid repetition of somewhat complex arguments.

Do I need to read this section? The material in this section is specialised and can be skipped until needed.

2.12.1 Haar measure on a locally compact topological group

2.12.2 The case of $\mathbb{T}$

2.12.3 The case of $\mathbb{R}$

In this section we consider the norm topology of the set of functions of bounded variation defined on a compact interval.

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Now we define a proposed norm $\|\cdot\|_{\mathrm{BV}}$ on $\mathrm{BV}([a,b];\mathbb{F})$ by $\|f\|_{\mathrm{BV}} = |f(a)| + \mathrm{TV}(f)$. We claim that this makes the set of functions of bounded variation a Banach space.

2.12.1 Theorem (BV([a, b]; $\mathbb{F}$) is a Banach space) BV([a, b]; $\mathbb{F}$) *is a Banach space.*

Proof First let $\lambda \in \mathbb{F}$ and $f \in \mathrm{BV}([a,b];\mathbb{F})$ and, for a partition with endpoints $(x_0, x_1, \dots, x_k)$, compute

$$|\lambda f(a)| + \sum_{j=1}^{k} |\lambda f(x_j) - \lambda f(x_{j-1})| = |\lambda| \Big(|f(a)| + \sum_{j=1}^{k} |f(x_j) - f(x_{j-1})| \Big).$$

Taking the supremum over all partitions of $[a,b]$ gives $\|\lambda f\|_{\mathrm{BV}} = |\lambda| \|f\|_{\mathrm{BV}}$. We also have, for $f, g \in \mathsf{BV}([a,b];\mathbb{F})$ and a partition with endpoints $(x_0, x_1, \dots, x_k)$,

$$|(f+g)(a)| + \sum_{j=1}^{k} |(f+g)(x_j) - (f+g)(x_{j-1})| \\ \leq |f(a)| + |g(a)| + \sum_{j=1}^{k} |f(x_j) - f(x_{j-1})| + \sum_{j=1}^{k} |g(x_j) - g(x_{j-1})|.$$

Taking the supremum over all partitions of $[a,b]$ gives $\|f+g\|_{\mathsf{BV}} \leq \|f\|_{\mathsf{BV}} + \|g\|_{\mathsf{BV}}$. Finally, it is clear that $\|f\|_{\mathsf{BV}} \geq 0$ for all $f \in \mathsf{BV}([a,b];\mathbb{F})$, and that $\|0\|_{\mathsf{BV}} = 0$. Moreover, suppose that $\|f\|_{\mathsf{BV}} = 0$. Then it follows that $|f(a)| = 0$ and $\mathrm{TV}\, f = 0$. From the definition of total variation, it follows easily that $\mathrm{TV}(f) > 0$ if f is not constant, and so we conclude that f is constant, and therefore takes is value at a, which is zero. This all shows that $\|\cdot\|_{\mathsf{BV}}$ is a norm.

Now let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathsf{BV}([a,b];\mathbb{F})$. We claim that, for each $x \in [a,b]$, the sequence $(f_j(x))_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathbb{R}$. Let $x \in [a,b]$, let $\epsilon > 0$, and let P be a partition with $\mathrm{EP}(P) = (a, x, b)$. Choose $N \in \mathbb{Z}_{>0}$ sufficiently large that

$$|(f_j - f_k)(a)| + \mathrm{TV}(f_j - f_k) < \epsilon$$

for $j, k \geq N$. This immediately gives $|(f_j - f_k)(a)| < \epsilon$. We also have, corresponding to the partition P,

$$|(f_j - f_k)(x) - (f_j - f_k)(a)| + |(f_j - f_k)(b) - (f_j - f_k)(x)| \leq \mathrm{TV}(f_j - f_k), \qquad j, k \geq N,$$

implying that

$$||(f_j - f_k)(x)| - |(f_j - f_k)(a)|| \leq \mathrm{TV}(f_j - f_k), \qquad j, k \geq N,$$

by Exercise I-2.2.8. In particular this gives

$$|(f_j - f_k)(x)| \leq |(f_j - f_k)(a)| + \mathrm{TV}(f_j - f_k) < \epsilon, \qquad j, k \geq N.$$

This shows that, not only is $(f_j)_{j\in\mathbb{Z}_{>0}}$ a Cauchy sequence for each $x \in [a,b]$, but that the sequence is Cauchy uniformly in x. Now let $f\colon [a,b] \to \mathbb{F}$ be the limit function for the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ and let $(x_0, x_1, \dots, x_k)$ be the endpoints of a partition of $[a,b]$. Let $N \in \mathbb{Z}_{>0}$ have the property that, $|f_j(x) - f(x)| < \frac{\epsilon}{2k}$ for $j \geq N$. Then we compute

$$\begin{aligned} \sum_{j=1}^{k} |f(x_j) - f(x_{j-1})| &\leq \sum_{j=1}^{k} \Big(|f(x_j) - f_N(x_j)| + |f(x_{j-1}) - f_N(x_{j-1})| + |f_N(x_j) - f_N(x_{j-1})| \Big) \\ &\leq \sum_{j=1}^{k} \Big(|f_N(x_j) - f_N(x_{j-1})| + \tfrac{\epsilon}{k} \Big) \\ &\leq \mathrm{TV}(f_N) + \epsilon. \end{aligned}$$

Since this holds for any partition P and any $\epsilon > 0$, we have $\mathrm{TV}(f) \leq \mathrm{TV}(f_N)$. Thus $f \in \mathsf{BV}([a,b];\mathbb{F})$, giving completeness as desired. ■

Let us also record some of the properties of the Banach space $\mathsf{BV}([a,b];\mathbb{F})$.

2.12.2 Proposition (Properties of BV([a, b]; 𝔽)) BV([a, b]; 𝔽) *is not separable.*

Proof We take the case when $\mathbb{F} = \mathbb{R}$ since this implies the case when $\mathbb{F} = \mathbb{C}$. For $c \in (a, b)$ let $f_c\colon [a, b] \to \mathbb{R}$ be defined by

$$f_c(x) = \begin{cases} 1, & x = c, \\ 0, & x \neq c. \end{cases}$$

By a direct computation one can see that $\|f_c\|_{\mathsf{BV}} = 2$ for $c \in (a, b)$ and that $\|f_{c_1} - f_{c_2}\|_{\mathsf{BV}} = 4$ provided that $c_1 \neq c_2$. For $c \in (a, b)$ note that $f \in \mathsf{B}(1, f_c)$ implies that

$$|\|f\|_{\mathsf{BV}} - \|f_c\|_{\mathsf{BV}}| \leq \|f - f_c\|_{\mathsf{BV}} \quad\Longrightarrow\quad |\|f\|_{\mathsf{BV}} - 2| \leq 1 \quad\Longrightarrow\quad \|f\|_{\mathsf{BV}} \leq 3,$$

using Exercise 3.1.3. Thus, for each $c \in (a, b)$, $\mathsf{B}(1, f_c) \subseteq \mathsf{B}(3, 0_{\mathsf{BV}([a,b];\mathbb{F})})$. Moreover, note that if $f_1 \in \mathsf{B}(1, f_{c_1})$ and $f_2 \in \mathsf{B}(1, f_{c_2})$ for $c_1 \neq c_2$, then

$$\|f_1 - f_{c_2}\|_{\mathsf{BV}} \geq |\|f_1 - f_{c_1}\|_{\mathsf{BV}} - \|f_{c_2} - f_{c_1}\|_{\mathsf{BV}}| \geq 3,$$

using Proposition 1.1.3. Thus $f_1 \notin \mathsf{B}(1, f_{c_2})$, and one similarly shows that $f_2 \notin \mathsf{B}(1, f_{c_1})$. Therefore, $\mathsf{B}(3, 0_{\mathsf{BV}([a,b];\mathbb{F})})$ contains the uncountable collection $(\mathsf{B}(1, f_c))_{c \in (a,b)}$ of disjoint open balls. In particular, if $(f_j)_{j \in \mathbb{Z}_{>0}}$ is any countable subset of $\mathsf{BV}([a, b]; \mathbb{R})$, then a countable number of the open balls $(\mathsf{B}(1, f_c))_{c \in (a,b)}$ can contain a function from the set $(f_j)_{j \in \mathbb{Z}_{>0}}$. Let the set of such open balls be denoted by $(\mathsf{B}(1, f_{c_\alpha}))_{\alpha \in A}$. Note that $\mathrm{cl}((f_j)_{j \in \mathbb{Z}_{>0}}) \subseteq \cup_{\alpha \in A} \overline{\mathsf{B}}(1, f_{c_\alpha})$. This precludes points in the balls

$$\{\mathsf{B}(1, f_c) \mid c \notin \{c_\alpha \mid \alpha \in A\}\}$$

from lying in $\mathrm{cl}((f_j)_{j \in \mathbb{Z}_{>0}})$. Thus any countable subset of $\mathsf{BV}([a, b]; \mathbb{R})$ is not dense. ■

We comment that there are other possible norms one can use on $\mathsf{BV}([a, b]; \mathbb{F})$ that are separable; we refer the reader to Section 3.8.10 for a few words on this, as well as references.

In the course of our discussion of the Riemann–Stieltjes integral in Section I-3.5, and particularly during the course of the constructions surrounding Theorem I-3.5.11, we ran into the idea of bounded linear functions on the Banach space $(\mathsf{C}^0([a, b]; \mathbb{R}), \|\cdot\|_\infty)$. In this section we complete the discussion that was started there by establishing a Banach isomorphism between the dual of $(\mathsf{C}^0([a, b]; \mathbb{R}), \|\cdot\|_\infty)$ and a closed subspace of $\mathsf{BV}([a, b]; \mathbb{R})$.

The first thing we need to do is define the appropriate subspace of $\mathsf{BV}([a, b]; \mathbb{R})$. The first step is an equivalence relation. Two functions $\varphi_1, \varphi_2 \in \mathsf{BV}([a, b]; \mathbb{R})$ are ***equivalent***, and we write $\varphi_1 \sim \varphi_2$, if

$$\int_a^b f(x)\,\mathrm{d}\varphi_1(x) = \int_a^b f(x)\,\mathrm{d}\varphi_2(x)$$

for every $f \in \mathsf{C}^0([a, b]; \mathbb{R})$. The following result describes the equivalence classes in $\mathsf{BV}([a, b]; \mathbb{R})$.

2.12.3 Lemma *Let $\varphi_1, \varphi_2 \in \mathsf{BV}([a,b];\mathbb{R})$. Then $\varphi_1 \sim \varphi_2$ if and only if*

(i) $\varphi_1(a) - \varphi_2(a) = \varphi_1(b) - \varphi_2(b)$ *and*

(ii) $\varphi_1(t+) - \varphi_2(t+) = \varphi_1(t-) - \varphi_2(t-) = \varphi_1(a) - \varphi_2(a)$ *for all* $t \in (a,b)$.

Proof By Proposition I-3.5.24 it suffices to determine the equivalence class of the zero function in $\mathsf{BV}([a,b];\mathbb{R})$.

First suppose that

$$\int_a^b f(x)\,\mathrm{d}\varphi(x) = 0$$

for all $f \in \mathsf{C}^0([a,b];\mathbb{R})$. Taking f to be the function $f(x) = 1$ for all $x \in [a,b]$ we have, by Exercise I-3.5.1,

$$\int_a^b \mathrm{d}\varphi(x) = \varphi(b) - \varphi(a),$$

giving (i). For $\xi \in [a,b)$ and $\epsilon > 0$ sufficiently small that $\xi + \epsilon < b$ define $f_\epsilon \in \mathsf{C}^0([a,b];\mathbb{R})$ by

$$f_{\xi,\epsilon}(x) = \begin{cases} 1, & x \in [a,\xi], \\ 1 - \frac{x-\xi}{\epsilon}, & x \in (\xi, \xi+\epsilon), \\ 0, & x \in [\xi+\epsilon, b]. \end{cases}$$

Now compute

$$\begin{aligned} \int_a^b f_{\xi,\epsilon}(x)\,\mathrm{d}\varphi(x) &= \varphi(\xi) - \varphi(a) + \int_\xi^{\xi+\epsilon} f_{\xi,\epsilon}(x)\,\mathrm{d}\varphi(x) \\ &= \varphi(\xi) - \varphi(a) - \varphi(\xi) + \frac{1}{\epsilon}\int_\xi^{\xi+\epsilon} \varphi(x)\,\mathrm{d}x, \end{aligned}$$

using integration by parts. We next claim that

$$\lim_{\epsilon\downarrow 0} \frac{1}{\epsilon}\int_\xi^{\xi+\epsilon} \mathrm{d}\varphi(x) = \varphi(\xi+).$$

Indeed, if we define $\tilde{\varphi}$ by

$$\tilde{\varphi}(x) = \begin{cases} \varphi(x), & x \in (\xi, b], \\ \varphi(\xi+), & x \in [a,\xi], \end{cases}$$

then ξ is a point of continuity for $\tilde{\varphi}$, and so by Theorem I-3.4.30,

$$\lim_{\epsilon\downarrow 0} \frac{1}{\epsilon}\int_\xi^{\xi+\epsilon} \varphi(x)\,\mathrm{d}x = \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon}\int_\xi^{\xi+\epsilon} \tilde{\varphi}(x)\,\mathrm{d}x = \tilde{\varphi}(\xi) = \varphi(\xi+).$$

Thus

$$\int_a^b f_{\xi,\epsilon}(x)\,\mathrm{d}\varphi(x) = \varphi(\xi+) - \varphi(a),$$

showing that $\varphi(\xi+) = \varphi(a)$. An altogether similar argument gives $\varphi(\xi-) = \varphi(b) = \varphi(a)$ for $\xi \in (a,b]$. Thus (ii) holds.

Now suppose that (i) and (ii) hold. Then at all points of continuity, i.e., almost everywhere by Theorem I-3.3.3(v), $\varphi(x) = \varphi(a)$. Then it is clear from the definition of the Riemann–Stieltjes integral that $\int_a^b f(x)\,\mathrm{d}\varphi(x) = 0$ for every $f \in \mathsf{C}^0([a,b];\mathbb{R})$. ■

We now provide a means of identifying a unique representative of each equivalence class.

2.12.4 Definition (Normalised functions in BV([a, b]; ℝ)) A function $\varphi \in \mathsf{BV}([a,b];\mathbb{R})$ is ***normalised*** if

(i) $\varphi(a) = 0$ and

(ii) $\varphi(x) = \varphi(x+)$ for each $x \in [a,b)$.

The set of normalised functions in $\mathsf{BV}([a,b];\mathbb{R})$ is denoted by $\overline{\mathsf{BV}}([a,b];\mathbb{R})$. •

2.12.5 Lemma *If $\varphi \in \mathsf{BV}([a,b];\mathbb{R})$ and if we define $\overline{\varphi}\colon [a,b] \to \mathbb{R}$ by*

$$\overline{\varphi}(x) = \begin{cases} 0, & x = a, \\ \varphi(x+) - \varphi(a), & x \in (a,b), \\ \varphi(b) - \varphi(a), & x = b, \end{cases}$$

then $\overline{\varphi} \in \overline{\mathsf{BV}}([a,b];\mathbb{R})$. Moreover, the following statements hold:

(i) for $\varphi \in \mathsf{BV}([a,b];\mathbb{R})$, $\overline{\varphi}$ is the unique element of $\overline{\mathsf{BV}}([a,b];\mathbb{R})$ for which $\varphi \sim \overline{\varphi}$;

(ii) $\mathrm{TV}(\overline{\varphi}) \le \mathrm{TV}(\varphi)$.

Proof That $\overline{\varphi} \in \overline{\mathsf{BV}}([a,b];\mathbb{R})$ is a simple matter of verifying the definition. That $\overline{\varphi}$ is the *unique* element of $\overline{\mathsf{BV}}([a,b];\mathbb{R})$ for which $\varphi \sim \overline{\varphi}$ follows from the fact that any normalised function from $\mathsf{BV}([a,b];\mathbb{R})$ that is equivalent to the zero function is the zero function. It remains, then, to prove the last assertion. Let $(x_0, x_1, \dots, x_k)$ be the endpoints of a partition of $[a,b]$ and, for $\epsilon > 0$, choose $y_j > x_j$, $j \in \{1, \dots, k-1\}$, such that φ is continuous at y_k and such that

$$|\varphi(x_j+) - \varphi(y_j)| < \tfrac{\epsilon}{2k}.$$

This is possible since the set of discontinuities of φ is countable by Theorem I-3.3.3(v) (why does this follow?). Also take $y_0 = a$ and $y_k = b$. Then we have

$$\begin{aligned}\sum_{j=1}^{k} |\overline{\varphi}(x_j) - \overline{\varphi}(x_{j-1})| &\le \sum_{j=1}^{k} \big(|\overline{\varphi}(x_j) - \varphi(y_j)| + |\overline{\varphi}(x_{j-1}) - \varphi(y_{j-1})| + |\varphi(y_j) - \varphi(y_{j-1})|\big) \\ &\le \sum_{j=1}^{k} |\varphi(y_j) - \varphi(y_{j-1})| + \epsilon.\end{aligned}$$

Since $\epsilon > 0$ and the partition are arbitrary, it follows that $\mathrm{TV}(\overline{\varphi}) \le \mathrm{TV}(\varphi)$, as desired. ■

2.12.6 Theorem ($(\mathsf{C}^0([a,b];\mathbb{R}))^*$) *The map* $\mathsf{L}\colon \overline{\mathsf{BV}}([\mathrm{a},\mathrm{b}];\mathbb{R}) \to (\mathsf{C}^0([\mathrm{a},\mathrm{b}];\mathbb{R}))^*$ *defined by*

$$\mathsf{L}(\overline{\varphi})\cdot \mathrm{f} = \int_{\mathrm{a}}^{\mathrm{b}} \mathrm{f}(\mathrm{x})\mathrm{d}\overline{\varphi}(\mathrm{x})$$

is a Banach isomorphism of $(\overline{\mathsf{BV}}([\mathrm{a},\mathrm{b}];\mathbb{R}),\overline{\mathrm{TV}})$ *with the topological dual of* $(\mathsf{C}^0([\mathrm{a},\mathrm{b}];\mathbb{R}),\|\cdot\|_\infty)$.

Proof Let $|||\cdot|||_\infty$ denote the norm on $(\mathsf{C}^0([a,b];\mathbb{R}))^*$ induced by the norms on $\|\cdot\|_\infty$ and $|\cdot|$. We also note that, by Theorem I-3.5.18, $f\in\mathsf{C}^0([a,b];\mathbb{R})$ is Riemann–Stieltjes integrable with respect to $\varphi\in\mathsf{BV}([a,b];\mathbb{R})$. So we do not have to prove that L is well-defined.

Let us first show that L is continuous. For a partition with endpoints $(x_0,x_1,\dots,x_k)$ we have

$$\begin{aligned}|\mathsf{L}(\varphi)(f)| &= \Big|\int_a^b f(x)\,\mathrm{d}\varphi(x)\Big| \le \int_a^b |f(x)|\,\mathrm{d}\varphi(x)\\ &\le \|f\|_\infty\int_a^b \mathrm{d}\varphi(x) \le \|f\|_\infty\sum_{j=1}^k(\varphi(x_j)-\varphi(x_{j-1}))\\ &\le \|f\|_\infty\sum_{j=1}^k|\varphi(x_j)-\varphi(x_{j-1})| \le \|f\|_\infty\overline{\mathrm{TV}}(\varphi).\end{aligned}$$

Therefore,

$$|||\mathsf{L}(\varphi)|||_\infty = \sup\Big\{\frac{|\mathsf{L}(\varphi)(f)|}{\|f\|_\infty}\ \Big|\ f\in\mathsf{C}^0([a,b];\mathbb{R}),\ \|f\|_\infty\ne 0\Big\} \le \overline{\mathrm{TV}}(\varphi),$$

giving continuity of L with the induced norm of L not exceeding 1.

Next let us show that L is surjective. Thus let $\alpha\in(\mathsf{C}^0([a,b];\mathbb{R}))^*$. By the Hahn–Banach Theorem we can extend α to a continuous linear function β_α on $\mathsf{L}^\infty([a,b];\mathbb{R})$ which satisfies $|||\beta_\alpha|||_\infty = |||\alpha|||_\infty$; we use the symbol $|||\cdot|||_\infty$ to denote the induced norm on $(\mathsf{L}^\infty([a,b];\mathbb{R}))^*$. For $\xi\in[a,b]$ define $g_\xi\in\mathsf{L}^\infty([a,b];\mathbb{R})$ by

$$g_\xi(x) = \begin{cases}1, & x\in[a,\xi],\\ 0, & x\in(\xi,b].\end{cases}$$

Define $\varphi_\alpha\colon[a,b]\to\mathbb{R}$ by $\varphi_\alpha(x)=\beta_\alpha(g_x)$. We will show that φ_α has bounded variation and that

$$\int_a^b f(x,\mathrm{d}\varphi_\alpha(x) = \alpha(f), \qquad f\in C^0([a,b];\mathbb{R}). \tag{2.24}$$

First we show that φ_α has bounded variation. Let $(x_0,x_1,\dots,x_k)$ be the endpoints of a partition P of $[a,b]$ and define a function $h_P\in\mathsf{L}^\infty([a,b];\mathbb{R})$ by

$$h_P(x) = \begin{cases}\operatorname{sign}(\varphi_\alpha(x_1)-\varphi_\alpha(x_0)), & x\in[x_0,x_1],\\ \operatorname{sign}(\varphi_\alpha(x_j)-\varphi_\alpha(x_{j-1})), & x\in(x_{j-1},x_j].\end{cases}$$

Note that, by the definition of the functions g_ξ, $\xi \in [a,b]$ and of the function φ_α we have

$$h_P(x) = \sum_{j=1}^{k} \operatorname{sign}(\varphi_\alpha(x_j) - \varphi_\alpha(x_{j-1}))(g_{x_j} - g_{x_{j-1}}).$$

Therefore,

$$\left|\!\left|\!\left| \sum_{j=1}^{k} \operatorname{sign}(\varphi_\alpha(x_j) - \varphi_\alpha(x_{j-1}))(g_{x_j} - g_{x_{j-1}}) \right|\!\right|\!\right|_\infty \le 1.$$

Now compute

$$\begin{aligned}
\sum_{j=1}^{k} |\varphi_\alpha(x_j) - \varphi_\alpha(x_{j-1})| &= \sum_{j=1}^{k} \operatorname{sign}(\varphi_\alpha(x_j) - \varphi_\alpha(x_{j-1}))(\varphi_\alpha(x_j) - \varphi_\alpha(x_{j-1})) \\
&= \sum_{j=1}^{k} \operatorname{sign}(\varphi_\alpha(x_j) - \varphi_\alpha(x_{j-1}))(\beta_\alpha(g_{x_j}) - \beta_\alpha(g_{x_{j-1}})) \\
&= \beta_\alpha\Big(\sum_{j=1}^{k} \operatorname{sign}(\varphi_\alpha(x_j) - \varphi_\alpha(x_{j-1}))(g_{x_j} - g_{x_{j-1}})\Big) \\
&= |||\beta_\alpha|||_\infty \left|\!\left|\!\left| \sum_{j=1}^{k} \operatorname{sign}(\varphi_\alpha(x_j) - \varphi_\alpha(x_{j-1}))(g_{x_j} - g_{x_{j-1}}) \right|\!\right|\!\right|_\infty \\
&\le |||\beta_\alpha|||_\infty.
\end{aligned}$$

This shows that $\mathrm{TV}(\varphi_\alpha) \le |||\beta_\alpha|||_\infty = |||\alpha|||_\infty$. Thus φ_α has bounded variation.

Now we show that (2.24) holds. For $k \in \mathbb{Z}_{>0}$ define a partition of $[a,b]$ with endpoints $(x_0, x_1, \dots, x_k)$ defined by $x_j = a + \frac{j(b-a)}{k}$, $j \in \{0, 1, \dots, k\}$. Thus the endpoints are evenly spaced. Now, for $f \in \mathsf{C}^0([a,b];\mathbb{R})$ define $f_k \in \mathsf{L}^\infty([a,b];\mathbb{R})$ by

$$f_k = \sum_{j=1}^{k} f(x_j)(g_{x_j} - g_{x_{j-1}}).$$

We claim that $\lim_{k\to\infty} f_k(x) = f(x)$ for all $x \in [a,b]$. Indeed, let $\epsilon > 0$ and choose $\delta > 0$ sufficiently small that, for each $x \in [a,b]$, $y \in \mathsf{B}(\delta, x) \cap [a,b]$ implies that $f(y) \in \mathsf{B}(\epsilon, f(x))$. This is possible since f is uniformly continuous by the Heine–Cantor Theorem. Now, if $N \in \mathbb{Z}_{>0}$ is chosen such that $\frac{b-a}{N} < \delta$, then we have $|f(x) - f_k(x)| < \epsilon$ for each $k \ge N$, giving our claim. Now, using we compute

commuting evaluation and limit

$$\begin{aligned}\alpha(f) &= \lim_{k\to\infty} \beta_\alpha(f_k) = \lim_{k\to\infty} \beta_\alpha\Big(\sum_{j=1}^k f(x_j)(g_{x_j} - g_{x_{j-1}})\Big)\\ &= \lim_{k\to\infty} \sum_{j=1}^k f(x_j)(\beta_\alpha(g_{x_j}) - \beta_\alpha(g_{x_{j-1}}))\\ &= \lim_{k\to\infty} \sum_{j=1}^k f(x_j)(\varphi_\alpha(x_j) - \varphi_\alpha(x_{j-1}))\\ &= \int_a^b f(x)\,\mathrm{d}\varphi_\alpha(x),\end{aligned}$$

as desired.

Note that φ_α may not be an element of $\overline{\mathsf{BV}}([a,b];\mathbb{R})$. We will now construct $\overline{\varphi}_\alpha \in \overline{\mathsf{BV}}([a,b];\mathbb{R})$ and show that $\mathsf{L}(\overline{\varphi}_\alpha) = \alpha$. We construct $\overline{\varphi}_\alpha \in \overline{\mathsf{BV}}([a,b];\mathbb{R})$ from φ_α as in Lemma 2.12.5. By Lemma 2.12.3

$$\int_a^b f(x)\,\mathrm{d}\varphi_\alpha - \overline{\varphi}_\alpha(x) = 0$$

for every $f \in \mathsf{C}^0([a,b];\mathbb{R})$. Therefore, $\mathsf{L}(\overline{\varphi}_\alpha) = \alpha$, and so L is surjective as desired.

Next we show the injectivity of L. Suppose that for $\overline{\varphi} \in \overline{\mathsf{BV}}([a,b];\mathbb{R})$ we have $\mathsf{L}(\overline{\varphi}) = 0$. By Lemma 2.12.3 and the fact that a normalised function equivalent to zero is itself zero, we then have $\varphi(x) = 0$ for every $x \in [a,b]$. Thus L is injective.

Thus we have shown that L is a continuous bijection. Moreover, during the course of the proof we have shown that $|||\mathsf{L}(\varphi)|||_\infty \le \overline{\mathrm{TV}}(\phi)$ and that $|||\mathsf{L}(\varphi)|||_\infty \ge \overline{\mathrm{TV}}(\phi)$. Thus L is also norm-preserving. ■

2.12.4 The case of $\mathbb{R}^n$

Chapter 3

Banach spaces

In Chapter I-5, particularly in Sections I-5.4 and I-5.8, we studied linear algebra over arbitrary fields. Here we relied on the notion, introduced in Section I-4.5, of a vector space. In many instances in applications, one is interested in the case where the field is either $\mathbb{R}$ or $\mathbb{C}$. In finite-dimensions, the story here is not too complicated; finite-dimensional vector spaces over $\mathbb{R}$ or $\mathbb{C}$ are fairly easy to understand and linear maps on these spaces are also fairly easy to understand. However, in applications, it turns out that infinite-dimensional vector spaces are often what is of most interest. We make no attempt to motivate this here, but refer the reader to Chapter IV-1. The reader will note that we were careful to understand the algebra of infinite-dimensional vector spaces in Section I-4.5 and linear maps between them in Section I-5.4. It turns out, though, that the key to understanding the infinite-dimensional vector spaces that arise in applications is through the various topologies one can put on these. This is the genesis of the huge subject of topological vector spaces which we spend the next three chapters introducing. The present chapter is devoted to topologies defined by a "norm." These are the most basic topologies, and suffice to cover many, but by no means all, areas of application.

Certain parts of what we say in this chapter have already been accounted for in Chapter 1. However, we it seems like a good idea to make the treatment here independent, for the most part, of the more general and abstract treatment in Chapter 1. Therefore, at the cost of repetitiveness we make treat all of the topological ideas for normed vector spaces independently of the fact that we have already considered them.

Do I need to read this chapter? This chapter is fundamental to understanding in any rigorous way topics like Fourier series, Fourier transforms, linear system theory, signal processing, etc. This makes at least the basic material in this chapter essential reading. Perhaps a reading of the detailed examples of dual spaces in Section 3.9 can be postponed until it is needed, although it is at least interesting. •

Contents

3.1 Definitions and properties of normed vector spaces 352

3.1.1 Norms and seminorms . . . 352
3.1.2 Open and closed subsets of normed vector spaces . . . 358
3.1.3 Subspaces, direct sums, and quotients . . . 363
Exercises . . . 365
3.2 Sequences in normed vector spaces . . . 367
3.2.1 Definitions and properties of sequences . . . 367
3.2.2 Algebraic operations on sequences . . . 369
3.2.3 Multiple sequences . . . 370
Exercises . . . 370
3.3 Completeness and completions . . . 371
3.3.1 Completeness (Banach spaces) . . . 371
3.3.2 Why completeness is important . . . 375
3.3.3 Completeness and direct sums and quotients . . . 377
3.3.4 Completions . . . 378
Exercises . . . 381
3.4 Series in normed vector spaces . . . 383
3.4.1 Definitions and properties of series . . . 383
3.4.2 Absolute and unconditional convergence . . . 385
3.4.3 Algebraic operations on series . . . 395
3.4.4 Multiple series . . . 397
3.4.5 Cesàro convergence of sequences and series . . . 398
3.4.6 Series in normed vector spaces with arbitrary index sets . . . 400
3.4.7 Notes . . . 405
Exercises . . . 405
3.5 Continuous maps between normed vector spaces . . . 407
3.5.1 General continuous maps between normed vector spaces . . . 407
3.5.2 Continuous linear maps between normed vector spaces . . . 410
3.5.3 Induced topologies on continuous linear maps . . . 416
3.5.4 Invertibility of continuous linear maps . . . 423
3.5.5 Spectral properties for continuous operators on Banach spaces . . . 425
3.5.6 The Open Mapping Theorem and Closed Graph Theorem . . . 428
Exercises . . . 429
3.6 Topology of normed vector spaces . . . 430
3.6.1 Properties of balls in normed vector spaces . . . 430
3.6.2 Interior, closure, boundary, etc. . . . 431
3.6.3 Compactness . . . 436
3.6.4 Closed subspaces . . . 439
3.6.5 Bases for normed vector spaces . . . 448
3.6.6 Notes . . . 450
Exercises . . . 451
3.7 Split subspaces? . . . 452
3.7.1 Split subspaces . . . 452
3.7.2 Questions and answers about closed and split subspaces . . . 453
3.7.3 Projection constants . . . 454
3.7.4 The proof of the Split Subspace Theorem . . . 454
3.7.5 Notes . . . 454

3.8 Examples of Banach spaces . . . 455
3.8.1 The p-norms on $\mathbb{F}^n$ . . . 455
3.8.2 Banach spaces of sequences . . . 463
3.8.3 Banach spaces of direct sums . . . 472
3.8.4 Banach spaces of continuous functions on $\mathbb{R}$ . . . 477
3.8.5 Banach spaces of continuous functions on metric spaces . . . 485
3.8.6 Banach spaces of continuous functions on locally compact topological spaces . . . 485
3.8.7 Banach spaces of integrable functions on $\mathbb{R}$ . . . 485
3.8.8 Banach spaces of integrable functions on measure spaces . . . 497
3.8.9 Banach spaces of measures . . . 497
3.8.10 Notes . . . 500
Exercises . . . 500
3.9 Topological duals of Banach spaces . . . 501
3.9.1 Reflexivity in Banach spaces . . . 502
3.10 Examples of duals of Banach spaces . . . 503
3.10.1 The dual of $\mathsf{L}^p(I;\mathbb{F})$ . . . 503

Section 3.1

Definitions and properties of normed vector spaces

The basic ingredient in this chapter is a norm on a vector space. While it is possible to introduce this notion for other classes of fields, we restrict our attention to vector spaces over either $\mathbb{R}$ or $\mathbb{C}$. It will often be convenient to be able to consider both of these cases together, and so let us introduce some notation for doing this.

3.1.1 Notation ($\mathbb{F}$) The symbol $\mathbb{F}$ will denote either $\mathbb{R}$ or $\mathbb{C}$. That is to say, whenever the symbol $\mathbb{F}$ is present, the statement can be read by replacing it with either $\mathbb{R}$ or $\mathbb{C}$. In order to use this convenient notation as much as possible we have the following conventions.

(i) If $\mathbb{F} = \mathbb{R}$ and if $a \in \mathbb{F}$ then $|a|$ denotes the absolute value of a.

(ii) If $\mathbb{F} = \mathbb{C}$ and if $a \in \mathbb{F}$ then $|a|$ denotes the modulus of a.

(iii) If $\mathbb{F} = \mathbb{R}$ and if $a \in \mathbb{F}$ then $\bar{a} = a$.

(iv) If $\mathbb{F} = \mathbb{C}$ if $a \in \mathbb{F}$ then $\bar{a}$ is the complex conjugate of a. •

Do I need to read this section? Accepting that normed vector spaces are important (they are), this section must then be important. •

3.1.1 Norms and seminorms

In this section we consider norms and seminorms. While the notion of a norm is the most important for us, we will see that seminorms come up in two natural ways. One is in Section 3.8.8 when we give an extremely important class of normed vector spaces. As we shall see, in the construction of this class it is natural to first define a seminorm. Thus, although one is interested in a norm in the end, a seminorm naturally arises along the way. In a completely different manner, seminorms will be important in Chapter 6 in their own right. As we shall see, particularly in the context of so-called "generalised signals" in Chapter IV-3, seminorms often arise in natural way independently of whether they are used to define a norm.

In any event, here are the definitions.

3.1.2 Definition (Seminorm, norm) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be an $\mathbb{F}$-vector space. A ***seminorm*** on V is a map $\mathsf{V} \ni v \mapsto \|v\| \in \mathbb{R}_{\geq 0}$ with the following properties:

(i) $\|av\| = |a|\|v\|$ for $a \in \mathbb{F}$ and $v \in \mathsf{V}$ (***homogeneity***);

(ii) $\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$ for $v_1, v_2 \in \mathsf{V}$ (***triangle inequality***).

A ***norm*** on V is a seminorm $v \mapsto \|v\|$ with the additional property that

(iii) $\|v\| = 0$ only if $v = 0_{\mathsf{V}}$ (***positive-definiteness***).

We shall often denote a seminorm by $\|\cdot\|$. •

Let us give some examples of norms and seminorms. Sometimes examples are illustrative and sometimes they are of great value in their own right. The examples below, with the exception of the first one, are all of great independent interest, as well as illustrating the concept of a norm.

3.1.3 Examples (Seminorm, norm)

1. For any $\mathbb{F}$-vector space V there is a useless seminorm defined by $v \mapsto 0$. Let us call this the ***trivial seminorm*** since it is good for giving trivial examples. Unless $\mathsf{V} = \{0_{\mathsf{V}}\}$, the trivial seminorm is never a norm.
2. On $\mathbb{F}^n$ define
$$\|\boldsymbol{v}\|_2 = \left(|v_1|^2 + \cdots + |v_n|^2\right)^{1/2}.$$
In the case when $\mathbb{F} = \mathbb{R}$ this is the standard norm on $\mathbb{R}^n$ as discussed in Section II-1.2. In particular, this norm defines the usual notion of length of a vector in $\mathbb{F}^n$, i.e., $\|\boldsymbol{v}\|$ is the distance from $\mathbf{0}_{\mathbb{F}^n}$ to $\boldsymbol{v}$. Note that we now use different notation for this norm. We shall also sometimes call it the **2-*norm*** on $\mathbb{F}^n$ rather than the standard norm. It is pretty evident that $\|\cdot\|_2$ satisfies the homogeneity and positive-definiteness properties required of a norm. It is also true that $\|\cdot\|_2$ satisfies the triangle inequality. We do not prove this here, although it was proved in the case when $\mathbb{F} = \mathbb{R}$ as part of Proposition II-1.1.8. The proof of this relies on the so-called "Cauchy–Bunyakovsky–Schwarz Inequality." This inequality holds because $\|\cdot\|_2$ is the norm derived from an inner product on $\mathbb{F}^n$. Thus we shall see how $\|\cdot\|_2$ satisfies the triangle inequality when we discuss inner products in Section 4.1. Moreover, we shall see this example come up in another general context in Section 3.8.1. The point is that we will subsequently see multiple proofs of the triangle inequality for $\|\cdot\|_2$.
3. Let us consider another norm on $\mathbb{F}^n$ which differs from the standard norm. For $\boldsymbol{v} = (v_1, \dots, v_n) \in \mathbb{F}^n$ define
$$\|\boldsymbol{v}\|_1 = |v_1| + \cdots + |v_n|.$$
All properties of the norm are readily verified, including the triangle inequality, as this now follows from the triangle inequality for $|\cdot|$. Although different from the standard norm, this norm is in some sense equivalent to it, and we refer to Exercise 3.1.6 for an exploration of this. This norm is called the **1-*norm***.
4. Let us consider a final (for now) norm on $\mathbb{F}^n$ given by
$$\|\boldsymbol{v}\|_\infty = \max\{|v_j| \mid j \in \{1, \dots, n\}\}.$$
This is in fact a norm, called the **∞-*norm***. The only not entirely trivial norm property to verify is the triangle inequality. For this, let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{F}^n$ and let $j, k, \ell \in \{1, \dots, n\}$ have the property that $\|\boldsymbol{u}\|_\infty = |u_j|$, $\|\boldsymbol{v}\|_\infty = |v_k|$, and $\|\boldsymbol{u} + \boldsymbol{v}\|_\infty = |u_\ell + v_\ell|$. We then have
$$\|\boldsymbol{u} + \boldsymbol{v}\|_\infty = |u_\ell + v_\ell| \le |u_\ell| + |v_\ell| \le |u_j| + |v_k| = \|\boldsymbol{u}\|_\infty + \|\boldsymbol{v}\|_\infty.$$

Note that this norm is also different from the standard norm, but it is equivalent in some sense; Exercise 3.1.6.

The above three examples of norms were all defined on the finite-dimensional $\mathbb{F}$-vector space $\mathbb{F}^n$. Let us now consider infinite-dimensional analogues of these norms.

5. Recall from Example I-4.5.2–4 that $\mathbb{F}_0^\infty$ denotes the sequences $(v_j)_{j\in\mathbb{Z}_{>0}}$ for which the set $\{j \in \mathbb{Z}_{>0} \mid v_j \neq 0\}$ is finite. Thus sequences in $\mathbb{F}_0^\infty$ are eventually zero. We define

$$\|(v_j)_{j\in\mathbb{Z}_{>0}}\|_2 = \Big(\sum_{j=1}^\infty |v_j|^2\Big)^{1/2},$$

noting that the sum makes sense since it is actually finite. That $\|\cdot\|_2$ satisfies the properties of a norm is straightforward. Let us verify just the triangle inequality, since its proof gives the idea of how the norm works. We let $(u_j)_{j\in\mathbb{Z}_{>0}}, (v_j)_{j\in\mathbb{Z}_{>0}} \in \mathbb{F}_0^\infty$ and let $N \in \mathbb{Z}_{>0}$ be such that $u_j = v_j = 0$ for $j \geq N$. Then

$$\begin{aligned}\|(u_j)_{j\in\mathbb{Z}_{>0}} + (v_j)_{j\in\mathbb{Z}_{>0}}\|_2 &= \Big(\sum_{j=1}^\infty |u_j|^2 + \sum_{j=1}^\infty |v_j|^2\Big)^{1/2} = \Big(\sum_{j=1}^\infty (|u_j|^2 + |v_j|^2)\Big)^{1/2}\\ &= \Big(\sum_{j=1}^N (|u_j|^2 + |v_j|^2)\Big)^{1/2} \leq \Big(\sum_{j=1}^N |u_j|^2\Big)^{1/2} + \Big(\sum_{j=1}^N |u_j|^2\Big)^{1/2}\\ &= \Big(\sum_{j=1}^\infty |u_j|^2\Big)^{1/2} + \Big(\sum_{j=1}^\infty |u_j|^2\Big)^{1/2}\\ &= \|(u_j)_{j\in\mathbb{Z}_{>0}}\|_2 + \|(v_j)_{j\in\mathbb{Z}_{>0}}\|_2,\end{aligned}$$

where we have used the triangle inequality for the 2-norm on $\mathbb{F}^N$. This norm is called the ***2-norm*** on $\mathbb{F}_0^\infty$.

6. We again consider the vector space $\mathbb{F}_0^\infty$ and now define

$$\|(v_j)_{j\in\mathbb{Z}_{>0}}\|_1 = \sum_{j=1}^\infty |v_j|,$$

this sum again making sense since it is finite. It is easy to verify, just as we did for the 2-norm above, that $\|\cdot\|_1$ is a norm, and we call it the ***1-norm***.

7. As a final norm on $\mathbb{F}_0^\infty$ we define

$$\|(v_j)_{j\in\mathbb{Z}_{>0}}\|_\infty = \sup\{|v_j| \mid j \in \mathbb{Z}_{>0}\}.$$

Because the sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ is finite, it is certainly bounded, and so the definition makes sense. Moreover, the norm properties follow, essentially from those of $\|\cdot\|_\infty$ on $\mathbb{F}^n$. This norm we call, of course, the ***∞-norm***.

Now we consider yet another generalisation of the three types of norms we have been considering, now thinking about, not sequences, but functions. The reader should note the very strong analogies between the definitions of the norms that follow and the norms above: the sums are replaced with integrals and the "max" is replaced with a "sup." Since the issues surrounding norms on infinite-dimensional vector spaces can be complex, one should cling to familiarity where possible.

8. We consider the $\mathbb{F}$-vector space $\mathsf{C}^0([a,b];\mathbb{F})$ of continuous $\mathbb{F}$-valued functions on the compact interval $[a,b]$. Provided that $b > a$ this is an infinite-dimensional vector space, cf. Example I-4.5.18–6. On this vector space we define

$$\|f\|_2 = \Big(\int_a^b |f(x)|^2\,\mathrm{d}x\Big)^{1/2}.$$

Note that continuous functions (and therefore their squares) on compact intervals are always Riemann integrable by Corollary I-3.4.12, and so the integral here is the friendly Riemann integral. It is easy to see that this possible norm satisfies the homogeneity and positive-definiteness properties of a norm (see Exercise I-3.4.1 for positive-definiteness). Thus, like its 2-norm brother on $\mathbb{F}^n$, the difficult norm property to verify is the triangle inequality. However, we shall see in that this norm is derived from an inner product, and so this will give the triangle inequality just like the 2-norm on $\mathbb{F}^n$. We shall also see this norm arise from the more general setting of Section 3.8.8. Again, the point is that we will subsequently prove the triangle inequality for $\|\cdot\|_2$ in a few different ways. where?

This norm will be called the ***2-norm*** on $\mathsf{C}^0([a,b];\mathbb{F})$.

9. On $\mathsf{C}^0([a,b];\mathbb{F})$ define

$$\|f\|_1 = \int_a^b |f(x)|\,\mathrm{d}x.$$

Again, the integral here is the Riemann integral. The three norm properties are easily verified. Only the triangle inequality is possibly nontrivial:

$$\begin{aligned}\|f+g\|_1 &= \int_a^b |f(x)+g(x)|\,\mathrm{d}x \le \int_a^b \big(|f(x)|+|g(x)|\big)\,\mathrm{d}x\\ &= \int_a^b |f(x)|\,\mathrm{d}x + \int_a^b |g(x)|\,\mathrm{d}x = \|f\|_1 + \|g\|_1.\end{aligned}$$

This norm, called the ***1-norm***, is different than the 2-norm. As the reader can explore in Exercise 3.1.6, for the 1- and 2-norms on $\mathbb{F}^n$, there is some sort of equivalence between these. However, for the 1- and 2-norms on $\mathsf{C}^0([a,b];\mathbb{F})$ this is no longer true. This is not perfectly obvious right now, and the reader will have to wait until to start understanding this. But this is where we start to see how things are more complicated for infinite-dimensional vector spaces. what?

10. As a final norm on $\mathsf{C}^0([a,b],\mathbb{F})$ we take

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in [a,b]\}.$$

Again, the triangle inequality is the troublesome property to verify. In this case the verification goes as follows:

$$\begin{aligned} \|f+g\|_\infty &= \sup\{|f(x)+g(x)| \mid x \in [0,1]\} \\ &\leq \sup\{|f(x)|+|g(x)| \mid x \in [0,1]\} \\ &\leq \sup\{|f(x)|+|g(y)| \mid x,y \in [0,1]\} \\ &\leq \sup\{|f(x)| \mid x \in [0,1]\} + \sup\{|g(y)| \mid y \in [0,1]\} \\ &= \|f\|_\infty + \|g\|_\infty. \end{aligned}$$

This norm is yet again different than the 1- and 2-norms. Moreover, it is yet again fundamentally not equivalent, distinguishing the infinite-dimensional case from the finite-dimensional case. This will be elucidated in . •

what?

An obvious question is whether a vector space always possesses a norm. The answer is, "Yes, it does," and the astute reader will have seen from Examples 5, 6, and 7 above how this can be done. We record this as the following result.

3.1.4 Proposition (Vector spaces always have at least one norm) *If* $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ *and if* V *is an* $\mathbb{F}$*-vector space then there is a norm on* V*.*

Proof By Theorem I-4.5.45 we know the vector space V possesses a basis which establishes an isomorphism ι of V with $\mathbb{F}_0^J$ for some set J. Let us first define a norm on $\mathbb{F}_0^J$. Writing a typical element of $\mathbb{F}_0^J$ as $(v_j)_{j\in J}$ we define

$$\|(v_j)_{j\in J}\|_J = \sum_{j\in J} |v_j|,$$

noting that this sum exists since all but finitely many of the v_j's are zero. To verify that this is a norm is straightforward, cf. Example 3.1.3–5. Now define $\|\cdot\|_{\mathsf{V}}$ by $\|v\|_{\mathsf{V}} = \|\iota(v)\|_J$. That this is indeed defines a norm follows from linearity of ι:

$$\begin{aligned} &\|av\|_{\mathsf{V}} = \|\iota(av)\|_J = \|a\iota(v)\|_J = |a|\|\iota(v)\| = |a|\|v\|_{\mathsf{V}}; \\ &\|v_1+v_2\|_{\mathsf{V}} = \|\iota(v_1)+\iota(v_2)\|_J \leq \|\iota(v_1)\|_J + \|\iota(v_2)\|_J = \|v_1\|_{\mathsf{V}} + \|v_2\|_{\mathsf{V}}. \end{aligned}$$

Also, if $\|v\|_{\mathsf{V}} = 0$ this $\|\iota(v)\|_J = 0$ which means that $\iota(v) = 0_{\mathbb{F}_0^J}$. Thus $v = 0_{\mathsf{V}}$ since ι is an isomorphism. ■

One needs to take care with the preceding result: (1) it does not say that there is a *unique* norm on a given vector space; (2) it does not say that there is a useful norm on a given vector space. Indeed, we will see in Corollary 3.6.27 that some vector spaces do not possess "useful" norms. Thus the result should be thought of as being in the interesting vein rather than the useful vein, particularly for infinite-dimensional normed vector spaces.

In terms of convenient lingo the following definition is helpful.

3.1.5 Definition (Seminormed vector space, normed vector space) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

(i) A ***seminormed*** $\mathbb{F}$***-vector space*** is a pair $(\mathsf{V}, \|\cdot\|)$ where V is a $\mathbb{F}$-vector space and $\|\cdot\|$ is a seminorm on V.

(ii) A ***normed*** $\mathbb{F}$***-vector space*** is a pair $(\mathsf{V}, \|\cdot\|)$ where V is a $\mathbb{F}$-vector space and $\|\cdot\|$ is a norm on V. •

3.1.6 Notation ((Semi)normed vector spaces) If a norm or seminorm is understood, we shall often say, "the (semi)normed $\mathbb{F}$-vector space V." One really needs to exercise caution with this abuse, however, since the same vector space can have multiple norms, and the behaviour can depend in a drastic way on the norm. •

Let us give some more or less trivial properties of normed vector spaces.

3.1.7 Proposition (Properties of seminormed and normed vector spaces) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $(\mathsf{V}, \|\cdot\|)$ *be a seminormed* $\mathbb{F}$*-vector space, and let* $\mathsf{U} \subseteq \mathsf{V}$ *be a subspace. Then the following statements hold:*

(i) the map $(v_1, v_2) \mapsto \|v_1 - v_2\|$ *is a semimetric on* V*, and is a metric when* $\|\cdot\|$ *is a norm;*

(ii) $\big|\|v_1\| - \|v_2\|\big| \le \|v_1 - v_2\|$ *for all* $v_1, v_2 \in \mathsf{V}$*;*

(iii) $\big|\|v_1 - v_3\| - \|v_2 - v_4\|\big| \le \|v_1 - v_2\| + \|v_3 - v_4\|$ *for all* $v_1, v_2, v_3, v_4 \in \mathsf{V}$*;*

(iv) the restriction of $\|\cdot\|$ *to* U *defines a seminorm on* U*, and this seminorm is a norm when* $\|\cdot\|$ *is a norm.*

Proof (i) This is just a matter of plugging in the definitions. Perhaps the only nontrivial fact is the triangle inequality:

$$\|v_1 - v_3\| = \|(v_1 - v_2) + (v_2 - v_3)\| \le \|v_1 - v_2\| + \|v_2 - v_3\|.$$

(ii) This is Exercise 3.1.3.

(iii) We use the triangle inequality and part (ii):

$$\begin{aligned}\big|\|v_1 - v_3\| - \|v_2 - v_4\|\big| &= \big|\|v_1 - v_3\| - \|v_2 - v_3\| + \|v_3 - v_2\| - \|v_2 - v_4\|\big| \\ &\le \big|\|v_1 - v_3\| - \|v_2 - v_3\|\big| + \big|\|v_3 - v_2\| + \|v_2 - v_4\|\big| \\ &\le \|v_1 - v_2\| + \|v_3 - v_4\|,\end{aligned}$$

as desired.

(iv) This is trivial. ■

Now we indicate how one can pass from a seminormed vector space to a normed vector space in a natural way. This mirrors our result Theorem 1.1.28 for semimetric spaces.

3.1.8 Theorem (Normed vector spaces from seminormed vector spaces) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \|\cdot\|)$ *be a seminormed* $\mathbb{F}$*-vector space. Then the following statements hold:*

(i) *the set* $\mathsf{V}_0 = \{v \in \mathsf{V} \mid \|v\| = 0\}$ *is a subspace of* V*;*

(ii) *the function* $\mathsf{V}/\mathsf{V}_0 \ni v + \mathsf{V}_0 \mapsto \|v\|$ *is a norm on* V/V_0*.*

Proof (i) If $u, v \in \mathsf{V}_0$ and if $a \in \mathbb{F}$ then

$$0 \le \|u + v\| \le \|u\| + \|v\| = 0$$

and

$$\|av\| = |a|\|v\| = 0,$$

giving $u + v, av \in \mathsf{V}_0$, as desired.

(ii) First let us show that the function is well-defined. Suppose that $v + \mathsf{V}_0 = v' + \mathsf{V}_0$ so that $v - v' \in \mathsf{V}_0$. Then

$$\|v'\| = \|v + (v' - v)\| \le \|v\| + \|v' - v\| = \|v\|$$

and

$$\|v\| = \|v' + (v - v')\| \le \|v'\| + \|v - v'\| = \|v'\|$$

using the triangle inequality. Thus $\|v'\| = \|v\|$, and the map is then well-defined. It clearly has the homogeneity and triangle inequality properties of a norm. To check the positive-definiteness, suppose that $\|v + \mathsf{V}_0\| = 0$. Then $\|v\| = 0$ and so $v \in \mathsf{V}_0$, giving $v + \mathsf{V}_0 = 0_{\mathsf{V}} + \mathsf{V}_0$, as desired. ■

3.1.2 Open and closed subsets of normed vector spaces

As we saw in Proposition 3.1.7 a seminorm (resp. norm) $\|\cdot\|$ on V determines a semimetric (resp. metric) on V by $d_{\|\cdot\|}(v_1, v_2) = \|v_1 - v_2\|$. A semimetric then determines a topology, and, if the semimetric is a metric, this topology is Hausdorff (see). Therefore, seminormed vector spaces are topological spaces, and normed vector spaces are Hausdorff topological spaces. In this section we describe this topology in more detail. Some of what we say is redundant since it follows from what we have already said for metric spaces. However, we aim to make our treatment of normed vector spaces as self-contained as possible. In this section we make statements that are valid for seminormed vector spaces, and not just normed vector spaces, although it is the latter that are of most immediate interest. We adopt the convention of writing "(semi)norm" when we mean that the object can be either a norm or a seminorm. Readers caring only about norms can omit the "(semi)" in their heads.

what

As with metrics, the building block of the topology of a normed vector space is the open ball.

3.1.9 Definition (Open, closed, and bounded sets in (semi)normed vector spaces) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space.

(i) The ***open ball*** of radius r about $v_0 \in \mathsf{V}$ is the set

$$\mathsf{B}(r, v_0) = \{v \in \mathsf{V} \mid \|v - v_0\| < r\}.$$

(ii) The ***closed ball*** of radius r about $v_0 \in \mathsf{V}$ is the set

$$\overline{\mathsf{B}}(r, v_0) = \{v \in \mathsf{V} \mid \|v - v_0\| \leq r\}.$$

(iii) A subset $U \subseteq \mathsf{V}$ is ***open*** if, for each $v \in U$, there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}(\epsilon, v) \subseteq U$. (The empty set is also open, by declaration.)

(iv) A subset $A \subseteq \mathsf{V}$ is ***closed*** is $\mathsf{V} \setminus A$ if open.

(v) A subset $A \subseteq \mathsf{V}$ is ***bounded*** if there exists $R \in \mathbb{R}_{>0}$ such that $A \subseteq \mathsf{B}(R, 0_{\mathsf{V}})$. •

One can easily show that the open ball is open (this is Exercise 3.1.1).

We shall not attempt to systematically distinguish notationally the rôle of $\|\cdot\|$ in the open ball $\mathsf{B}(r, v_0)$. If there is a potential cause of confusion we will handle it as it comes up. For example, if we are working with multiple (semi)normed vector spaces, we may use the notation $\mathsf{B}_{\mathsf{V}}(r, v_0)$ to specify that a ball is in V.

Let us give some properties of open sets.

3.1.10 Proposition (Properties of open subsets of (semi)normed vector spaces) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Then the following statements hold:*

(i) for $(\mathrm{U}_a)_{a \in A}$ an arbitrary family of open sets, $\cup_{a \in A} \mathrm{U}_a$ is open;

(ii) for $(\mathrm{U}_1, \ldots, \mathrm{U}_n)$ a finite family of open sets, $\cap_{j=1}^n \mathrm{U}_j$ is open.

Proof (i) Let $v \in \cup_{a \in A} U_a$. Then, since $v \in U_{a_0}$ for some $a_0 \in A$, there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}(\epsilon, v) \subseteq U_{a_0} \subseteq \cup_{a \in A} U_a$.

(ii) Let $v \in \cap_{j=1}^n U_j$. For each $j \in \{1, \ldots, n\}$, choose $\epsilon_j \in \mathbb{R}_{>0}$ such that $\mathsf{B}(\epsilon_j, v) \subseteq U_j$, and let $\epsilon = \min\{\epsilon_1, \ldots, \epsilon_n\}$. Then $\mathsf{B}(\epsilon, v) \subseteq U_j$, $j \in \{1, \ldots, n\}$, and so $\mathsf{B}(\epsilon, v) \subseteq \cap_{j=1}^n U_j$. ■

This result shows that the collection of open subsets of a (semi)normed vector space define a topology.

3.1.11 Definition ((Semi)norm topology) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. The topology on V whose open sets are the open sets defined by the (semi)norm $\|\cdot\|$ is the ***(semi)norm topology*** on V. •

One of the most important properties about the norm topology is that it is translation invariant. Let us see what this means. For $v_0 \in \mathsf{V}$ define $\tau_{v_0} \colon \mathsf{V} \to \mathsf{V}$ by $\tau_{v_0}(v) = v + v_0$. Thus τ_{v_0} is "translation by v_0." We then have the following result.

3.1.12 Proposition (Translation invariance of the (semi)norm topology) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Then a subset $\mathrm{U} \subseteq \mathsf{V}$ is open if and only if $\tau_{\mathrm{v}_0}(\mathrm{U})$ is open.*

Proof Suppose that $U \subseteq \mathsf{V}$ is open and let $v \in \tau_{v_0}(U)$. Then $\tau_{-v_0}(v) \in U$ and so there exists $\epsilon > 0$ such that $\mathsf{B}(\epsilon, v) \subseteq U$. Note that

$$\begin{aligned}
\tau_{v_0}(\mathsf{B}(\epsilon, v)) &= \tau_{v_0}(\{u \in \mathsf{V} \mid \|u - v\| < \epsilon\}) \\
&= \{v_0 + u \in \mathsf{V} \mid \|u - v\| < \epsilon\} \\
&= \{u' \in \mathsf{V} \mid \|u' - (v + v_0)\| < \epsilon\} \\
&= \mathsf{B}(\epsilon, \tau_{v_0}(v)).
\end{aligned}$$

Thus

$$\mathsf{B}(\epsilon, v) \subseteq U \implies \tau_{v_0}(\mathsf{B}(\epsilon, v)) \subseteq \tau_{v_0}(U) \implies \mathsf{B}(\epsilon, \tau_{v_0}(v)) \subseteq \tau_{v_0}(U).$$

Thus $\tau_{v_0}(U)$ is open.

Conversely, if $\tau_{v_0}(U)$ is open then, by the first part of the proof, $\tau_{-v_0}(\tau_{v_0}(U)) = U$ is open. ■

As the proof of the preceding result makes clear, the key to the translation invariance of the norm topology is the fact that $\tau_{v_0}(\mathsf{B}(r, v)) = \mathsf{B}(r, \tau_{v_0}(v))$ for every $r \in \mathbb{R}_{>0}$ and $v, v_0 \in \mathsf{V}$. This is a pretty obvious fact, but is so useful that it is worth pointing out explicitly.

The norm topology generally depends on the norm. However, it is possible that two different norms will give the same topology. The following definition captures this idea.

3.1.13 Definition (Equivalent norms) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be a $\mathbb{F}$-vector space. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ (the subscripts "1" and "2" have nothing to do with the 1- and 2-norms considered in Example 3.1.3) are ***equivalent*** if a subset $U \subseteq \mathsf{V}$ is open in the norm topology defined by $\|\cdot\|_1$ if and only if it is open in the norm topology defined by $\|\cdot\|_2$. •

We will not be interested in the notion of equivalence for seminorms.

In short, equivalent norms define the same open sets. It is useful to be able to characterise equivalent norms in a more computational manner, one that might be able to check in practice. The following result gives just such a characterisation.

3.1.14 Theorem (Characterisation of equivalent norms) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be a $\mathbb{F}$-vector space. Two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on V are equivalent if and only if there exists $\mathrm{C} \in \mathbb{R}_{>0}$ such that*

$$\mathrm{C}^{-1}\|\mathrm{v}\|_2 \le \|\mathrm{v}\|_1 \le \mathrm{C}\|\mathrm{v}\|_2$$

for all $\mathrm{v} \in \mathsf{V}$.

Proof First suppose that $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent. Let $\mathsf{B}_1(r, v_0)$ and $\mathsf{B}_2(r, v_0)$ denote the open balls of radius r centred at v_0 for $\|\cdot\|_1$ and $\|\cdot\|_2$, respectively. By , equivalence of the two norm topologies implies that for every $R \in \mathbb{R}_{>0}$ there exists $C_1, C_2 \in \mathbb{R}_{>0}$ such that

what?

$$\mathsf{B}_2(C_1, 0_{\mathsf{V}}) \subseteq \mathsf{B}_1(R, 0_{\mathsf{V}}) \subseteq \mathsf{B}_2(C_2, 0_{\mathsf{V}}).$$

Let us consider the inclusion $\mathsf{B}_2(C_1, 0_{\mathsf{V}}) \subseteq \mathsf{B}_1(R, 0_{\mathsf{V}})$. If $v \in \mathsf{V}$ is nonzero then this inclusion gives

$$\|v\|_2 \le 1 \implies \|C_1 v\|_2 \le C_1 \implies \|C_1 v\|_1 \le R \implies \frac{\|C_1 v\|_1}{\|C_1 v\|_2} \le \frac{R}{\|C_1 v\|_2}$$
$$\implies \frac{\|v\|_1}{\|v\|_2} \le \frac{R}{C_1} \implies \|v\|_1 \le \frac{R}{C_1}\|v\|_2.$$

Thus $\|v\|_1 \le \frac{R}{C_1}\|v\|_2$ holds if v is nonzero and if $\|v\|_2 \le 1$. Clearly the same equality holds for $v = 0_{\mathsf{V}}$. For $v \in \mathsf{V}$ nonzero we also have

$$\Big\|\frac{v}{\|v\|_2}\Big\|_1 \le \frac{R}{C_1}\Big\|\frac{v}{\|v\|_2}\Big\|_2 \implies \|v\|_1 \le \frac{R}{C_1}\|v\|_2.$$

Thus the relation $\|v\|_1 \le \frac{R}{C_1}\|v\|_2$ holds for all $v \in \mathsf{V}$.

An entirely similar argument shows that the inclusion $\mathsf{B}_1(R, 0_{\mathsf{V}}) \subseteq \mathsf{B}_2(C_2, 0_{\mathsf{V}})$ implies that $\|v\|_2 \le \frac{C_2}{R}\|v\|_1$ for all $v \in \mathsf{V}$. Thus we have

$$\frac{C_2}{R}\|v\|_2 \le \|v\|_1 \le \frac{R}{C_1}\|v\|_2$$

for all $v \in \mathsf{V}$. Taking $C = \max\{\frac{R}{C_1}, \frac{R}{C_2}\}$ gives

$$C^{-1}\|v\|_2 \le \|v\|_1 \le C\|v\|_2, \qquad v \in \mathsf{V},$$

as desired.

Now suppose that there exists $C \in \mathbb{R}_{>0}$ such that

$$C^{-1}\|v\|_2 \le \|v\|_1 \le C\|v\|_2$$

for all $v \in \mathsf{V}$. Let $R \in \mathbb{R}_{>0}$ and note that

$$v \in \mathsf{B}_1(R, 0_{\mathsf{V}}) \implies \|v\|_1 < R \implies \|v\|_2 \le RC \implies v\mathsf{B}_2(RC, 0_{\mathsf{V}}).$$

Thus $\mathsf{B}_1(R, 0_{\mathsf{V}}) \subseteq \mathsf{B}_2(RC, 0_{\mathsf{V}})$. Similarly we show that $\mathsf{B}_2(\frac{R}{C}, 0_{\mathsf{V}}) \subseteq \mathsf{B}_1(R, 0_{\mathsf{V}})$. Thus we have

$$\mathsf{B}_2(\tfrac{R}{C}, 0_{\mathsf{V}}) \subseteq \mathsf{B}_1(R, 0_{\mathsf{V}}) \subseteq \mathsf{B}_2(RC, 0_{\mathsf{V}})$$

for every $R \in \mathbb{R}_{>0}$. From the remarks following the proof of Proposition 3.1.12 it follows that

$$\mathsf{B}_2(\tfrac{R}{C}, v_0) \subseteq \mathsf{B}_1(R, v_0) \subseteq \mathsf{B}_2(RC, v_0)$$

for every $R \in \mathbb{R}_{>0}$ and every $v_0 \in \mathsf{V}$. The equivalence of the two norm topologies now follows from . ■ what?

The following result shows that, on a finite-dimensional normed vector space there is really only one norm topology, although one can use different norms to define it.

3.1.15 Theorem (Uniqueness of the norm topology on finite-dimensional normed vector spaces) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if V is a finite-dimensional $\mathbb{F}$-vector space, then any two norms on V are equivalent.*

Proof Let $\{e_1, \dots, e_n\}$ be a basis for V and let $\iota\colon \mathsf{V} \to \mathbb{F}^n$ be defined by

$$\iota(v_1e_1, \dots, v_ne_n) = (v_1, \dots, v_n).$$

Define norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on V by

$$\|v\|_1 = \|\iota(v)\|_1 = \sum_{j=1}^{n} |v_j|,$$

$$\|v\|_2 = \|\iota(v)\|_2 = \Big(\sum_{j=1}^{n} |v_j|^2\Big)^{1/2}.$$

Thus we are abusing notation and using $\|\cdot\|_1$ and $\|\cdot\|_2$ for norms both on V and $\mathbb{F}^n$. These do define norms on V by Example 3.1.3 (also, cf. the proof of Proposition 3.1.4). Since the notion of equivalence of norms is an equivalence relation (this is Exercise 3.1.5), it suffices to show that any other norm of V is equivalent to $\|\cdot\|_2$. Let $\|\cdot\|$ be another norm on V and write, for $u, v \in \mathsf{V}$,

$$u = u_1 e_1 + \dots + u_n e_n, \quad v = v_1 e_1 + \dots + v_n e_n.$$

We then have, by Exercise 3.1.3 and Proposition II-1.1.11, and the triangle inequality,

$$\big|\|u\| - \|v\|\big| \le \|u - v\| = \Big\|\sum_{j=1}^{n} (u_j - v_j) e_j\Big\| \le \sum_{j=1}^{n} |u_j - v_j| \|e_j\|$$

$$\le \max\{\|e_j\| \mid j \in \{1, \dots, n\}\} \|u - v\|_1 \le C\|v\|_2,$$

where $C = \max\{\|e_j\| \mid j \in \{1, \dots, n\}\}\sqrt{n}$. We claim that this implies that the function $\boldsymbol{v} \mapsto \|\iota^{-1}(\boldsymbol{v})\|$ on $\mathbb{F}^n$ is continuous with respect to the norm $\|\cdot\|_2$. Indeed, for $\epsilon \in \mathbb{R}_{>0}$ let $\delta = \frac{\epsilon}{C}$. For $\boldsymbol{v}_0 \in \mathbb{F}^n$ suppose that $\|\boldsymbol{v} - \boldsymbol{v}_0\|_2 < \delta$. Then, from our computations above,

$$\big|\|\iota^{-1}(\boldsymbol{v})\| - \|\iota^{-1}(\boldsymbol{v}_0)\|\big| \le C\|\boldsymbol{v} - \boldsymbol{v}_0\|_2 < \epsilon,$$

giving continuity of $\boldsymbol{v} \mapsto \|\iota^{-1}(\boldsymbol{v})\|$ at $\boldsymbol{v}_0$. Let $\overline{\mathsf{B}}_2(1, \boldsymbol{0}_{\mathbb{F}^n})$ be the unit ball with respect to the norm $\|\cdot\|_2$ centred at the origin in $\mathbb{F}^n$ and let $\overline{\mathsf{B}}_2(1, 0_{\mathsf{V}})$ be the unit ball with respect to the norm $\|\cdot\|_2$ centred at the origin in V. The boundary of $\overline{\mathsf{B}}_2(1, \boldsymbol{0}_{\mathbb{F}^n})$ is closed and bounded with respect to the norm $\|\cdot\|_2$ and its topology, and so is compact in $\mathbb{F}^n$ with
what respect to the usual topology by the Heine–Borel Theorem. Therefore, by , the function $\boldsymbol{v} \mapsto \|\iota^{-1}(\boldsymbol{v})\|$ attains a minimum value $m \in \mathbb{R}_{>0}$ and a maximum value $M \in \mathbb{R}_{>0}$ on $\mathrm{bd}(\overline{\mathsf{B}}_2(1, \boldsymbol{0}_{\mathbb{F}^n}))$. Thus, for $\boldsymbol{v} \in \overline{\mathsf{B}}_2(1, \boldsymbol{0}_{\mathbb{F}^n})$ we have

$$m \le \|\iota^{-1}(\boldsymbol{v})\| \le M$$

which is equivalent to saying that, for $v \in \mathrm{bd}(\overline{\mathsf{B}}_2(1, 0_{\mathsf{V}}))$ (boundary being taken with respect to the norm topology on V for the norm $\|\cdot\|_2$) we have

$$m \le \|v\| \le M.$$

For arbitrary $v \in \mathsf{V} \setminus \{0_{\mathsf{V}}\}$ this gives

$$m \le \Big\|\frac{v}{\|v\|_2}\Big\| \le M \quad \Longrightarrow \quad m\|v\|_2 \le \|v\| \le M\|v\|_2,$$

showing that $\|\cdot\|$ and $\|\cdot\|_2$ are equivalent if we take $C = \max\{M, m^{-1}\}$. ■

We will use this theorem to unambiguously talk about the norm topology on $\mathbb{F}^n$, or any finite-dimensional $\mathbb{F}$-vector space, as being the topology defined by *any* norm.

3.1.3 Subspaces, direct sums, and quotients

We have studied in Section I-4.5 the notions of subspace, direct sum, and quotient from an algebraic point of view. Let us see now how these notions interact with the structure of a norm.

For subspaces we record the following trivial result. We will have much more to say about subspaces of normed vector spaces in Section 3.6.4.

3.1.16 Proposition (Subspaces of (semi)normed vector spaces are (semi)normed vector spaces) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. If $\mathsf{U} \subseteq \mathsf{V}$ is a subspace then the map $\mathsf{U} \ni u \mapsto \|u\| \in \mathbb{R}_{\geq 0}$ is a (semi)norm on U.*

Now we consider direct sums of normed vector spaces. Let us first consider the general case, and then consider the case of finite direct sums as a special case. We recall from Definition I-4.5.39 that the direct sum of a family $(\mathsf{V}_i)_{i\in I}$ of vector spaces is the set of maps $\phi\colon I \to \cup_{i\in I}\mathsf{V}_i$ for which $\phi(i) \in \mathsf{V}_i$, $i \in I$, and for which the set $\{i \in I \mid \phi(i) \neq 0_{\mathsf{V}_i}\}$ is finite. This set has a natural vector space structure and is denoted $\bigoplus_{i\in I}\mathsf{V}_i$.

3.1.17 Theorem (Direct sums of (semi)normed vector spaces are (semi)normed vector spaces) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$ be a family of (semi)normed $\mathbb{F}$-vector spaces. For $\phi \in \bigoplus_{i\in I}\mathsf{V}_i$ define*

$$\|\phi\|_I = \sum_{i\in I}\|\phi(i)\|_i,$$

this sum being well-defined since it is finite. Then $(\bigoplus_{i\in I}\mathsf{V}_i, \|\cdot\|_I)$ is a (semi)normed $\mathbb{F}$-vector space, and is moreover a normed vector space if each of the components $(\mathsf{V}_i, \|\cdot\|_i)$, $i \in I$, is a normed vector space.

Proof Let $a \in \mathbb{F}$ and compute

$$\|a\phi\|_I = \sum_{i\in I}\|a\phi(i)\|_a = \sum_{i\in I}|a|\|\phi(i)\|_i = |a|\sum_{i\in I}\|\phi(i)\|_a = |a|\|\phi\|_I,$$

where all operations make sense since the sums are finite.

If $\phi, \psi \in \bigoplus_{i\in I}\mathsf{V}_i$ we compute

$$\|\phi + \psi\|_I = \sum_{i\in I}\|\phi(i) + \psi(i)\|_i \leq \sum_{i\in I}\|\phi(i)\|_i + \sum_{i\in I}\|\psi(i)\|_i = \|\phi\|_I + \|\psi\|_I,$$

as desired.

Finally, if

$$\|\phi\| = \sum_{i\in I}\|\phi(i)\|_i = 0$$

then we must have $\|\phi(i)\|_i = 0$ for each $i \in I$. If each of the seminorms $\|\cdot\|_i$, $i \in I$, are norms then this implies that $\phi(i) = 0_{\mathsf{V}_i}$, $i \in I$, implying that $\|\cdot\|_I$ is a norm. ■

We can now make the following definition.

3.1.18 Definition (Direct sum of (semi)normed vector spaces) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$ be a family of (semi)normed $\mathbb{F}$-vector spaces. The (semi)normed vector space $(\bigoplus_{i\in I} \mathsf{V}_i, \|\cdot\|_I)$ is the ***direct sum*** of $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$. •

Let us record how this works for the direct sum of two (semi)normed vector spaces. Thus let $(\mathsf{V}_1, \|\cdot\|_1)$ and $(\mathsf{V}_2, \|\cdot\|_2)$ be (semi)normed $\mathbb{F}$-vector spaces. Their direct sum is the vector space $\mathsf{V}_1 \oplus \mathsf{V}_2$, points in which we denote by (v_1, v_2), with the (semi)norm

$$\|(v_1, v_2)\|_{1,2} = \|v_1\|_1 + \|v_2\|_2.$$

Now we consider quotients of normed vector spaces by subspaces.

3.1.19 Proposition (The quotient of a (semi)normed vector space is a (semi)normed vector space) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space, and let U be a subspace. If we define*

$$\|\mathrm{v} + \mathsf{U}\|_{/\mathsf{U}} = \inf\{\|\mathrm{v} + \mathrm{u}\| \mid \mathrm{u} \in \mathsf{U}\}$$

then $\|\cdot\|_{/\mathsf{U}}$ is a seminorm on V/U. Moreover, if $\|\cdot\|$ is a norm and if U is closed, then $\|\cdot\|_{/\mathsf{U}}$ is a norm.

Proof It is evident that $\|v + \mathsf{U}\|_{/\mathsf{U}} \in \mathbb{R}_{>0}$. If $a = 0$ we have

$$\|0(v + \mathsf{U})\|_{/\mathsf{U}} = \|0v + \mathsf{U}\|_{/\mathsf{U}} = \inf\{\|0_{\mathsf{V}} + u\| \mid u \in \mathsf{U}\} = 0 = |a|\|v + \mathsf{U}\|_{/\mathsf{U}}.$$

For $a \in \mathbb{F} \setminus \{0\}$ we have

$$\begin{aligned}\|a(v + \mathsf{U})\|_{/\mathsf{U}} = \|av + \mathsf{U}\|_{/\mathsf{U}} &= \inf\{\|av + u\| \mid u \in \mathsf{U}\}\\ &= \inf\{\|av + au'\| \mid u' \in \mathsf{U}\} = \inf\{|a|\|v + u'\| \mid u' \in \mathsf{U}\}\\ &= |a| \inf\{\|v + u'\| \mid u' \in \mathsf{U}\} = |a|\|v + \mathsf{U}\|_{/\mathsf{U}}.\end{aligned}$$

For the triangle inequality we have

$$\begin{aligned}\|(v_1 + \mathsf{U}) + (v_2 + \mathsf{U})\|_{/\mathsf{U}} &= \|(v_1 + v_2) + \mathsf{U}\|_{/\mathsf{U}} = \inf\{\|v_1 + v_2 + u\| \mid u \in \mathsf{U}\}\\ &= \inf\{\|v_1 + v_2 + u_1 + u_2\| \mid u_1, u_2 \in \mathsf{U}\}\\ &\leq \inf\{\|v_1 + u_1\| + \|v_2 + u_2\| \mid u_1, u_2 \in \mathsf{U}\}\\ &= \inf\{\|v_1 + u_1\| \mid u_1 \in \mathsf{U}\} + \inf\{\|v_2 + u_2\| \mid u_2 \in \mathsf{U}\}\\ &= \|v_1 + \mathsf{U}\|_{/\mathsf{U}} + \|v_2 + \mathsf{U}\|_{/\mathsf{U}},\end{aligned}$$

as desired, where we have used Proposition I-2.2.27.

To prove the final assertion we rely on some facts about closed sets that we will not prove until Section 3.6.2. Let $v + \mathsf{U} \in \mathsf{V}/\mathsf{U}$ satisfy $\|v + \mathsf{U}\|_{/\mathsf{U}} = 0$. Thus

$$\inf\{\|v + u\| \mid u \in \mathsf{U}\} = 0.$$

Therefore, for $j \in \mathbb{Z}_{>0}$, there exists $u_j \in \mathsf{U}$ such that $\|v + u_j\| < \frac{1}{j}$. Thus the sequence $(v + u_j)_{j\in\mathbb{Z}_{>0}}$ converges to 0_{V}. By Proposition 3.2.6 it follows that the sequence $(u_j)_{j\in\mathbb{Z}_{>0}}$ converges to $-v$. Since the sequence is in U and since U is closed, by Proposition 3.6.8(ii) it follows that $-v \in \mathsf{U}$ and so $v \in \mathsf{U}$. Thus $v + \mathsf{U} = 0_{\mathsf{V}} + \mathsf{U}$, giving $\|\cdot\|_{/\mathsf{U}}$ as a norm. ■

One should be a little careful with the result. It does not say that $\|\cdot\|_{/\mathsf{U}}$ is a norm if $\|\cdot\|$ is a norm; this requires the additional assumption that U is closed.

Let us examine some properties of the canonical projection from V to V/U.

Let us examine some properties of the canonical projection from V to V/U. Here we refer ahead to Section 3.5 for notion of continuity and back to for the notion of the quotient topology. what

3.1.20 Proposition (The canonical projection onto the quotient is continuous) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space, and let U be a subspace. Then the canonical projection $\pi_{\mathsf{U}} \colon \mathsf{V} \to \mathsf{V}/\mathsf{U}$ is continuous. Moreover, the seminorm topology on V/U coincides with the quotient topology.*

Proof Let $v \in \mathsf{V}$ with $v + \mathsf{U}$ the projection to V/U. Let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in V converging to v. We claim that $(v_j + \mathsf{U})_{j\in\mathbb{Z}_{>0}}$ converges to $v + \mathsf{U}$. Indeed, if $\epsilon \in \mathbb{R}_{>0}$, take $N \in \mathbb{Z}_{>0}$ such that $\|v - v_j\| < \epsilon$ for $j \geq N$. Then

$$\|(v - v_j) + \mathsf{U}\|_{/\mathsf{U}} \leq \|v - v_j\| < \epsilon$$

for $j \geq N$, giving convergence as desired. Continuity of $v \mapsto v + \mathsf{U}$ now follows from Theorem 3.5.2.

Let $\pi \colon \mathsf{V} \to \mathsf{V}/\mathsf{U}$ denote the canonical projection. Now let $S \subseteq \mathsf{V}/\mathsf{U}$ be such that $\pi^{-1}(S)$ is a open. We claim that S is a open. For $v_0 + \mathsf{U} \in S$ let $\mathsf{B}_{\mathsf{V}}(\epsilon, v_0)$ be an open ball about v_0 contained in $\pi^{-1}(S)$. We have

$$\begin{aligned}\pi(\mathsf{B}_{\mathsf{V}}(\epsilon, v_0)) &= \{v + \mathsf{U} \mid \|v - v_0\| < \epsilon\} = \{v + \mathsf{U} \mid \|(v - v_0) + \mathsf{U}\|_{/\mathsf{U}} < \epsilon\} \\ &= \mathsf{B}_{\mathsf{V}/\mathsf{U}}(\epsilon, v_0 + \mathsf{U}).\end{aligned}$$

Since $\pi(\mathsf{B}_{\mathsf{V}}(\epsilon, v_0)) \subseteq S$ it follows that S is open. ■

Exercises

3.1.1 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Show that $\mathsf{B}(r, v_0)$ is open for every $r \in \mathbb{R}_{>0}$ and $v_0 \in \mathsf{V}$.

3.1.2 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a $\mathbb{F}$-vector space. Let $r_1, r_2 \in \mathbb{R}_{>0}$ satisfy $r_2 \leq r_1$ and let $v_1, v_2 \in \mathbb{R}^n$. Show that if $\overline{\mathsf{B}}(r_1, v_1) \cap \overline{\mathsf{B}}(r_2, v_2) \neq \emptyset$ then $\overline{\mathsf{B}}(r_2, v_2) \subseteq \overline{\mathsf{B}}(3r_1, v_1)$. Show that you understand your proof by drawing a picture.

3.1.3 In a (semi)normed vector space $(\mathsf{V}, \|\cdot\|)$ show that for each $v_1, v_2 \in \mathsf{V}$, $|\|v_1\| - \|v_2\|| \leq \|v_1 - v_2\|$.

3.1.4 Denote by $\mathsf{C}^1([0,1];\mathbb{R})$ the set of $\mathbb{R}$-valued functions on $[0,1]$ which are continuously differentiable, derivatives at 0 and 1 being taken from the right and left, respectively.

(a) For $f \in \mathsf{C}^1([0,1];\mathbb{R})$ define

$$\|f\| = \int_0^1 |f'(x)|\,\mathrm{d}x.$$

Show that $\|\cdot\|$ is a seminorm on $\mathsf{C}^1([0,1];\mathbb{R})$, but not a norm.

(b) For $f \in \mathsf{C}^1([0,1];\mathbb{R})$ define

$$\|f\| = |f(0)| + \int_0^1 |f'(x)|\,\mathrm{d}x.$$

Show that $\|\cdot\|$ is a norm on $\mathsf{C}^1([0,1];\mathbb{R})$.

3.1.5 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be an $\mathbb{F}$-vector space. Define a relation $\sim$ on the set of norms on V by saying that $\|\cdot\|_1 \sim \|\cdot\|_2$ if $\|\cdot\|_1$ and $\|\cdot\|_2$ are equivalent in the sense of Definition 3.1.13. Show that $\sim$ is an equivalence relation.

3.1.6 On $\mathsf{V} = \mathbb{R}^2$ consider the three norms $\|\cdot\|_2$, $\|\cdot\|_1$, and $\|\cdot\|_\infty$ given by Examples 3.1.3–2, 3.1.3–3, and 3.1.3–4, respectively.

(a) Draw the subsets $\mathsf{B}_2(r,\mathbf{0})$, $\mathsf{B}_1(r,\mathbf{0})$, and $\mathsf{B}_\infty(r,\mathbf{0})$ of $\mathbb{R}^2$ defined by

$$\overline{\mathsf{B}}_2(r,\mathbf{0}) = \{v \in \mathbb{R}^2 \mid \|v\|_2 \le r\}$$
$$\overline{\mathsf{B}}_1(r,\mathbf{0}) = \{v \in \mathbb{R}^2 \mid \|v\|_1 \le r\}$$
$$\overline{\mathsf{B}}_\infty(r,\mathbf{0}) = \{v \in \mathbb{R}^2 \mid \|v\|_\infty \le r\}.$$

(b) Using your drawings from part (a), argue that if and only if a sequence of points $(v_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}^2$ converges in one of the three norms, it converges in the other two norms.

3.1.7 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a finite-dimensional normed $\mathbb{F}$-vector space. Let $\{e_1, \dots, e_n\}$ be a basis for V for which $\|e_1\| = \cdots = \|e_n\| = 1$.

(a) For $v = v_1 e_1 + \cdots + v_n e_n \in \mathsf{V}$ define

$$\|v\|_1 = |v_1| + \cdots + |v_n|.$$

Show that $\|\cdot\|_1$ is a norm on V that satisfies $\|e_1\|_1 = \cdots = \|e_n\|_1 = 1$.

(b) Let $\mathsf{B}(1, 0_{\mathsf{V}})$ and $\mathsf{B}_1(1, 0_{\mathsf{V}})$ be the unit balls for the norms $\|\cdot\|$ and $\|\cdot\|_1$, respectively. Show that $\mathsf{B}_1(1, 0_{\mathsf{V}}) \subseteq \mathsf{B}(1, 0_{\mathsf{V}})$.

(The point is that the balls in the norm $\|\cdot\|_1$ are the smallest among the balls for all norms in which the basis vectors have unit length.)

Section 3.2

Sequences in normed vector spaces

Much of the structure of normed vector spaces can be captured by studying sequences in these spaces. Much of the presentation here follows the presentation of Section I-2.3. Indeed, many of the proofs are mere changes of notation of the analogous proofs for sequences in $\mathbb{R}$. However, we give all of the details of the presentation here for both (1) completeness and (2) because not all results are *exactly* the same as those for $\mathbb{R}$. This has the disadvantage of repetitiveness, but the advantage of making this section more self-contained.

Do I need to read this section? The ideas in this section are basic, so the definitions should be read and the results understood. Readers who are familiar with the material in Section I-2.3 will find this section reads pretty easily. •

3.2.1 Definitions and properties of sequences

Let V be a $\mathbb{F}$-vector space. A sequence in V is, in accordance with Definition I-1.6.8, a map from $\mathbb{Z}_{>0}$ to V, and we denote a sequence by $(v_j)_{j\in\mathbb{Z}_{>0}}$. For sequences we have the usual definitions corresponding to notions of convergence.

3.2.1 Definition (Convergence of sequences) Let$\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}$ and let $v_0 \in \mathsf{V}$. The sequence:

(i) is a ***Cauchy sequence*** if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $\|v_j - v_k\| < \epsilon$ for $j, k \geq N$;

(ii) ***converges to*** $\mathbf{v_0}$ if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $\|v_j - v_0\| < \epsilon$ for $j \geq N$;

(iii) ***diverges*** if it does not converge to any element in V;

(iv) is ***bounded*** if there exists $M \in \mathbb{R}_{>0}$ such that $\|v_j\| < M$ for each $j \in \mathbb{Z}_{>0}$.

If the sequence converges to v_0 then v_0 is the ***limit*** of the sequence and we write $v_0 = \lim_{j\to\infty} v_j$. •

3.2.2 Notation (Limits with general index sets) As in Section I-2.3.7 we can talk about limits of things more general than sequences. The setup where we will use this idea is the following. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces. We consider an open subset $O \subseteq \mathsf{U}$ and a map $\phi\colon O \to \mathsf{V}$. For $u_0 \in O$, we wish to define what we mean by $\lim_{u\to u_0} \phi(u)$. What we mean is this. If, there exists $v_0 \in \mathsf{V}$ such that, for any sequence $(u_j)_{j\in\mathbb{Z}_{>0}}$ converging to u_0, the sequence $(\phi(u_j))_{j\in\mathbb{Z}_{>0}}$ converges to $_0$, then we write $\lim_{u\to u_0} \phi(u_0) = v_0$. •

As for sequences in $\mathbb{Q}$, $\mathbb{R}$, or $\mathbb{C}$, convergent sequences are Cauchy.

3.2.3 Proposition (Convergent sequences are Cauchy) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. If $(\mathrm{v}_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence converging to v_0 then it is a Cauchy sequence.*

Proof Let $\epsilon \in \mathbb{R}_{>0}$ and choose $N \in \mathbb{Z}_{>0}$ such that $|v_j - v_0| < \frac{\epsilon}{2}$ for $j \geq N$. Then, for $j, k \geq N$ we have

$$\|v_j - v_k\| = \|v_j - v_0 - v_k + v_0\| = \|v_j - v_0\| + \|v_k - v_0\| < \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon,$$

using the triangle inequality. ■

Cauchy sequences are bounded.

3.2.4 Proposition (Cauchy sequences are bounded) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. If $(\mathrm{v}_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence, then it is bounded.*

Proof Choose $N \in \mathbb{Z}_{>0}$ such that $\|v_j - v_k\| < 1$ for $j, k \in \mathbb{Z}_{>0}$. Then take M_N to be the largest of the nonnegative real numbers $\|v_1\|, \dots, \|v_N\|$. Then, for $j \geq N$ we have, using the triangle inequality,

$$\|v_j\| = \|v_j - v_N + v_N\| \leq \|v_j - v_N\| + \|v_N\| < 1 + M_N,$$

giving the result by taking $M = M_N + 1$. ■

Since we often deal simultaneously with seminorms rather than just norms, it is useful to record what is different about the two cases. What we lose for seminorms is the uniqueness of limits for convergent sequences.

3.2.5 Proposition ((Non)uniqueness of limits for (semi)normed vector spaces) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a seminormed $\mathbb{F}$-vector space. If a sequence $(\mathrm{v}_j)_{j\in\mathbb{Z}_{>0}}$ converges to limits u_0 and v_0, then*

$$\mathrm{u}_0 - \mathrm{v}_0 \in \mathsf{V}_0 \triangleq \{\mathrm{v} \in \mathsf{V} \mid \ \|\mathrm{v}\| = 0\}.$$

In particular, if $\|\cdot\|$ is a norm then convergent sequences have unique limits.

Proof Suppose that the sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ converges to u_0 and v_0 and let $\epsilon \in \mathbb{R}_{>0}$. Choose $N \in \mathbb{Z}_{>0}$ such that

$$\|u_0 - v_j\| \leq \tfrac{\epsilon}{2}, \quad \|v_0 - v_j\| < \tfrac{\epsilon}{2}, \qquad j \geq N.$$

For $j \geq N$ we then have

$$\|u_0 - v_0\| = \|u_0 - v_j - (v_0 - v_j)\| \leq \|u_0 - v_j\| + \|v_0 - v_j\| \leq \epsilon.$$

Therefore, $\|u_0 - v_0\| = 0$, giving the result. ■

As is the case in our previous discussions of sequences in $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, one can wonder whether all Cauchy sequences converge. In cases where they do, we call the normed vector space complete (see Definition 3.3.2). In Section 3.3 we shall see that all finite-dimensional normed vector spaces are complete (Theorem 3.3.3) but that there are easy examples of infinite-dimensional normed vector spaces that are not complete (Example 3.3.1). This is one of the factors that tends to make the theory of infinite-dimensional normed vector spaces significantly more complicated than the finite-dimensional theory. For sequences in $\mathbb{R}$ and $\mathbb{C}$ there are useful tests for convergence. There are no significant analogues for sequences in normed vector spaces.

3.2.2 Algebraic operations on sequences

Convergence is compatible with the standard algebraic operations on vector spaces.

3.2.6 Proposition (Algebraic operations on sequences) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \|\cdot\|)$ *be a (semi)normed* $\mathbb{F}$*-vector space. Let* $(\mathrm{u}_j)_{j\in\mathbb{Z}}$ *and* $(\mathrm{v}_j)_{j\in\mathbb{Z}_{>0}}$ *be sequences in* V *converging to* u_0 *and* v_0*, respectively, let* $(\mathrm{a}_j)_{j\in\mathbb{Z}_{>0}}$ *be a sequence in* $\mathbb{F}$ *converging to* a_0*, and let* $\mathrm{a} \in \mathbb{F}$*. Then the following statements hold:*

(i) the sequence $(\mathrm{a}\mathrm{v}_j)_{j\in\mathbb{Z}_{>0}}$ *converges to* $\mathrm{a}\mathrm{v}_0$*;*

(ii) the sequence $(\mathrm{u}_j + \mathrm{v}_j)_{j\in\mathbb{Z}_{>0}}$ *converges to* $\mathrm{u}_0 + \mathrm{v}_0$*;*

(iii) the sequence $(\mathrm{a}_j\mathrm{v}_j)_{j\in\mathbb{Z}_{>0}}$ *converges to* $\mathrm{a}_0\mathrm{v}_0$*.*

Proof (i) The result is trivially true for $a = 0$, so let us suppose that $a \neq 0$. Let $\epsilon > 0$ and choose $N \in \mathbb{Z}_{>0}$ such that $\|v_j - v_0\| < \frac{\epsilon}{|a|}$. Then, for $j \geq N$,

$$\|av_j - av_0\| = |a|\|v_j - v_0\| < \epsilon.$$

(ii) Let $\epsilon > 0$ and take $N_1, N_2 \in \mathbb{Z}_{>0}$ such that

$$\|u_j - u_0\| < \tfrac{\epsilon}{2}, \quad j \geq N_1, \qquad \|v_j - v_0\| < \tfrac{\epsilon}{2}, \quad j \geq N_2.$$

Then, for $j \geq \max\{N_1, N_2\}$,

$$\|u_j + v_j - (u_0 + v_0)\| \leq \|u_j - u_0\| + \|v_j - v_0\| = \epsilon,$$

using the triangle inequality.

(iii) Let $\epsilon > 0$ and define $N_1, N_2, N_3 \in \mathbb{Z}_{>0}$ such that

$$\begin{aligned} &|a_j - a_0| < 1, \qquad j \geq N_1, \quad \implies \quad |a_j| < |a_0| + 1, \qquad j \geq N_1, \\ &|a_j - a_0| < \frac{\epsilon}{2(|a_0| + 1)}, \qquad j \geq N_2, \\ &\|v_j - v_0\| < \frac{\epsilon}{2(\|v_0\| + 1)}, \qquad j \geq N_2. \end{aligned}$$

Then, for $j \geq \max\{N_1, N_2, N_3\}$,

$$\begin{aligned} \|a_jv_j - a_0v_0\| &= \|a_jv_j - a_jv_0 + a_jv_0 - a_0v_0\| \\ &= \|a_j(v_j - v_0) + (a_j - a_0)v_0\| \\ &\leq |a_j|\|v_j - v_0\| + |a_j - a_0|\|v_0\| \\ &\leq (|a_0| + 1)\frac{\epsilon}{2(|a_0| + 1)} + \frac{\epsilon}{2(\|v_0\| + 1)}(\|v_0\| + 1) = \epsilon, \end{aligned}$$

as desired. ■

3.2.3 Multiple sequences

Finally, let us introduce the notion of a double sequence in a normed vector space.

3.2.7 Definition (Double sequence) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be an $\mathbb{F}$-vector space. A ***double sequence*** in V is a family of elements of V indexed by $\mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$. We denote a double sequence by $(v_{jk})_{j,k\in\mathbb{Z}_{>0}}$, where v_{jk} is the image of $(j,k) \in \mathbb{Z}_{>0} \times \mathbb{Z}_{>0}$ in V. •

For double sequences we have the following notions of convergence.

3.2.8 Definition (Convergence of double sequences) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \|\cdot\|$ be a (semi)normed $\mathbb{F}$-vector space, and let $v_0 \in \mathsf{V}$. A double sequence $(v_{jk})_{j,k\in\mathbb{Z}_{>0}}$:

(i) ***converges to*** $\mathbf{v_0}$, and we write $\lim_{j,k\to\infty} v_{jk} = v_0$, if, for each $\epsilon > 0$, there exists $N \in \mathbb{Z}_{>0}$ such that $\|v_0 - v_{jk}\| < \epsilon$ for $j, k \geq N$;

(ii) has v_0 as a ***limit*** if it converges to v_0.

(iii) is ***convergent*** if it converges to some member of V;

(iv) ***diverges*** if it does not converge. •

Some results here?

Exercises

3.2.1 In the $\mathbb{F}$-vector space $\mathbb{F}_0^\infty$, if possible find sequences with the following properties:

(a) Cauchy in the ∞-norm but not the 2-norm;

(b) Cauchy in the 2-norm but not the 1-norm;

(c) Cauchy in the 1-norm;

(d) Cauchy in the 1-norm but not the 2-norm;

(e) Cauchy in the 2-norm but not the ∞-norm.

3.2.2 Give an example of a sequence in $\mathsf{C}^0([0,1]; \mathbb{R})$ that is Cauchy with respect to the norm $\|\cdot\|_1$ but not with respect to the norm $\|\cdot\|_2$.
Hint: *Consider the function* $f\colon [0,1] \to \mathbb{R}$ *defined by*

$$f(x) = \begin{cases} x^{-1/2}, & x \in (0,1], \\ 0, & x = 0. \end{cases}$$

3.2.3 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. Let $(u_j)_{j\in\mathbb{Z}_{>0}}$ and $(v_j)_{j\in\mathbb{Z}_{>0}}$ be Cauchy sequences in V, let $(a_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathbb{F}$, and let $a \in \mathbb{F}$.

(a) Show that $(av_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence.

(b) Show that $(a_j v_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence.

(c) Show that $(u_j + v_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence.

Section 3.3

Completeness and completions

In Theorem I-2.3.5 we showed that the set of real numbers is complete in that every Cauchy sequence of real numbers converges. In Theorem II-1.2.5 we used the completeness of $\mathbb{R}$ to conclude that $\mathbb{R}^n$ is complete. As we shall see in Theorem 3.3.3, every finite-dimensional normed vector space is complete. This is not true for infinite-dimensional normed vector spaces, and so for these spaces the notion of completeness has teeth: in infinite-dimensional normed vector spaces there may well be Cauchy sequences that do not converge.

For reasons that are may not be perfectly clear initially, completeness is an essential property for a normed vector space to possess. If one is confronted with a normed vector space that is not complete, the first thing one does is complete it. We have already seen in Section 1.1.7 how this works for metric spaces, and the same ideas apply for normed vector spaces. Completions are easier to understand in general than they are in specific cases. This will become painfully clear in some of the examples in Section 3.8.

Do I need to read this section? Completeness is important, so the basic ideas in this section should be understood. The technicalities can be glossed over on a first reading. •

3.3.1 Completeness (Banach spaces)

Let us begin with two examples that illustrate that for normed vector spaces, the notions of Cauchy sequences and convergent sequences are not the same.

3.3.1 Examples (Nonconvergent Cauchy sequences)

1. First consider the normed vector space $(\mathbb{F}_0^\infty, \|\cdot\|_1)$ of Example 3.1.3–6. Consider the sequence $(s_k)_{k\in\mathbb{Z}_{>0}}$ in $\mathbb{F}_0^\infty$ defined by asking that s_k be the sequence $(v_{kj})_{j\in\mathbb{Z}_{>0}}$ with
$$v_{kj} = \begin{cases} \frac{1}{j^2}, & j \in \{1,\dots,k\}, \\ 0, & j > k. \end{cases}$$
Thus the sequence s_k is the truncation to k terms of the sequence $(\frac{1}{j^2})_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}$. We claim that this is a Cauchy sequence. Indeed, let $\epsilon > 0$ and choose $N \in \mathbb{Z}_{>0}$ sufficiently large that, for $k, l \geq N$ with $l > k$,
$$\sum_{j=k+1}^{l} \frac{1}{j^2} < \epsilon.$$
This is possible since the series $\sum_{j=1}^\infty \frac{1}{j^2}$ is convergent by Example I-2.4.2–4, and so its sequence of partial sums is Cauchy. Now let $k, l \geq N$ with $l > k$ and

compute

$$\|s_l - s_k\|_1 = \sum_{j=1}^{\infty} |v_{lj} - v_{kj}| = \sum_{j=k+1}^{l} \frac{1}{j^2} < \epsilon.$$

Thus the sequence $(s_k)_{k\in\mathbb{Z}_{>0}}$ is indeed Cauchy. However, it does not converge, as we now show. Suppose that $\sigma = (v_j)_{j\in\mathbb{Z}_{>0}}$ is an element of $\mathbb{F}_0^\infty$ such that $\lim_{k\to\infty}\|\sigma - s_k\|_1 = 0$, i.e., such that $(s_k)_{k\in\mathbb{Z}_{>0}}$ converges to σ. We claim that this implies that $v_j = \frac{1}{j^2}$ for each $j \in \mathbb{Z}_{>0}$. Indeed, suppose that $v_{j_0} \neq \frac{1}{j_0^2}$ for some $j_0 \in \mathbb{Z}_{>0}$. Then

$$\|\sigma - s_k\|_1 = \sum_{j=1}^{\infty} |v_j - s_k| \geq \left|v_{j_0} - \tfrac{1}{j_0^2}\right|$$

for every $k \in \mathbb{Z}_{>0}$. This implies that if $v_{j_0} \neq \frac{1}{j_0^2}$ for some $j_0 \in \mathbb{Z}_{>0}$ then $(s_k)_{k\in\mathbb{Z}_{>0}}$ cannot converge to σ. However, the sequence $(\frac{1}{j^2})_{j\in\mathbb{Z}_{>0}}$ is not in $\mathbb{F}_0^\infty$, as so we conclude that the sequence $(s_k)_{k\in\mathbb{Z}_{>0}}$ does not converge.

2. We work next with the normed vector space $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_1)$ of Example 3.1.3–9. In this vector space, consider the sequence of functions $(f_j)_{j\in\mathbb{Z}_{>0}}$ given by

$$f_j(x) = \begin{cases} 0, & x \in [0, \frac{1}{2} - \frac{1}{2j}], \\ 2jx + 1 - j, & x \in (\frac{1}{2} - \frac{1}{2j}, \frac{1}{2}), \\ 1, & x \in [\frac{1}{2}, 1]. \end{cases}$$

In Figure 3.1 a few terms in this sequence are graphed. Suppose that $k \geq j$ so

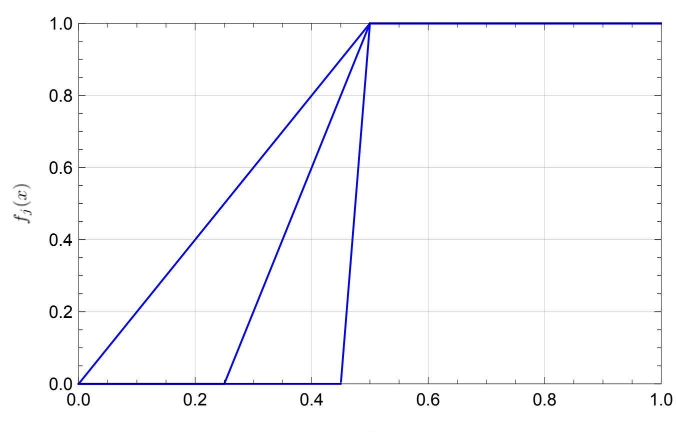


Figure 3.1 A Cauchy sequence (f_1, f_2, and f_{10} are shown) in $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_1)$

that the function $f_j - f_k$ is positive. A simple computation gives

$$\begin{aligned}\|f_j - f_k\|_1 &= \int_0^1 |f_j(x) - f_k(x)|\,\mathrm{d}x \\ &= \int_0^1 (f_j(x) - f_k(x))\,\mathrm{d}x \\ &= \int_0^1 f_j(x)\,\mathrm{d}x - \int_0^1 f_k(x)\,\mathrm{d}x \\ &= \frac{1}{2} + \frac{1}{4j} - \frac{1}{2} - \frac{1}{4k} = \frac{1}{4j} - \frac{1}{4k}.\end{aligned}$$

Now let $\epsilon > 0$ and take $N = \lceil \frac{1}{2\epsilon} \rceil$. This means that for any $j \geq N$ we have

$$j \geq N \geq \frac{1}{2\epsilon} \quad \Longrightarrow \quad \frac{1}{2j} \leq \epsilon.$$

We then have, for $j, k \geq N$,

$$\|f_j - f_k\|_1 = \left|\frac{1}{4j} - \frac{1}{4k}\right| < \frac{1}{4j} + \frac{1}{4k} \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

This shows that $(f_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence. However, it is evident that for any $x \in [0,1]$ we have

$$\lim_{j\to\infty} f_j(x) = f(x)$$

where $f\colon [0,1] \to \mathbb{R}$ is the function

$$f(x) = \begin{cases} 0, & x \in [0, \frac{1}{2}), \\ 1, & x \in [\frac{1}{2}, 1]. \end{cases}$$

Note that $f \notin \mathsf{C}^0([0,1];\mathbb{R})$. One might want to conclude that the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ does not converge since it converges pointwise to a discontinuous function. However, we should not really feel so comfortable with our knowledge of the normed vector space $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_1)$ at this point. Thus we prove a lemma that really settles that $(f_j)_{j\in\mathbb{Z}_{>0}}$ does not, in fact, converge.

1 **Lemma** *Consider the sequence* $(\mathrm{f_j})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *as above. If* $\mathrm{g} \in \mathsf{C}^0([0,1];\mathbb{R})$ *is such that the sequence* $(\mathrm{f_j})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *converges to* g *in the norm* $\|\cdot\|_1$*, then*

$$\mathrm{g}(\mathrm{x}) = \begin{cases} 0, & \mathrm{x} \in (0, \frac{1}{2}), \\ 1, & \mathrm{x} \in (\frac{1}{2}, 1). \end{cases}$$

Proof Suppose that $g(x_0) > 0$ for some $x_0 \in [0, \frac{1}{2})$. Then, by continuity of g, there exists $\delta \in \mathbb{R}_{>0}$ such that

$$(x_0 - \delta, x_0 + \delta) \subseteq (0, \tfrac{1}{2})$$

and such that $g(x) > \frac{1}{2}g(x_0)$ for all $x \in (x_0 - \delta, x_0 + \delta)$ has the same sign as $g(x_0)$. Let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $f_N(x) = 0$ for all $x \in (x_0 - \delta, x_0 + \delta)$. It then holds that for $j \geq N$ we have

$$\begin{aligned} \|g - f_j\|_1 &= \int_0^1 |g(x) - f_j(x)| \, \mathrm{d}x \geq \int_{x_0-\delta}^{x_0+\delta} |g(x) - f_j(x)| \, \mathrm{d}x \\ &= \int_{x_0-\delta}^{x_0+\delta} |g(x)| \, \mathrm{d}x \geq \delta g(x_0). \end{aligned}$$

This shows that the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ cannot converge to g if $g(x_0) > 0$ for some $x_0 \in (0, \frac{1}{2})$. A completely similar argument shows that the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ cannot converge to g if $g(x_0) < 0$ for some $x_0 \in (0, \frac{1}{2})$.

Now suppose that $g(x_0) > 1$ for some $x_0(\frac{1}{2}, 1)$. Then there exists $\delta > 0$ such that

$$(x_0 - \delta, x_0 + \delta) \subseteq (\tfrac{1}{2}, 1)$$

and such that $g(x) - 1 > \frac{1}{2}(g(x_0) - 1)$ for all $x \in (x_0 - \delta, x_0 + \delta)$. Then, for any $j \in \mathbb{Z}_{>0}$,

$$\begin{aligned} \|g - f_j\|_1 &= \int_0^1 |g(x) - f_j(x)| \, \mathrm{d}x \geq \int_{x_0-\delta}^{x_0+\delta} |g(x) - f_j(x)| \, \mathrm{d}x \\ &= \int_{x_0-\delta}^{x_0+\delta} |g(x) - 1| \, \mathrm{d}x \geq \delta(g(x_0) - 1). \end{aligned}$$

This shows that the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ cannot converge to g if $g(x_0) > 1$ for some $x_0 \in (\frac{1}{2}, 1)$. A completely similar argument shows that the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ cannot converge to g if $g(x_0) < 1$ for some $x_0 \in (\frac{1}{2}, 1)$. ▼

There is obviously no continuous function satisfying the conditions of the lemma. Thus we have found a Cauchy sequence in $(\mathsf{C}^0([0, 1]; \mathbb{R}), \|\cdot\|_1)$ that does not converge. •

The examples show something very important: that there is a genuine distinction between Cauchy sequences and convergent sequences. Moreover, normed vector spaces where the two notions agree are important.

3.3.2 Definition (Completeness, Banach space) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. A normed $\mathbb{F}$-vector space $(\mathsf{V}, \|\cdot\|)$ is ***complete*** if every Cauchy sequence in V converges. A ***$\mathbb{F}$-Banach space*** is a complete normed $\mathbb{F}$-vector space. •

The following result is important in the same way that completeness of $\mathbb{R}$ is important.

3.3.3 Theorem (Completeness of finite-dimensional normed vector spaces) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and if* $(\mathsf{V}, \|\cdot\|)$ *is a finite-dimensional normed* $\mathbb{F}$*-vector space, then* V *is complete.*

Proof Let $\{e_1, \dots, e_n\}$ be a basis for V which defines an isomorphism $\iota \colon \mathsf{V} \to \mathbb{F}^n$ by

$$\iota(v_1 e_1 + \cdots + v_n e_n) = (v_1, \dots, v_n).$$

Define a norm $\|\cdot\|_2$ on V by $\|v\|_2 = \|\iota(v)\|_2$ where $\|\cdot\|_2$ also denotes the standard norm on $\mathbb{F}^n$. This is a norm, cf. the proof of Proposition 3.1.4. By Theorem 3.1.15 it follows that there exists $C \in \mathbb{R}_{>0}$ such that

$$C^{-1}\|v\|_2 \leq \|v\| \leq C\|v\|_2.$$

Now let $(v_j)_{j \in \mathbb{Z}_{>0}}$ be a Cauchy sequence in V. Let's write

$$v_j = v_{j1} e_1 + \cdots + v_{jn} e_n$$

for $v_{jl} \in \mathbb{F}$, $j \in \mathbb{Z}_{>0}$, $l \in \{1, \dots, n\}$. For $\epsilon \in \mathbb{R}_{>0}$ let $N \in \mathbb{Z}_{>0}$ by such that $\|v_j - v_k\| < C^{-1}\epsilon$ for $j, k \in \mathbb{Z}_{>0}$. We then have

$$C^{-1}\epsilon > \|v_j - v_k\| \geq C^{-1}\|v_j - v_k\|_2 = C^{-1}\Big(\sum_{l=1}^{n} |v_{jl} - v_{kl}|\Big)^{1/2} \geq C^{-1}|v_{jl_0} - v_{kl_0}|$$

for $j, k \geq N$ and for each $l_0 \in \{1, \dots, n\}$. Thus $|v_{jl_0} - v_{kl_0}| < \epsilon$ for $j, k \geq N$ and for each $l_0 \in \{1, \dots, n\}$. Thus $(v_{jl_0})_{j \in \mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathbb{F}$ for each $l_0 \in \{1, \dots, n\}$. Since $\mathbb{F}$ is complete by Theorem II-1.2.5 it follows that there exists $v_{l_0} \in \mathbb{F}$, $l_0 \in \{1, \dots, n\}$, such that $\lim_{j \to \infty} v_{jl_0} v_{l_0}$. Now define $v = v_1 e_1 + \cdots + v_n e_n$. We claim that $(v_j)_{j \in \mathbb{Z}_{>0}}$ converges to v. To see this, for $\epsilon \in \mathbb{R}_{>0}$ let $N \in \mathbb{Z}_{>0}$ be such that $\|v_l - v_{jl}\| < \frac{\epsilon}{C\sqrt{n}}$. Then

$$\|v - v_j\| \leq C\|v - v_j\|_2 = C\Big(\sum_{l=1}^{n} |v_l - v_{jl}|^2\Big)^{1/2} \leq C\Big(\sum_{l=1}^{n} \Big(\frac{\epsilon}{C\sqrt{n}}\Big)^2\Big)^{1/2} = \epsilon,$$

as desired. ■

3.3.2 Why completeness is important

We have now seen completeness arise in three important cases. The first was with the incompleteness of the rational numbers and the second and third were in Example 3.3.1. It is fair to ask, "Who cares whether a normed vector space is complete?" In this section we address this.

First let us consider the simple case of the incompleteness of the rational numbers. Rational numbers are fairly simple to define and pretty easy to understand. Real numbers are somewhat more difficult to define, and we think we understand them only because we live in a world where the notion of a real number has been accepted for so long that they are as integral a part of science as are the integers. However, it is worth reflecting that the notion of numbers that were not rational numbers has not always been as acceptable as it is now. Indeed, the development of mathematics is marked by strong resistance to any of the "unusual" kinds of new numbers that arose, whether they be negative numbers, real numbers, or complex numbers. As concerns real numbers, many Greek mathematicians were dedicated to the existence only of rational numbers. There is an amusing story—completely unsubstantiated by any historical record and thus almost certainly false—that a student of Pythagoras was thrown into the sea for proving that $\sqrt{2}$ was not rational. It is also worth reflecting that, if one is only interested in computation, rational numbers are all one can represent in a digital computer. Thus it is difficult to justify the construction of the real numbers from a purely practical point of view. So why are the real numbers important? They are important precisely because they are complete. It is completeness that makes true "obvious" statements like, "every bounded increasing sequence converges." Relatively simple ideas like continuity and differentiability of functions, the Riemann integral, convergence of sequences of functions, all rely on the completeness of the real numbers for their power. Scientific life would be very difficult and complicated without the completeness of the real numbers.

The point of the above paragraph is this:

1. The real numbers arise in a natural way from the incompleteness of the rational numbers.
2. The completeness of the real numbers is not important for the purposes of computation.
3. The completeness of the real numbers is important for the very basic ideas we use every day concerning real variables and functions of a real variable.
4. You are probably comfortable with the real numbers, but this is only because of societal norms.

Now let us think about the notion of completeness in normed vector spaces. Indeed, let us think specifically about Example 3.3.1–2. In that example we saw that there is a simple Cauchy sequence in $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_1)$ that does not converge. But the sequence of functions certainly converges to a perfectly nice, albeit discontinuous, function. So why not just include this limit function in our set and move on? Well, one can certainly do this, but it also leads to the question, "What are the functions that we need to add to $\mathsf{C}^0([0,1];\mathbb{R})$ in order to be sure that all Cauchy sequences of continuous functions converge?" This is a little like saying that, since $\sqrt{2}$ is irrational, why not just add it to our collection of numbers and move on (the result would be the field extension $\mathbb{Q}(\sqrt{2})$). One could do this, but then eventually one would need to address the matter of what other kinds of numbers need to be added to the rational numbers. Thinking about things in this sort of *ad hoc* way is not satisfying, and is really just faking your way around the real issue, which is this: *one should be sure to always be dealing with complete normed vector spaces*.

The difficulty that arises, as we shall see in Section 3.8.7, is that it is difficult to describe the set of functions that need to be added to $\mathsf{C}^0([0,1];\mathbb{R})$ in order to ensure completeness with respect to the norm $\|\cdot\|_1$. But the point is that just because it is difficult does not mean that it is not important to do. It *is* important to do. Indeed, at some point one *must* do it.

3.3.3 Completeness and direct sums and quotients

In this section we consider how completeness interacts with direct sums and quotients. We first consider direct sums. Recall from Theorem 3.1.17 that if $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$ is a family of normed vector spaces then we define a norm $\|\cdot\|_I$ on the direct sum $\bigoplus_{i\in I}\mathsf{V}_i$ by

$$\|\phi\|_I = \sum_{i\in I}\|\phi(i)\|_i,$$

the sum making sense since it is finite.

3.3.4 Proposition (Completeness of direct sums of Banach spaces) *Let $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and let $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$ be a family of $\mathbb{F}$-Banach spaces. Then $(\bigoplus_{i\in I}\mathsf{V}_i, \|\cdot\|_I)$ is complete if and only if I is finite.*

Proof First suppose that I is finite and so take $I = \{1,\dots,k\}$. Let us denote elements of $\bigoplus_{l=1}^k \mathsf{V}_j$ as $(v_1,\dots,v_k)$. Let $((v_{1j},\dots,v_{kj}))_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\bigoplus_{l=1}^k\mathsf{V}_j$. We claim that, for each $l\in\{1,\dots,k\}$, $(v_{lj})_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in V_l. Let $\epsilon\in\mathbb{R}_{>0}$ and take $N\in\mathbb{Z}_{>0}$ sufficiently large that

$$\|(v_{1j},\dots,v_{kj}) - (v_{1m},\dots,v_{km})\|_I < \epsilon, \qquad j,m\geq N.$$

Since

$$\|(v_{1j},\dots,v_{kj}) - (v_{1m},\dots,v_{km})\|_I = \|v_{1j}-v_{1m}\|_1 + \dots + \|v_{kj}-v_{km}\|_k$$

it follows that

$$\|v_{lj}-v_{lm}\|_l < \epsilon, \qquad j,m\geq N,$$

and so the sequence $(v_{lj})_{j\in\mathbb{Z}_{>0}}$ is indeed Cauchy. Therefore, since V_l is a Banach space, the sequence converges to $v_l\in\mathsf{V}_l$. We next claim that the sequence $((v_{1j},\dots,v_{kj}))_{j\in\mathbb{Z}_{>0}}$ converges to $(v_1,\dots,v_k)$. Indeed, let $\epsilon\in\mathbb{R}_{>0}$ and take $N\in\mathbb{Z}_{>0}$ sufficiently large that

$$\|v_{lj}-v_l\|_l < \tfrac{\epsilon}{k}, \qquad l\in\{1,\dots,k\},\ j\geq N.$$

Then

$$\|(v_{1j},\dots,v_{kj}) - (v_1,\dots,v_k)\|_I = \|v_{1j}-v_1\|_1 + \dots + \|v_{kj}-v_k\|_k < \epsilon,$$

for $j\geq N$, giving the desired convergence.

Next suppose that I is infinite and, for each $i\in I$, choose $v_i\in\mathsf{V}_i$ such that $\|v_i\|_i = 1$. Let $\{i_l\}_{l\in\mathbb{Z}_{>0}}$ be a set of distinct elements of I and then define a sequence $(\phi_k)_{k\in\mathbb{Z}_{>0}}$ in $\bigoplus_{i\in I}\mathsf{V}_i$ by

$$\phi_k(i) = \begin{cases} 2^{-j}v_{i_j}, & i = i_j,\ j\in\{1,\dots,k\},\\ 0, & \text{otherwise.}\end{cases}$$

We claim that $(\phi_k)_{k\in\mathbb{Z}_{>0}}$ is a Cauchy sequence. Indeed, let $\epsilon > 0$ and let $N\in\mathbb{Z}_{>0}$ be such that for $k,m\geq N$ with $m>k$ we have $\sum_{j=k+1}^m < \epsilon$. This is possible since the series

$\sum_{j=1}^{\infty} 2^{-j}$ converges by Example I-2.4.2–1. Now note that, for $k, m \geq N$ with $m > k$ we have

$$\|\phi_k - \phi_m\|_I = \sum_{j=k+1}^{m} \|2^{-j} v_{i_j}\|_{i_j} = \sum_{j=k+1}^{m} 2^{-j} < \epsilon,$$

showing that the sequence $(\phi_k)_{k\in\mathbb{Z}_{>0}}$ is indeed Cauchy. However, the sequence does not converge. Indeed, if $\phi \in \bigoplus_{i\in I} \mathsf{V}_i$ has the property that $\lim_{k\to\infty}\|\phi - \phi_k\|_I = 0$ then this implies that $\phi(i_j) = 2^{-j} v_{i_j}$ for $j \in \mathbb{Z}_{>0}$, cf. Example 3.3.1–1. But then $\phi \notin \bigoplus_{i\in I} \mathsf{V}_i$. ■

In Section 3.8.3 we will revisit the matter of the completeness of direct sums.

For now we turn to quotients. We recall from Proposition 3.1.19 the definition of the norm $\|\cdot\|_{/\mathsf{U}}$ on V/U.

3.3.5 Proposition (Quotients of Banach spaces by closed subspaces are Banach spaces) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, if $(\mathsf{V}, \|\cdot\|)$ is an $\mathbb{F}$-Banach space, and if U is a closed subspace of V, then $(\mathsf{V}/\mathsf{U}, \|\cdot\|_{/\mathsf{U}})$ is an $\mathbb{F}$-Banach space.*

Proof We already know from Proposition 3.1.19 that $(\mathsf{V}/\mathsf{U}, \|\cdot\|_{/\mathsf{U}})$ is a normed vector space, so it is completeness that e must prove here. Let $(v_j + \mathsf{U})_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence. By passing to a subsequence if necessary we can suppose that $\|(v_{j+1} - v_j) + \mathsf{U}\|_{/\mathsf{U}} < 2^{-j}$, $j \in \mathbb{Z}_{>0}$. By definition of $\|\cdot\|_{/\mathsf{U}}$ this means that there exists $u_2 \in \mathsf{U}$ such that $\|v_2 + u_2 - v_1\| < 2^{-1}$. Define $v_2' = v_2 + u_2$. Similarly, there exists $u_3 \in \mathsf{U}$ such that $\|v_3 + u_3 - v_2\| < 2^{-2}$. Define $v_3' = v_3 + u_3$. Proceeding in this way we define a sequence $(v_j')_{j\in\mathbb{Z}_{>0}}$ such that $\|v_{j+1}' - v_j'\| < 2^{-j}$ and such that $v_j' + \mathsf{U} = v_j + \mathsf{U}$ for $j \in \mathbb{Z}_{>0}$. In particular, the sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ is Cauchy and so converges to some $v \in \mathsf{V}$ since V is complete. Then, by Theorem 3.5.2,

$$\lim_{j\to\infty}(v_j' + \mathsf{U}) = (\lim_{j\to\infty} v_j') + \mathsf{U} = v + \mathsf{U}$$

since the projection from V to V/U is continuous. ■

3.3.4 Completions

Having been confronted in Section 3.3.1 with the reality of normed vector spaces that are not complete, and having seen evidence of the importance of completeness in Section 3.3.2, it becomes important to know the answer to this question: "What do we do when we have an incomplete normed vector space?" The answer is: "We complete it!"

The notion of a completion was discussed in detail in Section 1.1.7 for metric spaces. Since normed vector spaces are metric spaces by Proposition 3.1.7, that entire discussion can be transported here to define the completion of a normed vector space. However, we will develop at least some of this discussion independently.

The main result is the following. In the statement of the result we make reference to the notion of an isomorphism of normed vector spaces. We will not formally get to this idea until Section 3.5.2, but let us just say here that an isomorphism of normed vector spaces is an invertible linear map that is continuous and has a continuous inverse.

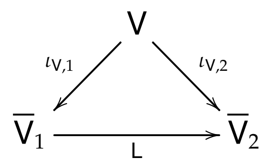
3.3.6 Theorem (Completion of a normed vector space) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. Then there exists a Banach space $(\overline{\mathsf{V}}, \overline{\|\cdot\|})$ with the following properties:*

(i) there exists an injective linear map $\iota_{\mathsf{V}}\colon \mathsf{V} \to \overline{\mathsf{V}}$ such that $\overline{\|\iota_{\mathsf{V}}(\mathrm{v})\|} = \|\mathrm{v}\|$ for every $\mathrm{v} \in \mathsf{V}$;

(ii) for each $\overline{\mathrm{v}} \in \overline{\mathsf{V}}$ there exists a sequence $(\mathrm{v}_j)_{j\in\mathbb{Z}_{>0}}$ in V such that $(\iota_{\mathsf{V}}(\mathrm{v}_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\overline{\mathrm{v}}$.

*Such a Banach space $(\overline{\mathsf{V}}, \overline{\|\cdot\|})$ is a **completion** of $(\mathsf{V}, \|\cdot\|)$.*

Furthermore, if $(\overline{\mathsf{V}}_1, \overline{\|\cdot\|}_1)$ and $(\overline{\mathsf{V}}_2, \overline{\|\cdot\|}_2)$ are two completions of $(\mathsf{V}, \|\cdot\|)$ with $\iota_{\mathsf{V},1}\colon \mathsf{V} \to \overline{\mathsf{V}}_1$ and $\iota_{\mathsf{V},2}\colon \mathsf{V} \to \overline{\mathsf{V}}_2$ being the corresponding injective linear maps, then there exists an isomorphism $\mathsf{L}\colon \overline{\mathsf{V}}_1 \to \overline{\mathsf{V}}_2$ of Banach space such that the following diagram commutes:



Proof Many of the details of this proof follow that of Theorem 1.1.34, and we therefore omit them, only making reference to the existing proof.

We let $\mathrm{CS}(\mathsf{V})$ denote the collection of Cauchy sequences in V. If we define vector addition and scalar multiplication by

$$(u_j)_{j\in\mathbb{Z}_{>0}} + (v_j)_{j\in\mathbb{Z}_{>0}} = (u_j + v_j)_{j\in\mathbb{Z}_{>0}}, \quad a(v_j)_{j\in\mathbb{Z}_{>0}} = (av_j)_{j\in\mathbb{Z}_{>0}},$$

then $\mathrm{CS}(\mathsf{V})$ is an $\mathbb{F}$-vector space by Exercise 3.2.3.

For a Cauchy sequences $(v_j)_{j\in\mathbb{Z}_{>0}}$ let us define

$$\widetilde{\|(v_j)_{j\in\mathbb{Z}_{>0}}\|} = \lim_{j\to\infty} \|v_j\|.$$

To make the connection with the proof of Theorem 1.1.34 we note that we can define

$$\tilde{d}((u_j)_{j\in\mathbb{Z}_{>0}}, (v_j)_{j\in\mathbb{Z}}) = \lim_{j\to\infty} \|u_j - v_j\|.$$

Then we obviously have

$$\widetilde{\|(v_j)_{j\in\mathbb{Z}_{>0}}\|} = \tilde{d}((v_j)_{j\in\mathbb{Z}_{>0}}, (0)_{j\in\mathbb{Z}_{>0}}).$$

This identity can be used to easily prove many of the assertions we are about to make about $\widetilde{\|\cdot\|}$. In particular, the definition of $\widetilde{\|\cdot\|}$ is shown to make sense in that the limit exists. Moreover, $\widetilde{\|\cdot\|}$ is readily seen to be a seminorm on $\mathrm{CS}(\mathsf{V})$. For example, we compute

$$\widetilde{\|a(v_j)_{j\in\mathbb{Z}_{>0}}\|} = \lim_{j\to\infty} \|av_j\| = |a| \lim_{j\to\infty} \|v_j\| = |a| \widetilde{\|(v_j)_{j\in\mathbb{Z}_{>0}}\|}.$$

(Note that in the third step we make use of continuity of the norm which we will prove as Proposition 3.5.4.) The remaining seminorm properties follow just as do the corresponding assertions from Theorem 1.1.34.

We now let $(\overline{\mathsf{V}}, \overline{\|\cdot\|})$ be the normed vector space associated with $(\mathsf{CS}(\mathsf{V}), \widetilde{\|\cdot\|})$ as in Theorem 3.1.8. Note that $(\overline{\mathsf{V}}, \overline{\|\cdot\|})$ as in Theorem 3.1.8 is the normed vector space whose associated metric space is the metric space $(\overline{\mathsf{V}}, \overline{d})$ of Theorem 1.1.28. From Exercise 3.3.4 it immediately follows that $(\overline{\mathsf{V}}, \overline{\|\cdot\|})$ is a Banach space.

Recalling from Theorem 3.1.8 that $\overline{\mathsf{V}}$ is a quotient of $\mathsf{CS}(\mathsf{V})$ by a subspace, denote by $\pi_{\mathsf{V}}\colon \mathsf{CS}(\mathsf{V}) \to \overline{\mathsf{V}}$ the canonical projection. Now define $\iota_{\mathsf{V}}\colon \mathsf{V} \to \overline{\mathsf{V}}$ by $\iota_{\mathsf{V}}(v) = \pi_{\mathsf{V}}((v)_{j\in\mathsf{V}})$. As for the corresponding assertion from Theorem 1.1.34, we readily show that $\overline{\|\iota_{\mathsf{V}}(v)\|} = \|v\|$ for each $v \in \mathsf{V}$. Since the injection ι_{V} of V into $\overline{\mathsf{V}}$ is the same as the injection in the proof of Theorem 1.1.34, it follows from Theorem 1.1.34 that for any $\overline{v} \in \overline{\mathsf{V}}$ there is a sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ for which $(\iota_{\mathsf{V}}(v_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\overline{v}$.

Now we prove the final assertion of the theorem, letting $(\overline{\mathsf{V}}_1, \overline{\|\cdot\|}_1)$ and $(\overline{\mathsf{V}}_1, \overline{\|\cdot\|}_2)$ be completions of $(\mathsf{V}, \|\cdot\|)$. Let $\overline{v}_1 \in \overline{\mathsf{V}}_1$ and let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence for which $(\iota_{\mathsf{V},1})_{j\in\mathbb{Z}_{>0}}$ converges to $\overline{v}_1$. Thus $(\iota_{\mathsf{V},1}(v_j))_{j\in\mathbb{Z}_{>0}}$ is Cauchy. Since $\iota_{\mathsf{V},1}$ preserves the norm, one easily shows that $(v_j)_{j\in\mathbb{Z}_{>0}}$ is Cauchy. Since $\iota_{\mathsf{V},2}$ also preserves the norm, the sequence $(\iota_{\mathsf{V},2}(v_j))_{j\in\mathbb{Z}_{>0}}$ is Cauchy, and so converges since $\overline{\mathsf{V}}_2$ is complete. Let $\overline{v}_2$ denote its limit. We define $\mathsf{L}\colon \overline{\mathsf{V}}_1 \to \overline{\mathsf{V}}_2$ by $\mathsf{L}(\overline{v}_1) = \overline{v}_2$, according to the preceding construction. As with the corresponding assertion in the proof of Theorem 1.1.34, one can show that this definition is independent of the choice of sequence converging to $\overline{v}_1$. Moreover, just as in the proof of Theorem 1.1.34, we can show that L is a bijection and an isometry. Therefore, it is continuous and has a continuous inverse. All that remains is to show that L is linear. To see this, let $\overline{u}_1, \overline{v}_1 \in \overline{\mathsf{V}}_1$ and let $a \in \mathbb{F}$. Let $(u_j)_{j\in\mathbb{Z}_{>0}}$ and $(v_j)_{j\in\mathbb{Z}_{>0}}$ be sequences in V for which $\lim_{j\to\infty} \iota_{\mathsf{V},1}(u_j) = \overline{u}_1$ and $\lim_{j\to\infty} \iota_{\mathsf{V},1}(u_j) = \overline{u}_1$. We then have

Is this obvious?

$$\mathsf{L}(a\overline{v}_1) = \lim_{j\to\infty} \iota_{\mathsf{V},2}(av_j) = a \lim_{j\to\infty} \iota_{\mathsf{V},2}(v_j) = a\mathsf{L}(\overline{v}_1)$$

and

$$\mathsf{L}(\overline{u}_1 + \overline{v}_1) = \lim_{j\to\infty} \iota_{\mathsf{V},2}(u_j + v_j) = \lim_{j\to\infty} \iota_{\mathsf{V},2}(u_j) + \lim_{j\to\infty} \iota_{\mathsf{V},2}(v_j) = \mathsf{L}(\overline{u}_1) + \mathsf{L}(\overline{v}_1),$$

where we have used the continuity properties of the norm as in Proposition 3.5.4 below. ■

The preceding theorem is nice in that the proof is constructive. The completion consists of equivalence classes of Cauchy sequences, just as was the case for the construction of $\mathbb{R}$ in Section I-2.1.2. The problem is that it may not be so easy to understand what elements in the completion "look like." For example, in Example 3.3.1 we gave two instances of incomplete normed vector spaces. For the incomplete normed vector space $(\mathbb{F}_0^\infty, \|\cdot\|_1)$ it is fairly easy to understand the completion; we do this in Section 3.8.2. However, for the incomplete normed vector space $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_1)$ the completion is harder to understand. Indeed, try to imagine what might be the set of limits of all Cauchy sequences in $\mathsf{C}^0([0,1];\mathbb{R})$. Surely these limits can be pretty complicated! And we shall see in Section 3.8.7 that to describe these limits is possible by using Lebesgue's integral that we dedicated so much effort to in Chapter 2. Indeed, many of the examples of Banach spaces in

Section 3.8 are constructed as completions. The diversity of the examples in that section should, alone, convince the reader of the importance of completeness and completions.

Exercises

3.3.1 For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ show that $(\mathbb{F}_0^\infty, \|\cdot\|_\infty)$ (see Example 3.1.3–7) is not complete.

3.3.2 Consider the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ of signals in $C^0([0,1];\mathbb{R})$ as defined in Example 3.3.1–2. In this exercise, use the norm $\|\cdot\|_\infty$.

(a) Show by explicit calculation that the sequence is not a Cauchy sequence.

(b) Is it possible to deduce that the sequence is not Cauchy without doing any calculations?

3.3.3 Consider the sequence $\{f_j\}_{j\in\mathbb{Z}_{>0}}$ of functions in $\mathsf{C}^0([0,1];\mathbb{R})$ defined by $f_j(x) = x^j$. For the vector space $\mathsf{C}^0([0,1];\mathbb{R})$ consider two norms, $\|\cdot\|_\infty$ and $\|\cdot\|_1$, defined by:

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in [0,1]\},$$
$$\|f\|_1 = \int_0^1 |f(x)|\,\mathrm{d}x.$$

Answer the following questions.

(a) Sketch the graphs of the first few functions in the sequence $\{f_j\}_{j\in\mathbb{Z}_{>0}}$.

(b) Is the sequence $\{f_j\}_{j\in\mathbb{Z}_{>0}}$ a Cauchy sequence in $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_\infty)$?

(c) Is the sequence $\{f_j\}_{j\in\mathbb{Z}_{>0}}$ a Cauchy sequence in $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_1)$?

(d) Does the sequence $\{f_j\}_{j\in\mathbb{Z}_{>0}}$ converge in $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_\infty)$?

(e) If the sequence $\{f_j\}_{j\in\mathbb{Z}_{>0}}$ does not converge in $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_\infty)$, does it converge in the completion of $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_\infty)$? If so, to what function does it converge?

(f) Does the sequence $\{f_j\}_{j\in\mathbb{Z}_{>0}}$ converge in $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_1)$?

(g) If the sequence $\{f_j\}_{j\in\mathbb{Z}_{>0}}$ does not converge in $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_1)$, does it converge in the completion of $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_1)$? If so, to what function does it converge?

3.3.4 Show that a normed vector space $(\mathsf{V}, \|\cdot\|)$ is complete if and if the associated metric space (from Proposition 3.1.7) is complete.

3.3.5 Let $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$ be a family of normed vector spaces with $(\bigoplus_{i\in I} \mathsf{V}_i, \|\cdot\|_I)$ the corresponding direct sum normed vector space. Show that, if $(\bigoplus_{i\in I} \mathsf{V}_i, \|\cdot\|_I)$ is complete, then $(\mathsf{V}_i, \|\cdot\|_i)$ is complete for each $i \in I$.

3.3.6 Let $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$ be a family of Banach spaces and define the norm $\|\cdot\|_{I,\infty}$ on $\bigoplus_{i\in I} \mathsf{V}_i$ by

$$\|\phi\|_{I,\infty} = \max\{|\phi(i)| \mid i \in I\}.$$

Show that $(\bigoplus_{i\in I} \mathsf{V}_i, \|\cdot\|_{I,\infty})$ is incomplete if I is infinite.

3.3.7 On the vector space $\mathsf{AC}([a,c];\mathbb{F})$ of $\mathbb{F}$-valued absolutely continuous functions on $[a,b]$, define the function $f \mapsto \|f\|$ by

$$\|f\| = \int_a^b |f(x)|\,\mathrm{d}x.$$

Answer the following questions.

(a) Show that $(\mathsf{AC}([a,b];\mathbb{F}), \|\cdot\|)$ is a normed vector space.

(b) Show that you understand why $(\mathsf{AC}([a,b];\mathbb{F}), \|\cdot\|)$ is not a Banach space by providing a nonconvergent Cauchy sequence.

Section 3.4

Series in normed vector spaces

We now consider series in normed vector spaces. While some of the development here bears a strong resemblance to that for series in $\mathbb{R}$ given in Section I-2.4, there are some significant differences. In particular, we introduce two new notions of convergence, condition and unconditional convergence. The latter of these is equivalent for series in $\mathbb{R}$ to absolute convergence, as we show in Proposition 3.4.5. However, in infinite-dimensions the two notions are not equivalent, and we prove this as the nontrivial Theorem 3.4.8. Much of the rest of the development follows in the same vein as that for series in $\mathbb{R}$.

Do I need to read this section? The reader should understand the notion of a series in a normed vector space since this will be important to us in Section 4.4, which in turn is important in the theory of Fourier series. The material in Section 3.4.2, while interesting, is also somewhat technical and can be skipped at a first reading. The material in Sections 3.4.5 and 3.4.6 can likewise be overlooked until it is needed. •

3.4.1 Definitions and properties of series

A ***series*** in an $\mathbb{F}$-vector space is an expression of the form

$$\sum_{j=1}^{\infty} v_j,$$

where $v_j \in \mathsf{V}$, $j \in \mathbb{Z}_{>0}$. As with series in $\mathbb{R}$ or $\mathbb{C}$, this expression is merely symbolic (but still sensible as a formal expression) unless something can be said about its convergence. For vector spaces without any structure, series can be nothing more than formal. Fortunately, (semi)normed vector spaces have topologies defined on them, and so notions of convergence can be defined. These are as follows.

3.4.1 Definition (Convergence, absolute convergence, and conditional convergence of series) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in V and consider the series

$$S = \sum_{j=1}^{\infty} v_j.$$

The corresponding sequence of ***partial sums*** is the sequence $(S_k)_{k\in\mathbb{Z}_{>0}}$ in V defined by

$$S_k = \sum_{j=1}^{k} v_j.$$

Let $v_0 \in \mathsf{V}$. The series:

(i) is ***Cauchy*** if the sequence of partial sums is a Cauchy sequence;

(ii) ***converges to*** $\mathbf{v_0}$, and we write $\sum_{j=1}^{\infty} v_j = v_0$, if the sequence of partial sums converges to v_0;

(iii) has v_0 as a ***limit*** if it converges to v_0;

(iv) is ***convergent*** if it converges to some member of V;

(v) ***converges absolutely***, or is ***absolutely convergent***, if the series

$$\sum_{j=1}^{\infty} \|v_j\|$$

converges;

(vi) is ***unconditionally Cauchy*** if, for every bijection $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$, the series $S_\phi = \sum_{j=1}^{\infty} v_{\phi(j)}$ is Cauchy;

(vii) ***converges unconditionally***, or is ***unconditionally convergent***, if, for every bijection $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$, the series $S_\phi = \sum_{j=1}^{\infty} v_{\phi(j)}$ converges;

(viii) is ***conditionally Cauchy*** if it is not unconditionally Cauchy;

(ix) ***converges conditionally***, or is ***conditionally convergent***, if it is not unconditionally convergent;

(x) ***diverges*** if it does not converge. •

There are a few differences between the definitions we give here and those for given in Definition I-2.4.1 for series of real numbers. These differences have real substance, so let us record why they arise.

1. In Definition I-2.4.1 we did not have the notion of Cauchy series. This is because this is not necessary for series in $\mathbb{R}$ since Cauchy sequences converge. However, in infinite-dimensional normed vector spaces there may well be non-convergent Cauchy sequences. Therefore, it is useful to distinguish between Cauchy sequences of partial sums and convergent sequences of partial sums. Whenever possible we state results for Cauchy series rather than convergent series, keeping in mind that convergent series are Cauchy.
2. There is a difference between the notions of conditional convergence for series in normed vector spaces and for real numbers as given in Definition I-2.4.1. There is some substance to this difference, and we shall explore this in Section 3.4.2, particularly Theorem 3.4.8.

Just as for series of real numbers and complex numbers, there is a useful relationship between the norm of a sum and the sum of the norms.

3.4.2 Proposition (Swapping summation and norm) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. For a sequence $(\mathrm{v}_j)_{j \in \mathbb{Z}_{>0}}$, if the series $\mathrm{S} = \sum_{j=1}^{\infty} \mathrm{v}_j$ is absolutely convergent, then*

$$\Big\|\sum_{j=1}^{\infty} \mathrm{v}_j\Big\| \leq \sum_{j=1}^{\infty} \|\mathrm{v}_j\|.$$

Proof Define

$$S_m^1 = \Big\|\sum_{j=1}^{m} v_j\Big\|, \quad S_m^2 = \sum_{j=1}^{m} \|v_j\|, \qquad m \in \mathbb{Z}_{>0}.$$

By Exercise 3.4.1 we have $S_m^1 \leq S_m^2$ for each $m \in \mathbb{Z}_{>0}$. Moreover, by Proposition 3.4.5 the sequences $(S_m^1)_{m \in \mathbb{Z}_{>0}}$ and $(S_m^2)_{m \in \mathbb{Z}_{>0}}$ are Cauchy sequences in $\mathbb{R}$ and so converge. It is then clear that

$$\lim_{m \to \infty} S_m^1 \leq \lim_{m \to \infty} S_m^2,$$

which is the result. ■

While we do not have for series in normed vector spaces the bevy of tests for convergence, we do have the obvious sufficient condition.

3.4.3 Proposition (Sufficient condition for a series to diverge) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. If the sequence $(\|\mathrm{v}_j\|)_{j \in \mathbb{Z}_{>0}}$ does not converge to zero, then the series $\sum_{j=1}^{\infty} \mathrm{v}_j$ diverges.*

Proof Suppose that the series $\sum_{j=1}^{\infty} v_j$ converges to v_0 and let $(S_k)_{k \in \mathbb{Z}_{>0}}$ be the sequence of partial sums. Then $v_k = S_k - S_{k-1}$. Then

$$\lim_{k \to \infty} v_k = \lim_{k \to \infty} S_k - \lim_{k \to \infty} S_{k-1} = v_0 - v_0 = 0_{\mathsf{V}},$$

as desired. ■

3.4.2 Absolute and unconditional convergence

In this section we explore the relationship between absolute and unconditional convergence. For finite-dimensional normed vector spaces we will see that the two notions are equivalent.

Let us begin by showing why unconditional convergence is useful, in the same way we showed that absolute convergence is useful in Theorem I-2.4.5.

3.4.4 Proposition (Unconditional limits are rearrangement independent) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. If the series $\sum_{j=1}^{\infty} \mathrm{v}_j$ is unconditionally convergent and converges to v_0, then, for any bijection $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$, the series $\sum_{j=1}^{\infty} \mathrm{v}_{\phi(j)}$ also converges to v_0.*

Proof In order to avoid duplication of part of the proof, we make use of the implication (ii) $\Longrightarrow$ (i) of Theorem 3.4.20. We do this in the following way. Let $S = \sum_{j=1}^{\infty} v_j$. Since S is unconditionally convergent it is unconditionally Cauchy by Proposition 3.2.3. By the implication (ii) $\Longrightarrow$ (i) of Theorem 3.4.20 it follows that

$\sum_{j\in\mathbb{Z}_{>0}} v_j$ is Cauchy in the sense of Definition 3.4.16. Now let $\epsilon \in \mathbb{R}_{>0}$ and let $I \subseteq \mathbb{Z}_{>0}$ be a finite set with the property that

$$\Big\|\sum_{j\in J} v_j\Big\| < \frac{\epsilon}{2}$$

for any finite set J such that $J \cap I = \emptyset$. Now let $N_1 \in \mathbb{Z}_{>0}$ be such that

$$\Big\|\sum_{j=1}^{k} v_j - v_0\Big\| < \frac{\epsilon}{2}$$

for every $k \geq N_1$ (this being possible since $\sum_{j=1}^{\infty} v_j$ converges to v_0) and such that $I \subseteq \{1, \dots, N_1\}$. Let $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a bijection and choose $N_2 \in \mathbb{Z}_{>0}$ sufficiently large that $\{1, \dots, N_1\} \subseteq \{\phi(1), \dots, \phi(N_2)\}$. Then we write

$$\{\phi(1), \dots, \phi(N_2)\} = \{1, \dots, N_1\} \cup J$$

where $J \cap \{1, \dots, N_1\} = \emptyset$. Note that $J \cap I = \emptyset$ since $I \subseteq \{1, \dots, N_1\}$. Therefore, we compute

$$\Big\|\sum_{j=1}^{N_2} v_{\phi(j)} - v_0\Big\| = \Big\|\sum_{j=1}^{N_1} v_j + \sum_{j\in J} v_j - v_0\Big\| \leq \Big\|\sum_{j=1}^{N_1} v_j - v_0\Big\| + \Big\|\sum_{j\in J} v_j\Big\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

giving convergence of $\sum_{j=1}^{\infty} v_{\phi(j)}$ to v_0. ■

As with series in $\mathbb{R}$, one of the essential features of absolutely convergent series is that their convergence is independent of rearrangement of terms. This mirrors the situation for series in $\mathbb{R}$.

3.4.5 Proposition (Absolute convergence implies unconditional Cauchy) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. For a sequence $(\mathrm{v_j})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ consider the series $\mathrm{S} = \sum_{\mathrm{j}=1}^{\infty} \mathrm{v_j}$. If S is absolutely convergent then it is unconditionally Cauchy. Moreover, if S converges then, for any bijection $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$, the series $\mathrm{S}_\phi = \sum_{\mathrm{j}=1}^{\infty} \mathrm{v}_{\phi(\mathrm{j})}$ converges absolutely to the same limit as S.*

Proof Let $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a bijection. First let us show that S_ϕ is absolutely convergent. Since S is absolutely convergent the sequence $(|S|_k)_{k\in\mathbb{Z}_{>0}}$ defined by

$$|S|_k = \sum_{j=1}^{k} \|v_j\|$$

is bounded and monotonically increasing. Thus there exists $M \in \mathbb{R}_{>0}$ such that $|S|_k \leq M$ for every $k \in \mathbb{Z}_{>0}$. Now define the sequence $(|S_\phi|_k)_{k\in\mathbb{Z}_{>0}}$ by

$$|S_\phi|_k = \sum_{j=1}^{k} v_{\phi(j)}.$$

For $k \in \mathbb{Z}_{>0}$ there exists $N \in \mathbb{Z}_{>0}$ such that $\{\phi(1), \dots, \phi(k)\} \subseteq \{1, \dots, N\}$. Then

$$|S_\phi|_k \leq \sum_{j=1}^{N} \|v_j\| \leq M.$$

Thus $(|S_\phi|_k)_{k \in \mathbb{Z}_{>0}}$ is bounded and monotonically increasing, and so convergent. Thus S_ϕ is absolutely convergent.

Next we show that if S is absolutely convergent then it is unconditionally Cauchy. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N_1 \in \mathbb{Z}_{>0}$ be such that

$$\sum_{j=N_1}^{\infty} \|v_j\| < \epsilon.$$

Now let $N_2 \in \mathbb{Z}_{>0}$ be such that $\{\phi(1), \dots, \phi(N_1)\} \subseteq \{1, \dots, N_2\}$. Let $k, l \geq N_2$ with $l > k$ and note that if $j \in \{k+1, \dots, l\}$ then $\phi^{-1}(j) \geq N_1$. Thus

$$\Big\| \sum_{j=k+1}^{l} v_{\phi(j)} \Big\| \leq \sum_{j=k+1}^{l} \|v_{\phi(j)}\| \leq \sum_{j=N_1}^{\infty} \|v_j\| < \epsilon,$$

showing that S_ϕ is Cauchy.

Now suppose that S converges to v_0 and let us show that S_ϕ converges to v_0. For $\epsilon \in \mathbb{R}_{>0}$ let $N_1 \in \mathbb{Z}_{>0}$ be such that

$$\Big\| \sum_{j=1}^{N_1} v_j - v_0 \Big\| < \frac{\epsilon}{2}$$

(this is possible since S converges to v_0) and such that

$$\sum_{j=N_1}^{\infty} \|v_j\| < \frac{\epsilon}{2} \tag{3.1}$$

(this is possible since S is absolutely convergent). There then exists $N_2 \in \mathbb{Z}_{>0}$ such that $\{\phi(1), \dots, \phi(N_1)\} \subseteq \{1, \dots, N_2\}$. Then

$$\sum_{j=1}^{N_2} v_{\phi(j)} = \sum_{j=1}^{N_1} v_j + \sum_{j \in J} v_{\phi(j)},$$

where $J = \{1, \dots, N_2\} \setminus \{\phi(1), \dots, \phi(N_1)\}$. Note that

$$\sum_{j \in J} \|v_{\phi(j)}\| \leq \sum_{j=N_1}^{\infty} \|v_j\| < \frac{\epsilon}{2}$$

by (3.1). Then

$$\Big\|\sum_{j=1}^{N_2} v_{\phi(j)} - v_0\Big\| = \Big\|\sum_{j=1}^{N_1} v_j + \sum_{j\in J} v_{\phi(j)} - v_0\Big\| \leq \Big\|\sum_{j=1}^{N_1} v_j - v_0\Big\| + \Big\|\sum_{j\in J} v_{\phi(j)}\Big\| \leq \frac{\epsilon}{2} + \sum_{j\in J} \|v_{\phi(j)}\| \leq \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon,$$

giving convergence of S_ϕ to v_0 as desired. ■

Thus Proposition 3.4.5 says that absolute convergence implies unconditional convergence. We shall see below in Theorem 3.4.8 that the two notions are equivalent if and only if the normed vector space is finite-dimensional. Thus the notion of unconditional convergence is the more general notion, and one may wonder whether absolute convergence is important. It is, and here is why.

3.4.6 Theorem (Absolute convergence and completeness) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. A normed $\mathbb{F}$-vector space $(\mathsf{V}, \|\cdot\|)$ is complete if and only if every absolutely convergent series in V converges.*

Proof Suppose that V is complete and let $\sum_{j=1}^{\infty} v_j$ be an absolutely convergent series. From Proposition 3.4.5 it follows that $\sum_{j=1}^{\infty} v_j$ is Cauchy, and so it converges since V is complete.

Now suppose that every absolutely convergent series converges, and let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence. Choose a subsequence $(v_{j_k})_{k\in\mathbb{Z}_{>0}}$ for which $\|u_{j_{k+1}} - u_{j_k}\| < \frac{1}{2^{k+1}}$. Then define $u_1 = v_{k_1}$ and $u_k = v_{j_k} - v_{j_{k-1}}$ so that the series $\sum_{k=1}^{\infty} u_k$ is absolutely convergent, and so convergent. This means therefore that

$$\lim_{k\to\infty} \|u_k\| = \lim_{k\to\infty} \|v_{j_k} - v_{j_{k-1}}\| = 0.$$

Thus the sequence $(v_{j_k})_{k\in\mathbb{Z}_{>0}}$ is convergent. Suppose it converges to v. Now, for $\epsilon > 0$ choose k and j sufficiently large that $\|v_j - v_{j_k}\| < \frac{\epsilon}{2}$ and $\|v - v_{j_k}\| < \frac{\epsilon}{2}$. Then we have

$$\|v - v_j\| \leq \|v - v_{j_k}\| + \|v_{j_k} - v_j\| < \epsilon,$$

so showing that $(v_j)_{j\in\mathbb{Z}_{>0}}$ converges to v. ■

The following trivial corollary is sometimes useful by itself.

3.4.7 Corollary (Absolutely convergent sequences in Banach spaces converge) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. If $(\mathsf{V}, \|\cdot\|)$ is a $\mathbb{F}$-Banach space and if $\sum_{\mathrm{j}=1}^{\infty} \mathrm{v_j}$ is an absolutely convergent series in V, then $\sum_{\mathrm{j}=1}^{\infty} \mathrm{v_j}$ is convergent.*

Now let us explore the possibility of a converse to Proposition 3.4.5. That is, let us consider the question, "Is it true that an unconditionally convergent series is absolutely convergent?" In Theorem I-2.4.5 we saw that this was true for series in $\mathbb{R}$. However, this is not generally true in normed vector spaces, but holds if and only if the vector space is finite-dimensional. This is an instance of where the difference between finite- and infinite-dimensions shows up.

3.4.8 Theorem (Absolute convergence and unconditional Cauchy agree (only) in finite-dimensions) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. Then the set of absolutely convergent series and the set of unconditionally Cauchy series coincide if and only if V is finite-dimensional.*

Proof From Proposition 3.4.5 we know that absolutely convergent series are always unconditionally convergent. Suppose that V is finite-dimensional and that $\sum_{j=1}^{\infty} v_j$ is unconditionally convergent. Let us also suppose that $\mathbb{F} = \mathbb{R}$ for the moment. Choose a basis $\{e_1, \dots, e_n\}$ for V and write

$$v_j = v_j^1 e_1 + \dots + v_j^n e_n$$

for $v_j^l \in \mathbb{F}$, $j \in \mathbb{Z}_{>0}$, $l \in \{1, \dots, n\}$. By Theorem 3.1.15 we can use any norm on V we wish to discuss convergence, so let us use the ∞-norm induced by the basis:

$$\|v^1 e_1 + \dots + v^n e_n\| = \max\{|v^1|, \dots, |v^n|\}.$$

Let $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a bijection so that $\sum_{j=1}^{\infty} v_{\phi(j)}$ converges, say to $v_0 \in \mathsf{V}$. Let us write

$$v_0 = v_0^1 e_1 + \dots + v_0^n e_n.$$

Now let $\epsilon \in \mathbb{R}_{>0}$ and choose $N \in \mathbb{Z}_{>0}$ such that

$$\Big\|\sum_{j=1}^{N} v_{\phi(j)} - v_0\Big\| < \epsilon.$$

Then

$$\Big|\sum_{j=1}^{N} v_{\phi(j)}^l - v_0^l\Big| \le \Big\|\sum_{j=1}^{N} v_{\phi(j)} - v_0\Big\| < \epsilon.$$

Thus $\sum_{j=1}^{\infty} v_{\phi(j)}^l$ converges to v_0^l for each $l \in \{1, \dots, n\}$. Thus $\sum_{j=1}^{\infty} v_j^l$ is unconditionally convergent, and so absolutely convergent by Theorem I-2.4.5. Now again let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that

$$\sum_{j=N+1}^{\infty} |v_j^l| < \epsilon, \qquad l \in \{1, \dots, n\},$$

this being possible by absolute convergence of $\sum_{j=1}^{\infty} v_j^l$. Then, for any $l \in \{1, \dots, n\}$,

$$\sum_{j=N+1}^{\infty} \|v_j\| \le \sum_{j=N+1}^{\infty} |v_j^l| < \epsilon,$$

giving absolute convergence of $\sum_{j=1}^{\infty} v_j$.

If V is a finite-dimensional $\mathbb{C}$-vector space, then it is also a finite-dimensional $\mathbb{R}$-vector space of twice the dimension, and so the above arguments can be used to show that an unconditionally convergent sum is absolutely convergent.

It remains to show that if V is infinite-dimensional then there exists an unconditionally convergent series that is not absolutely convergent. We do this via a sequence of lemmata, the first of which seems to have nothing to do with the problem at hand. Let us suppose that $\mathbb{F} = \mathbb{R}$.

The following lemma is crucial, and is called the ***Dvoretzky–Rogers Lemma***.

1 Lemma *Let* $\mathrm{C} \subseteq \mathbb{R}^{\mathrm{n}}$ *be a compact convex set with nonempty interior and with centre at* $\mathbf{0}_{\mathbb{R}^{\mathrm{n}}}$ *and let* $\mathrm{k} \in \{1, \dots, \mathrm{n}\}$*. Then there exists* $\mathbf{x}_1, \dots, \mathbf{x}_{\mathrm{n}} \in \mathrm{bd}(\mathrm{C})$ *such that, for any* $\lambda_1, \dots, \lambda_{\mathrm{k}} \in \mathbb{R}$*,*

$$\lambda_1 \mathbf{x}_1 + \dots + \lambda_{\mathrm{k}} \mathbf{x}_{\mathrm{k}} \in \lambda \mathrm{C} \triangleq \{\lambda \mathbf{x} \mid \mathbf{x} \in \mathrm{C}\},$$

where

$$\lambda^2 = \left(2 + \tfrac{\mathrm{k}(\mathrm{k}-1)}{\mathrm{n}}\right)(\lambda_1^2 + \dots + \lambda_{\mathrm{k}}^2).$$

Proof By Theorem 5.1.6 let E be the ellipsoid with largest volume contained in C. If $\boldsymbol{A} \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ is invertible then hypotheses of the lemma hold for the convex set $\boldsymbol{A}(C)$ and the conclusions hold for the points $\boldsymbol{A}\boldsymbol{x}_1, \dots, \boldsymbol{A}\boldsymbol{x}_n$. Thus we can apply an invertible linear transformation of $\mathbb{R}^n$ to the problem without changing either the hypotheses or the conclusions. Let us suppose that $\boldsymbol{A}$ has been chosen such that $\boldsymbol{A}(E) = \overline{\mathsf{B}}(1, \mathbf{0}_{\mathbb{R}^n})$, the closed unit ball in the 2-norm in $\mathbb{R}^n$. For the remainder of the proof we work with the transformed problem.

We next claim that there exists an orthogonal matrix $\boldsymbol{R}$, thought of as a linear mapping from $\mathbb{R}^n$ to itself, and points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n \in \overline{\mathsf{B}}(1, \mathbf{0}_{\mathbb{R}^n}) \cap C$ such that

$$\boldsymbol{y}_j \triangleq \boldsymbol{R}\boldsymbol{x}_j = (y_j^1, \dots, y_j^j, 0, \dots, 0), \qquad j \in \{1, \dots, n\}, \tag{3.2}$$

(i.e., the last $n - j$ components of $\boldsymbol{R}\boldsymbol{x}_j$ are zero) and such that

$$(y_j^1)^2 + \dots + (y_j^{j-1})^2 = 1 - (y_j^j)^2 \le \tfrac{j-1}{n}, \qquad j \in \{1, \dots, n\}. \tag{3.3}$$

We construct the points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ inductively. For $j = 1$ we take $\boldsymbol{x}_1 \in \overline{\mathsf{B}}(1, \mathbf{0}_{\mathbb{R}^n}) \cap C$ (this is possible by our initial definition of E). We then make an orthogonal change of basis for which $\boldsymbol{x}_1$ is the first basis vector. This defines an orthogonal transformation $\boldsymbol{R}_1$ satisfying (3.2) and (3.3) for $j = 1$. Suppose now, for $k - 1 < n$, that we have defined $\boldsymbol{R}_{k-1}$ and $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1} \in \overline{\mathsf{B}}(1, \mathbf{0}_{\mathbb{R}^n}) \cap C$ satisfying (3.2) and (3.3) for $j \in \{1, \dots, k-1\}$. Define $f_k \colon \mathbb{R}_{\ge 0} \times \mathbb{R}^n \to \mathbb{R}$ by

$$f_k(\epsilon, \boldsymbol{x}) = (1 + \epsilon)^{n-k+1}((y^1)^2 + \dots + (y^{k-1})^2) + (1 + \epsilon + \epsilon^2)^{-k+1}((y^k)^2 + \dots + (y^n)^2),$$

where $\boldsymbol{y} = \boldsymbol{R}_{k-1}\boldsymbol{x}$. For $\epsilon \in \mathbb{R}_{\ge 0}$ define

$$E_\epsilon = \{\boldsymbol{x} \in \mathbb{R}^n \mid f(\epsilon, \boldsymbol{x}) \le 1\}.$$

Thus E_ϵ is an ellipsoid. We claim that for $\epsilon \in \mathbb{R}_{>0}$ the volume of E_ϵ exceeds that of $\overline{\mathsf{B}}(1, \mathbf{0}_{\mathbb{R}^n})$. To see this, consider the linear transformation $\boldsymbol{T}_\epsilon$ of $\mathbb{R}^n$ defined by

$$\boldsymbol{T}_\epsilon(y^1, \dots, y^n) = \Big(\sqrt{(1+\epsilon)^{n-k+1}}y^1, \dots, \sqrt{(1+\epsilon)^{n-k+1}}y^{k-1}, \\ \sqrt{(1+\epsilon+\epsilon^2)^{-k+1}}y^k, \dots, \sqrt{(1+\epsilon+\epsilon^2)^{-k+1}}y^n\Big).$$

Thus $\boldsymbol{T}_\epsilon(E_\epsilon) = \overline{\mathsf{B}}(1, \mathbf{0}_{\mathbb{R}^n})$. Using the change of variables formula for the integral in $\mathbb{R}^n$ we have the volume of E_ϵ as $\det \boldsymbol{T}_\epsilon^{-1}$ times the volume of $\mathsf{B}(1, \mathbf{0}_{\mathbb{R}^n})$. Since

$$\det \boldsymbol{T}_\epsilon^{-1} = \Big(\frac{1+\epsilon+\epsilon^2}{1+\epsilon}\Big)^{(n-k+1)(k-1)/2} > 1$$

for $\epsilon \in \mathbb{R}_{>0}$, we indeed have the volume of E_ϵ as exceeding that of $\overline{\mathsf{B}}(1, \mathbf{0}_{\mathbb{R}^n})$.

Now, since $\mathsf{B}(1, \mathbf{0}_{\mathbb{R}^n})$ is the largest ellipsoid contained in C, there exists a point $\boldsymbol{x}_\epsilon \in \mathrm{bd}(C) \cap E_\epsilon$. Since $\boldsymbol{x}_\epsilon \in \mathrm{bd}(C)$ and since $\overline{\mathsf{B}}(1, \mathbf{0}_{\mathbb{R}^n}) \subseteq C$ it follows that $\|\boldsymbol{x}_\epsilon\| \geq 1$ (where $\|\cdot\|$ is the 2-norm on $\mathbb{R}^n$). Letting $\boldsymbol{y}_\epsilon = \boldsymbol{R}_{k-1}\boldsymbol{x}_\epsilon$ we have

$$((y_\epsilon^1)^2 + \dots + (y_\epsilon^{k-1})^2) + ((y_\epsilon^k)^2 + \dots + (y_\epsilon^n)^2) \geq 1.$$

Subtracting this inequality from the inequality $f(\epsilon, \boldsymbol{x}_\epsilon) \leq 1$ gives

$$((1+\epsilon)^{n-k+1} - 1)((y_\epsilon^1)^2 + \dots + (y_\epsilon^{k-1})^2) \\ + ((1+\epsilon+\epsilon^2)^{-k+1} - 1)((y_\epsilon^k)^2 + \dots + (y_\epsilon^n)^2) \leq 0. \quad (3.4)$$

Let $(\epsilon_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}_{>0}$ converging to zero. The resulting sequence $(\boldsymbol{x}_{\epsilon_j})_{j\in\mathbb{Z}_{>0}}$ is in $\mathrm{bd}(C)$ which is compact, being a closed subset of a compact set (Corollary II-1.2.36). Therefore, by the Bolzano–Weierstrass Theorem, there exists a subsequence of $(\boldsymbol{x}_{\epsilon_j})_{j\in\mathbb{Z}_{>0}}$ converging to some $\boldsymbol{x}_0 \in \overline{\mathsf{B}}(1, \mathbf{0}_{\mathbb{R}^n}) \cap \mathrm{bd}(C)$. Moreover, denoting $\boldsymbol{y}_0 = \boldsymbol{R}\boldsymbol{x}_0$, (3.4) gives

$$\begin{aligned} & \frac{1}{\epsilon}((1+\epsilon)^{n-k+1} - 1)((y_\epsilon^1)^2 + \dots + (y_\epsilon^{k-1})^2) \\ & \quad + ((1+\epsilon+\epsilon^2)^{-k+1} - 1)((y_\epsilon^k)^2 + \dots + (y_\epsilon^n)^2) \leq 0 \\ \implies \quad & \lim_{\epsilon\downarrow 0} \frac{1}{\epsilon}((1+\epsilon)^{n-k+1} - 1)((y_\epsilon^1)^2 + \dots + (y_\epsilon^{k-1})^2) \\ & \quad + ((1+\epsilon+\epsilon^2)^{-k+1} - 1)((y_\epsilon^k)^2 + \dots + (y_\epsilon^n)^2) \leq 0 \\ \implies \quad & (n-k+1)((y_0^1)^2 + \dots + (y_0^{k-1})^2) + (-k+1)((y_0^k)^2 + \dots + (y_0^n)^2) \leq 0. \end{aligned}$$

Now define $\boldsymbol{x}_k = \boldsymbol{x}_0$. If $\boldsymbol{R}_{k-1}\boldsymbol{x}_k \in \mathrm{span}_{\mathbb{R}}(\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1})$ then clearly the last $n-k$ components of $\boldsymbol{R}_{k-1}\boldsymbol{x}_k$ are zero in the basis defined by $\boldsymbol{R}_{k-1}$. If not, then the vectors $\{\boldsymbol{R}_{k-1}^{-1}\boldsymbol{e}_1, \dots, \boldsymbol{R}_{k-1}^{-1}\boldsymbol{e}_{k-1}, \boldsymbol{x}_k\}$ span a subspace of dimension k and by choosing an orthogonal complement in this subspace to $\mathrm{span}_{\mathbb{R}}(\boldsymbol{R}_{k-1}^{-1}\boldsymbol{e}_1, \dots, \boldsymbol{R}_{k-1}^{-1}\boldsymbol{e}_{k-1})$ we get an orthonormal basis for $\mathbb{R}^n$ where the first $k-1$ basis vectors are those defined by $\boldsymbol{R}_{k-1}$ and the first k basis vectors span a subspace containing $\boldsymbol{x}_k$. Thus the last $n-k$ components of $\boldsymbol{x}_k$ in this basis will be zero, and the components of $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1}$ will be unchanged from those in the basis defined by $\boldsymbol{R}_{k-1}$. This new orthonormal basis defines an orthogonal matrix $\boldsymbol{R}_k$. This gives condition (3.2). Moreover, if we abuse notation slightly and denote by $(y^1,\dots,y^n)$ the coordinates in the basis defined by $\boldsymbol{R}_k$, the point $\boldsymbol{y}_k = \boldsymbol{R}_k\boldsymbol{x}_k$ satisfies

$$(n-k+1)((y_k^1)^2+\dots+(y_k^{k-1})^2)+(-k+1)(y_k^k)^2 \le 0.$$

Since we also have $(y_k^1)^2+\dots+(y_k^k)^2=1$ we then get

$$(y_k^1)^2+\dots+(y_k^{k-1})^2 = 1-(y_k^k)^2 = \frac{k-1}{n},$$

and so (3.3) also holds.

Finally, let $\lambda_1,\dots,\lambda_k\in\mathbb{R}$. We compute the square of the length of $\lambda_1\boldsymbol{x}_1+\dots+\lambda_k\boldsymbol{x}_k$ as

$$\begin{aligned}\sum_{j=1}^k\Big(\sum_{l=1}^k\lambda_l y_l^j\Big)^2 &\le \sum_{j=1}^k\Big(2\lambda_j^2(y_j^j)^2+2\Big(\sum_{l=j+1}^k\lambda_l y_l^j\Big)^2\Big)\\ &\le 2\sum_{j=1}^k\Big(\lambda_j^2(y_j^j)^2+\Big(\sum_{l=j+1}^k\lambda_l^2\Big)\Big(\sum_{m=j+1}^k(y_m^j)^2\Big)\Big)\\ &= 2\sum_{j=1}^k\Big((y_j^j)^2+\sum_{l=1}^k\sum_{m=1}^{\min\{j-1,l-1\}}(y_l^m)^2\Big)\lambda_j^2.\end{aligned}$$

Since (3.3) holds we have

$$(y_j^j)^2+\sum_{l=1}^k\sum_{m=1}^{\min\{j-1,l-1\}}(y_l^m)^2 \le 1+\sum_{l=1}^k\frac{l-1}{n},\qquad j\in\{1,\dots,k\}.$$

Therefore, the length of $\lambda_1\boldsymbol{x}_1+\dots+\lambda_k\boldsymbol{x}_k$ is bounded above by

$$\sum_{j=1}^k\Big(1+\sum_{l=1}^k\frac{l-1}{n}\Big)\lambda_j^2 = \Big(2+\frac{k(k-1)}{n}\Big)\sum_{j=1}^k\lambda_j^2.$$

In other words, $\lambda_1\boldsymbol{x}_1+\dots+\lambda_k\boldsymbol{x}_k\in\overline{\mathsf{B}}(\lambda,\boldsymbol{0}_{\mathbb{R}^n})$ where

$$\lambda^2=\Big(2+\frac{k(k-1)}{n}\Big)\sum_{j=1}^k\lambda_j^2.$$

Thus $\lambda_1\boldsymbol{x}_1+\dots+\lambda_k\boldsymbol{x}_k\in\lambda C$ since $\overline{\mathsf{B}}(1,\boldsymbol{0}_{\mathbb{R}^n})\subseteq C$. ▼

2 Lemma *Let* $(\mathsf{V}, \|\cdot\|)$ *be an infinite-dimensional normed* $\mathbb{R}$*-vector space, let* $k \in \mathbb{Z}_{>0}$*, and let* $c_1, \dots, c_k \in \mathbb{R}_{>0}$*. Then there exists* $v_1, \dots, v_k \in \mathsf{V}$ *such that*

(i) $\|v_j\|^2 = c_j$, $j \in \{1, \dots, k\}$, *and*

(ii) $\left\|\sum_{j\in J} v_j\right\|^2 \le 3\sum_{j\in J} c_j$ *for every subset* $J \subseteq \{1, \dots, k\}$.

Proof Let $n = k(k-1)$ and let $u_1, \dots, u_n \in \mathsf{V}$ be linearly independent. Define

$$C = \left\{(x^1, \dots, x^n) \in \mathbb{R}^n \;\middle|\; \|x^1 u_1 + \dots + x^n u_n\| \le 1\right\}.$$

We claim that C is convex, compact, has nonempty interior, and has centre $\mathbf{0}_{\mathbb{R}^n}$. One sees this as follows. The map

$$\mathsf{L}\colon (x^1, \dots, x^n) \mapsto x^1 u_1 + \dots + x^n u_n$$

is a linear injection of $\mathbb{R}^n$ onto the n-dimensional subspace spanned by $u_1, \dots, u_n$. One can then define a norm on $\mathbb{R}^n$ to be the norm induced from the restriction of the norm in V to the subspace $\mathsf{L}(\mathbb{R}^n)$. The closed unit ball in this norm is simply C. Then $\mathsf{L}(C)$ is the intersection of the closed unit ball in V with the subspace $\mathsf{L}(\mathbb{R}^n)$. Thus $\mathsf{L}(C)$ is the intersection of convex sets and so is convex by Exercise II-1.9.4. Moreover, $\mathsf{L}(C)$ is clearly a closed and bounded subset of $\mathsf{L}(\mathbb{R}^n)$ and so is compact by the Heine–Borel Theorem. The unit ball in any norm clearly has nonempty interior (see Exercise 3.1.1). Also, $\mathbf{0}_{\mathbb{R}^n}$ is the centre of C since $\boldsymbol{x} \in C$ if and only if $-\boldsymbol{x} \in C$.

Let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ be as in Lemma 1 and define

$$v_j = \sqrt{c_j}\mathsf{L}(\boldsymbol{x}_j), \qquad j \in \{1, \dots, k\},$$

where $\mathsf{L}\colon \mathbb{R}^n \to \mathsf{V}$ is the map from the preceding paragraph. Then

$$\|v_j\|^2 = c_j\|x_j^1 u_1 + \dots + x_j^n u_n\| = c_j, \qquad j \in \{1, \dots, k\},$$

since $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathrm{bd}(C)$. Now let $J \subseteq \{1, \dots, k\}$. Then, by Lemma 1,

$$\sum_{j\in J} \sqrt{c_j}\boldsymbol{x}_j \in \lambda C$$

where

$$\lambda^2 = \left(2 + \frac{k(k-1)}{n}\right)\sum_{j\in J} c_j = 3\sum_{j\in J} c_j.$$

This implies that

$$\mathsf{L}\Big(\sum_{j\in J} \sqrt{c_j}\boldsymbol{x}_j\Big) \in \mathsf{L}(\lambda C) \quad\Longrightarrow\quad \Big\|\sum_{j\in J} v_j\Big\| \le \Big(3\sum_{j\in J} c_j\Big)^{1/2},$$

as claimed. ▼

3 Lemma *Let $(\mathsf{V}, \|\cdot\|)$ be an infinite-dimensional normed $\mathbb{R}$-vector space and let $\sum_{j=1}^{\infty} c_j$ be a convergent series in $\mathbb{R}_{>0}$. Then there exists an unconditionally Cauchy series $\sum_{j=1}^{\infty} v_j$ in V such that $\|v_j\|^2 = c_j$, $j \in \mathbb{Z}_{>0}$.*

Proof Define $n_0 = 0$ and define n_1 such that

$$\Big(\sum_{j=n_1+1}^{\infty} c_j\Big)^{1/2} < 1,$$

this being possible since $\sum_{j=1}^{\infty} c_j$ is a convergent series of positive terms. Then define $n_2 > n_1$ such that

$$\Big(\sum_{j=n_2+1}^{\infty} c_j\Big)^{1/2} < \frac{1}{4}.$$

Carrying on in this way we define an increasing sequence $(n_j)_{j\in\mathbb{Z}_{\geq 0}}$ such that

$$\Big(\sum_{j=n_k+1}^{n_{k+1}} c_j\Big)^{1/2} < \Big(\sum_{j=n_k+1}^{\infty} c_j\Big)^{1/2} < \frac{1}{k^2}, \qquad k \in \mathbb{Z}_{>0}.$$

The series

$$\sum_{k=0}^{\infty}\Big(\sum_{j=n_k+1}^{n_{k+1}} c_j\Big)^{1/2}$$

then converges by Example I-2.4.2–4. Take $k \in \mathbb{Z}_{\geq 0}$. By Lemma 2 let v_j, $j \in \{n_k + 1, \dots, n_{k+1}\}$, be such that $\|v_j\|^2 = c_j$ and such that

$$\Big\|\sum_{j\in J} v_j\Big\|^2 \leq 3\sum_{j\in J} c_j$$

for any $J \subseteq \{n_k + 1, \dots, n_{k+1}\}$. Let $\epsilon \in \mathbb{R}_{>0}$ and choose $N_1 \in \mathbb{Z}_{>0}$ such that

$$\sum_{k=N_1}^{\infty}\Big(\sum_{j=n_k+1}^{n_{k+1}} c_j\Big)^{1/2} < \frac{\epsilon}{3}.$$

Let $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a bijection and choose $N_2 \in \mathbb{Z}_{>0}$ such that

$$\{1, \dots, n_{N_1}\} \subseteq \{\phi(1), \dots, \phi(N_2)\}.$$

Thus

$$(v_{\phi(j)})_{j=N_2}^{\infty} \subseteq (v_j)_{j=N_1+1}^{\infty}.$$

Let $N_3 > N_2$ and let $k \geq N_1$. Denote by $J_k \subseteq \{n_k + 1, \dots, n_{k+1}\}$ the indices such that $j \in J_k$ if and only if $\phi(j) \in \{N_2, \dots, N_3\}$. Then we have

$$\Big\|\sum_{j=N_2}^{N_3} v_{\phi(j)}\Big\| = \Big\|\sum_{k=N_1}^{\infty}\sum_{j\in J_k} v_j\Big\| \leq \sum_{k=N_1}^{\infty}\Big\|\sum_{j\in J_k} v_j\Big\| \leq \sum_{k=N_1}^{\infty}\Big(3\sum_{j=n_k+1}^{n_k} c_j\Big)^{1/2} < \epsilon.$$

Thus the norm of the N_3rd partial sum minus the N_2nd partial sum for the series $\sum_{j=1}^{\infty} v_{\phi(j)}$ is less than ϵ. Thus this series is Cauchy and so $\sum_{j=1}^{\infty} v_j$ is unconditionally Cauchy. ▼

Now let us prove the theorem. Consider the sequence $\left(c_j = \frac{1}{j^2}\right)_{j\in\mathbb{Z}_{>0}}$ and by Lemma 3 let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence for which $\|v_j\|^2 = c_j$ and for which the series $\sum_{j=1}^{\infty} v_j$ is unconditionally Cauchy. But $\sum_{j=1}^{\infty}\|v_j\| = \sum_{j=1}^{\infty}\frac{1}{j}$ is divergent by Example I-2.4.2–2 and so $\sum_{j=1}^{\infty} v_j$ is not absolutely convergent. This proves the theorem for normed $\mathbb{R}$-vector spaces. For normed $\mathbb{C}$-vector spaces we note that these are also normed $\mathbb{R}$-vector spaces. Since none of the constructions in the proof alter when complex scalars are replaced with real scalars, the proof is also valid for normed $\mathbb{C}$-vector spaces. ■

3.4.3 Algebraic operations on series

Let us close by indicating that convergence of series respects the algebraic structure of vector spaces. We first give two definitions of products of series of scalars and vectors.

3.4.9 Definition (Scalar multiplication of series) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Let $S = \sum_{j=0}^{\infty} v_j$ be a series in V and let $s = \sum_{j=0}^{\infty} a_j$ be series in $\mathbb{R}$.

(i) The ***product*** of s and S is the double series $\sum_{j,k=0}^{\infty} a_j v_k$.

(ii) The ***Cauchy product*** of s and S is the series $\sum_{k=0}^{\infty}\left(\sum_{j=0}^{k} a_j v_{k-j}\right)$. •

Now we can state the interaction between convergence of series and the vector space operations.

3.4.10 Proposition (Algebraic operations on series) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Let $\mathrm{S} = \sum_{j=0}^{\infty} \mathrm{u_j}$ and $\mathrm{T} = \sum_{j=0}^{\infty} \mathrm{v_j}$ be series in V converging to U_0 and V_0, respectively, let $\mathrm{s} = \sum_{j=0}^{\infty} \mathrm{a_j}$ be a series in $\mathbb{F}$ converging to A_0, and let $\mathrm{a} \in \mathbb{F}$. Then the following statements hold:*

(i) the series $\sum_{j=0}^{\infty} \mathrm{av_j}$ converges to aV_0;

(ii) the series $\sum_{j=0}^{\infty}(\mathrm{u_j} + \mathrm{v_j})$ converges to $\mathrm{U}_0 + \mathrm{V}_0$;

(iii) if s and T are absolutely convergent, then the product of s and T is absolutely convergent and converges to $\mathrm{A}_0\mathrm{V}_0$;

(iv) if s and T are absolutely convergent, then the Cauchy product of s and T is absolutely convergent and converges to $\mathrm{A}_0\mathrm{V}_0$;

(v) if s or T are absolutely convergent, then the Cauchy product of s and T is convergent and converges to $\mathrm{A}_0\mathrm{V}_0$.

Proof (i) Since $\sum_{j=0}^{k} av_j = a\sum_{j=0}^{k} v_j$, this follows from part (i) of Proposition 3.2.6.

(ii) Since $\sum_{j=0}^{\infty}(u_j + v_j) = \sum_{j=0}^{k} u_j + \sum_{j=0}^{k} v_j$, this follows from part (ii) of Proposition 3.2.6.

(iii) and (iv) To prove these parts of the result, we first make a general argument. We note that $\mathbb{Z}_{\geq 0}\times\mathbb{Z}_{\geq 0}$ is a countable set (e.g., by Proposition I-1.7.16), and so there exists a bijection, in fact many bijections, $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{\geq 0}\times\mathbb{Z}_{\geq 0}$. For such a bijection ϕ, suppose that we are given a double sequence $(v_{jk})_{j,k\in\mathbb{Z}_{\geq 0}}$ and define a sequence $(v_j^{\phi})_{j\in\mathbb{Z}_{>0}}$ by $v_j^{\phi} =$

x_{kl} where $(k,l) = \phi(j)$. We then claim that, for any bijection $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$, the double series $A = \sum_{k,l=1}^{\infty} v_{kl}$ converges absolutely if and only if the series $A^{\phi} = \sum_{j=1}^{\infty} v_j^{\phi}$ converges absolutely.

Indeed, suppose that the double series $\|A\| = \sum_{k,l=1}^{\infty}\|v_{kl}\|$ converges to $\beta \in \mathbb{R}$. For $\epsilon > 0$ the set

$$\{(k,l) \in \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0} \mid \; |\|A\|_{kl} - \beta| \geq \epsilon\}$$

is then finite. Therefore, there exists $N \in \mathbb{Z}_{>0}$ such that, if $(k,l) = \phi(j)$ for $j \geq N$, then $|\|A\|_{kl} - \beta| < \epsilon$. It therefore follows that $|\|A^{\phi}\|_j - \beta| < \epsilon$ for $j \geq N$, where $\|A^{\phi}\|$ denotes the series $\sum_{j=1}^{\infty}|v_j^{\phi}|$. This shows that the series $\|A^{\phi}\|$ converges to β.

For the converse, suppose that the series $\|A^{\phi}\|$ converges to β. Then, for $\epsilon > 0$ the set

$$\{j \in \mathbb{Z}_{>0} \mid \; \|A^{\phi}|_j - \beta| \geq \epsilon\}$$

is finite. Therefore, there exists $N \in \mathbb{Z}_{>0}$ such that

$$\{(k,l) \in \mathbb{Z}_{\geq 0} \mid \; k,l \geq N\} \cap \{(k,l) \in \mathbb{Z}_{\geq 0} \mid \; \|A^{\phi}|_{\phi^{-1}(k,l)} - \beta| \geq \epsilon\} = \emptyset.$$

It then follows that for $k,l \geq N$ we have $\|A|_{kl} - \beta| < \epsilon$, showing that $|A|$ converges to β.

Thus we have shown that A is absolutely convergent if and only if A^{ϕ} is absolutely convergent for any bijection $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$. From Proposition 3.4.5 we know that the limit of an absolutely convergent series or double series is independent of the manner in which the terms in the series are arranged.

Consider now a term in the product of s and T. It is easy to see that this term appears exactly once in the Cauchy product of s and T. Conversely, each term in the Cauchy product appears exactly one in the product. Thus the product and Cauchy product are simply rearrangements of one another. Moreover, each term in the product and the Cauchy product appears exactly once in the expression

$$\Big(\sum_{j=0}^{N} a_j\Big)\Big(\sum_{k=0}^{N} v_k\Big)$$

as we allow N to go to ∞. That is to say,

$$\sum_{j,k=0}^{\infty} a_j v_k = \sum_{k=0}^{\infty}\Big(\sum_{j=k}^{k} a_j v_{k-j}\Big) = \lim_{N\to\infty}\Big(\sum_{j=0}^{N} a_j\Big)\Big(\sum_{k=0}^{N} v_k\Big).$$

However, this last limit is exactly $A_0 V_0$, using part (iii) of Proposition 3.2.6.

(v) Suppose that s converges absolutely. Let $(s_k)_{k\in\mathbb{Z}_{>0}}$, $(T_k)_{k\in\mathbb{Z}_{>0}}$, and $((sT)_k)_{k\in\mathbb{Z}_{>0}}$ be the sequences of partial sums for s, T, and the Cauchy product, respectively. Also define $\tau_k = T_k - V_0$, $k \in \mathbb{Z}_{\geq 0}$. Then

$$\begin{aligned}(sT)_k &= a_0 v_0 + (a_0 v_1 + a_1 v_0) + \cdots + (a_0 v_k + \cdots + a_k v_0)\\ &= a_0 T_k + a_1 T_{k-1} + \cdots + a_k T_0\\ &= a_0(V_0 + \tau_k) + a_1(V_0 + \tau_{k-1}) + \cdots + a_k(V_0 + \tau_0)\\ &= s_k V_0 + a_0\tau_k + a_1\tau_{k-1} + \cdots + a_k\tau_0.\end{aligned}$$

Since $\lim_{k\to\infty} s_k V_0 = A_0 V_0$ by Proposition I-2.4.30(i), this part of the result will follow if we can show that

$$\lim_{k\to\infty} (a_0\tau_k + a_1\tau_{k-1} + \cdots + a_k\tau_0) = 0. \tag{3.5}$$

Denote

$$\sigma = \sum_{j=0}^{\infty} |a_j|,$$

and for $\epsilon > 0$ choose $N_1 \in \mathbb{Z}_{>0}$ such that $\|\tau_j\| \leq \frac{\epsilon}{2\sigma}$ for $j \geq N_1$, this being possible since $(\tau_j)_{j\in\mathbb{Z}_{>0}}$ clearly converges to zero. Then, for $k \geq N_1$,

$$\begin{aligned}\|a_0\tau_k + a_1\tau_{k-1} + \cdots + a_k\tau_0\| &\leq \|a_0\tau_k + \cdots + a_{k-N_1-1}\tau_{N_1-1}\| + \|a_{k-N_1}\tau_{N_1} + \cdots + a_k\tau_0\| \\ &\leq \tfrac{\epsilon}{2} + \|a_{k-N_1}\tau_{N_1} + \cdots + a_k\tau_0\|.\end{aligned}$$

Since $\lim_{k\to\infty} a_k = 0$, choose $N_2 \in \mathbb{Z}_{>0}$ such that

$$\|a_{k-N_1}\tau_{N_1} + \cdots + a_k\tau_0\| < \tfrac{\epsilon}{2}$$

for $k \geq N_2$. Then

$$\begin{aligned}&\limsup_{k\to\infty} \|a_0\tau_k + a_1\tau_{k-1} + \cdots + a_k\tau_0\| \\ &\qquad = \lim_{k\to\infty} \sup\{\|a_0\tau_j + a_1\tau_{j-1} + \cdots + a_j\tau_0\| \mid j \geq k\} \\ &\qquad \leq \lim_{k\to\infty} \sup\{\tfrac{\epsilon}{2} + \|a_{k-N_1}\tau_{N_1} + \cdots + a_k\tau_0\| \mid j \geq k\} \\ &\qquad \leq \sup\{\tfrac{\epsilon}{2} + \|a_{k-N_1}\tau_{N_1} + \cdots + a_k\tau_0\| \mid j \geq N_2\} \leq \epsilon.\end{aligned}$$

Thus

$$\limsup_{k\to\infty} \|a_0\tau_k + a_1\tau_{k-1} + \cdots + a_k\tau_0\| \leq 0,$$

and since clearly

$$\liminf_{k\to\infty} \|a_0\tau_k + a_1\tau_{k-1} + \cdots + a_k\tau_0\| \geq 0,$$

we infer that (3.5) holds by Proposition I-2.3.17.

If T converges absolutely, the above argument can be modified by defining

$$\sigma = \sum_{j=0}^{\infty} \|v_j\|$$

and swapping the rôles of s and T in the remainder of the proof. ■

3.4.4 Multiple series

One also has the notion of double series in normed vector spaces.

3.4.11 Definition (Double series) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be a $\mathbb{F}$-vector space. A ***double series*** in V is a sum of the form $\sum_{j,k=1}^{\infty} v_{jk}$ where $(v_{jk})_{j,k\in\mathbb{Z}_{>0}}$ is a double sequence in V. •

We then have the following notions of convergence of double series.

3.4.12 Definition (Convergence and absolute convergence of double series) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a $\mathbb{F}$-vector (semi)normed space. Let $(v_{jk})_{j,k\in\mathbb{Z}_{>0}}$ be a double sequence in V and consider the double series

$$S = \sum_{j,k=1}^{\infty} v_{jk}.$$

The corresponding sequence of ***partial sums*** is the double sequence $(S_{jk})_{j,k\in\mathbb{Z}_{>0}}$ defined by

$$S_{jk} = \sum_{l=1}^{j} \sum_{m=1}^{k} v_{lm}.$$

Let $v_0 \in \mathsf{V}$. The double series:

(i) ***converges to*** $\mathbf{v_0}$, and we write $\sum_{j,k=1}^{\infty} v_{jk} = v_0$, if the double sequence of partial sums converges to v_0;

(ii) has v_0 as a ***limit*** if it converges to v_0;

(iii) is ***convergent*** if it converges to some member of V;

(iv) ***converges absolutely***, or is ***absolutely convergent***, if the series

$$\sum_{j,k=1}^{\infty} \|v_{jk}\|$$

converges;

(v) ***converges conditionally***, or is ***conditionally convergent***, if it is convergent, but not absolutely convergent;

(vi) ***diverges*** if it does not converge. •

3.4.5 Cesàro convergence of sequences and series

If a sequence diverges, all hope may not be lost. Indeed, it is possible that convergence may not actually be what one was interested in. This seems a somewhat absurd proposition at first glance, but it actually forms the first steps towards a powerful theory of Fourier series, as we shall see in Section IV-5.2.7. The point is that when one has a divergent sequence or series, one should not just throw in the towel. It is possible that by modifying one's notion of convergence, useful information can still be extracted.

The idea of Cesàro convergence is that one should average the sequence and see if the averaged sequence converges. The same idea can be applied to sums via their partial sums.

3.4.13 Definition (Cesàro[1] convergence) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space, and let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in V.

(i) The ***Cesàro means*** for the sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ is the sequence $(\bar{v}^1_k))_{k\in\mathbb{Z}_{>0}}$ where

$$\bar{v}^1_k = \frac{1}{k}\sum_{j=1}^{k} v_k.$$

(ii) The ***Cesàro means*** for the series $S = \sum_{j=1}^{\infty} v_j$ is the sequence $(\bar{S}^1_k)_{k\in\mathbb{Z}_{>0}}$ of Cesàro means for the sequence of partial sums. Thus

$$\bar{S}^1_k = \frac{1}{k}\sum_{j=1}^{k} S_j = \frac{1}{k}\sum_{j=1}^{k}\sum_{l=1}^{j} v_l.$$

(iii) The sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ is ***Cesàro convergent*** if the sequence $(\bar{v}^1_k)_{k\in\mathbb{Z}_{>0}}$ of Cesàro means converges.

(iv) The series $S = \sum_{j=1}^{\infty} v_j$ is ***Cesàro convergent*** or ***Cesàro summable*** if the sequence $(\bar{S}^1_k)_{k\in\mathbb{Z}_{>0}}$ of Cesàro means converges. •

The us give some examples to illustrate the concept.

3.4.14 Examples (Cesàro convergence)

1. The sequence $(x_j \triangleq (-1)^{j+1})_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}$ is oscillatory and so does not converge. However, the sequence is Cesàro convergent since the Cesàro means are given by

$$\bar{x}^1_j = \begin{cases} \frac{1}{j}, & j \text{ odd},\\ 0, & j \text{ even},\end{cases}$$

and so the sequence is Cesàro convergent.

2. Let us consider the sum $S = \sum_{j=1}^{\infty}(-1)^{j+1}$ in $\mathbb{R}$. The sequence of partial sums is $(S_k)_{k\in\mathbb{Z}_{>0}}$ with

$$S_k = \begin{cases} 1, & k \text{ odd},\\ 0, & k \text{ even}.\end{cases}$$

Thus this series is oscillatory. The Cesàro means for the series are $(\bar{S}^1_k)_{k\in\mathbb{Z}_{>0}}$ with

$$\bar{S}^1_k = \begin{cases} \frac{k+1}{2k}, & k \text{ odd},\\ \frac{1}{2}, & k \text{ even}.\end{cases}$$

Thus the series is Cesàro convergent and the Cesàro means converge to $\frac{1}{2}$. •

[1] Ernesto Cesàro (1859–1906) was an Italian mathematician who made contributions to analysis, number theory, and differential geometry.

The examples illustrate that when one has a divergent sequence or series, it is possible to have Cesàro convergence. This is a useful property that one would ask of a modified version of convergence. The other natural notion is that it should actually generalise the standard notion of convergence. Thus a convergent sequence should still converge with any modified version of convergence. Cesàro convergence possesses this property.

3.4.15 Theorem (Convergence implies Cesàro convergence) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \|\cdot\|)$ *be a normed* $\mathbb{F}$*-vector space. If a sequence* $(\mathrm{v}_j)_{j\in\mathbb{Z}_{>0}}$ *(resp. a series* $\sum_{j=1}^{\infty} \mathrm{v}_j$*) converges to* $\mathrm{v}_0 \in \mathsf{V}$ *then the sequence (resp. series) converges to* v_0 *in the sense of Cesàro convergence.*

Proof Since the statement for series follows, by definition, from the statement for sequences, we only show that a convergent sequence is Cesàro convergent with the same limit.

Define $\bar{v}_k^1 = \frac{1}{k}(v_1 + \cdots + v_k)$. Let $\epsilon \in \mathbb{R}_{>0}$ and take $N_1 \in \mathbb{Z}_{>0}$ such that $\|v_j - v_0\| < \frac{\epsilon}{2}$ for $j \geq N_1$. Also take $N_2 \in \mathbb{Z}_{>0}$ sufficiently large that

$$\tfrac{1}{N_2}(\|v_1\| + \cdots + \|v_{N_1}\| + N_1\|v_0\|) < \tfrac{\epsilon}{2}.$$

Then, for $j \geq \{N_1, N_2\}$, we have

$$\begin{aligned}
\|\bar{v}_k^1 - v_0\| &= \left\|\tfrac{1}{k}(v_1 + \cdots + v_k) - v_0\right\| = \tfrac{1}{k}\|(v_1 - v_0) + \cdots + (v_k - v_0)\| \\
&\leq \tfrac{1}{k}\|(v_1 - v_0) + \cdots + (v_{N_1} - v_0)\| + \tfrac{1}{k}\|(v_{N_1+1} - v_0) + \cdots + (v_k - v_0)\| \\
&\leq \tfrac{1}{k}(\|v_1\| + \cdots + \|v_{N_1}\| + N_1\|v_0\|) + \tfrac{1}{k}(\|v_{N_1+1} - v_0\| + \cdots + \|v_k - v_0\|) \\
&\leq \frac{\epsilon}{2} + \frac{k - N_1}{k}\frac{\epsilon}{2} < \frac{\epsilon}{2},
\end{aligned}$$

mean in convex hull interpretation

giving the result. ■

Note that the Cesàro means for a sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ form a sequence $(\bar{v}_j^1)_{j\in\mathbb{Z}_{>0}}$. If this sequence diverges one can ask whether *its* sequence of Cesàro means converges. That is, we can define

$$\bar{v}_k^2 = \frac{1}{k}\sum_{j=1}^{k} \bar{v}_j^1 = \frac{1}{k}\sum_{j=1}^{k} \frac{1}{j}\sum_{l=1}^{j} v_j,$$

and consider the convergence of the sequence $(\bar{v}_k^2)_{k\in\mathbb{Z}_{>0}}$. This can clearly be iterated any finite number of times. This is interesting, although we shall not consider it here. We refer to the notes in Section 3.4.7 for references.

3.4.6 Series in normed vector spaces with arbitrary index sets

In Section I-2.4.7 we presented the notion of a series in $\mathbb{R}$ with an arbitrary index set. Such series were useful in discussion saltus functions. Here we discuss series in normed vector spaces with arbitrary index sets. This will be helpful for us in Section 4.4 when we discuss Hilbert bases in general inner product spaces. In any case, much of the treatment mirrors to some extent that for arbitrary series in $\mathbb{R}$.

Let us begin with the definition.

3.4.16 Definition (Convergence of series with arbitrary index sets) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Let A be an index set, consider a family $(v_a)_{a\in A}$ in V, and denote $S = \sum_{a\in A} v_a$. Let $v_0 \in \mathsf{V}$.

(i) The series S ***converges*** to v_0 if, for any $\epsilon \in \mathbb{R}_{>0}$, there exists a finite set $I \subseteq A$ such that

$$\Big\|\sum_{a\in J} v_a - v_0\Big\| < \epsilon$$

for every finite subset $J \subseteq A$ for which $I \subseteq J$.

(ii) The series S is ***Cauchy*** if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists a finite set $I \subseteq A$ such that

$$\Big\|\sum_{a\in J} v_a\Big\| < \epsilon$$

for every finite subset $J \subseteq A$ for which $J \cap I = \varnothing$. •

We already have one point of difference with the results in Section I-2.4.7 in that here we have the notion of Cauchy series. This is because we need to allow for the possibility of sums that seem like they should converge, but do not. The next result is analogous to the fact that convergent sequences are always Cauchy, but Cauchy sequences need not converge, but only generally converge when the normed vector space is complete.

3.4.17 Theorem (Relationship between convergent series and Cauchy series) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. For a series $\mathrm{S} = \sum_{\mathrm{a}\in\mathrm{A}} \mathrm{v_a}$ the following statements hold:*

(i) if S is convergent then it is Cauchy;

(ii) if V is complete and if S is Cauchy then it is convergent.

Proof (i) Let $\epsilon \in \mathbb{R}_{>0}$ and let $I \subseteq A$ be a finite subset such that

$$\Big\|\sum_{a\in J} v_a - v_0\Big\| < \frac{\epsilon}{2}$$

for every finite subset J for which $I \subseteq J$. Let $K \subseteq A$ be finite and such that $K \cap I = \varnothing$. Then

$$\begin{aligned}\Big\|\sum_{a\in K} v_a\Big\| &= \Big\|\sum_{a\in K} v_a + \Big(\sum_{a\in I} v_a - v_0\Big) - \Big(\sum_{a\in I} v_a - v_0\Big)\Big\| \\ &\le \Big\|\sum_{a\in K\cup I} v_a - v_0\Big\| + \Big\|\sum_{a\in I} v_a - v_0\Big\| \\ &\le \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon,\end{aligned}$$

as desired.

(ii) Let $k \in \mathbb{Z}_{>0}$ and let $I_k \subseteq A$ be a finite subset such that

$$\Big\|\sum_{a\in J} v_a\Big\| < \frac{1}{k}$$

for every finite subset J for which $J \cap I_k = \emptyset$. Then define

$$u_k = \sum_{a \in I_k} v_a.$$

We claim that the sequence $(u_k)_{k \in \mathbb{Z}_{>0}}$ is Cauchy. Indeed, let $N \in \mathbb{Z}_{>0}$ be such that $\frac{1}{N} < \frac{\epsilon}{2}$. Then, for $j, k \geq N$, we have

$$\begin{aligned}\|u_j - u_k\| &= \Big\|\sum_{a \in I_j} v_a - \sum_{a \in I_k} v_a\Big\| = \Big\|\sum_{a \in I_j - I_k} v_a - \sum_{a \in I_k - I_j} v_a\Big\| \\ &\leq \Big\|\sum_{a \in I_j - I_k} v_a\Big\| + \Big\|\sum_{a \in I_k - I_j} v_a\Big\| = \tfrac{1}{j} + \tfrac{1}{k} < \epsilon,\end{aligned}$$

giving $(u_k)_{k \in \mathbb{Z}_{>0}}$ as a Cauchy sequence. Since V is complete there exists a limit u_0 of $(u_k)_{k \in \mathbb{Z}_{>0}}$. Thus, for $\epsilon \in \mathbb{R}_{>0}$, there exists $N_1 \in \mathbb{Z}_{>0}$ such that $\|u_j - u_0\| < \frac{\epsilon}{2}$ for $j \geq N_1$. If $N_2 = \max\{N_1, \frac{2}{\epsilon}\}$ then

$$\begin{aligned}\Big\|\sum_{a \in J} v_a - u_0\Big\| &= \Big\|\sum_{a \in I_{N_2}} v_a - u_0 + \sum_{a \in J \setminus I_{N_2}} v_a\Big\| \\ &\leq \Big\|\sum_{a \in I_{N_2}} v_a - u_0\Big\| + \Big\|\sum_{a \in J \setminus I_{N_2}} v_a\Big\| \leq \tfrac{\epsilon}{2} + \frac{1}{N_2} < \epsilon,\end{aligned}$$

where J is any finite set for which $I_{N_2} \subseteq J$. Thus S converges to u_0. ■

The theorem illustrates the difference between a convergent series and a Cauchy series. The most important fact is that the two notions are equivalent when V is a Banach space.

Just as with arbitrary sums of real numbers, any convergent arbitrary sum in normed vector space can have only countably many nonzero elements.

3.4.18 Proposition (There are only countably many nonzero terms in a convergent series) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \|\cdot\|)$ *be a normed* $\mathbb{F}$*-vector space. If* $S = \sum_{a \in A} v_a$ *is a convergent series then the set* $\{a \in A \mid v_a \neq 0_{\mathsf{V}}\}$ *is countable.*

Proof By Theorem 3.4.17, since S converges, for any $k \in \mathbb{Z}_{>0}$ there exists a finite set $I_k \subseteq A$ such that

$$\Big\|\sum_{a \in J} v_a\Big\| < \frac{1}{k}$$

for any finite set J such that $J \cap I_k = \emptyset$. Let $I = \cup_{k \in \mathbb{Z}_{>0}} I_k$ so that I is countable by Proposition I-1.7.16. If $a \notin I$ then $a \notin I_k$ for all $k \in \mathbb{Z}_{>0}$, i.e., $\{a\} \cap I_k = \emptyset$ for all $k \in \mathbb{Z}_{>0}$. Therefore, $\|v_a\| < \frac{1}{k}$ for all $k \in \mathbb{Z}_{>0}$ and so $\|v_a\| = 0$. Thus $v_a = 0_{\mathsf{V}}$ for all $a \notin I$. ■

Note that Definition 3.4.16 is not the generalisation of Definition I-2.4.31, or at least not obviously. Let us prove that the two definitions are, in fact, consistent.

3.4.19 Proposition (Consistency of two notions of arbitrary sums) *Let* A *be an index set and let* $S = \sum_{a \in A} x_a$ *be a series in* $\mathbb{R}$*. This series converges according to Definition* I-2.4.31 *if and only if it converges according to Definition* 3.4.16*, and in case the series converge, they converge to the same limit.*

Proof It suffices to consider the case when the numbers x_a, $a \in A$, are nonnegative (why?). First suppose that S converges according to Definition I-2.4.31. Thus

$$\sup\Big\{\sum_{a \in I} x_a \;\Big|\; I \subseteq A \text{ is finite}\Big\} = L < \infty.$$

Let $\epsilon \in \mathbb{R}_{>0}$ and let $I \subseteq A$ be a finite set such that

$$L - \epsilon \leq \sum_{a \in I} x_a \leq L.$$

Therefore, for any finite set $J \subseteq A$ for which $I \subseteq J$ it holds that

$$L - \epsilon \leq \sum_{a \in I} x_a \leq \sum_{a \in J} x_a \leq L$$

since the elements in the family $(x_a)_{a \in A}$ are nonnegative. This implies that

$$\Big\|\sum_{a \in J} x_a - L\Big\| < \epsilon$$

for any finite set J for which $I \subseteq J$, giving convergence of S to R in the sense of Definition 3.4.16.

The argument above can be essentially reversed to show that if S converges to L in the sense of Definition 3.4.16 then it converges to L in the sense of Definition I-2.4.31. ■

For arbitrary series in $\mathbb{R}$ we saw that convergence amounted to absolute convergence in the case when the index set was $\mathbb{Z}_{>0}$. The same is true for arbitrary series in formed vector spaces. For the following result, recall from Proposition 3.4.4 that limits of unconditionally convergent series are independent of rearrangement.

3.4.20 Theorem (A convergent series with index set $\mathbb{Z}_{>0}$ is unconditionally convergent) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \|\cdot\|)$ *be a (semi)normed* $\mathbb{F}$*-vector space. For a sequence* $(v_j)_{j \in \mathbb{Z}_{>0}}$ *the statements are equivalent:*

(i) the series $\sum_{j \in \mathbb{Z}_{>0}} v_j$ *is Cauchy in the sense of Definition* 3.4.16*;*

(ii) the series $\sum_{j=1}^{\infty} v_j$ *is unconditionally Cauchy.*

Moreover, for $v_0 \in \mathsf{V}$*, the following statements are also equivalent:*

(iii) the series $\sum_{j \in \mathbb{Z}_{>0}} v_j$ *converges to* v_0*;*

(iv) the series $\sum_{j=1}^{\infty}$ *converges unconditionally to* v_0*.*

Proof (i) $\Longrightarrow$ (ii) Let $\epsilon \in \mathbb{R}_{>0}$ and let $I \subseteq \mathbb{Z}_{>0}$ be a finite subset such that

$$\Big\|\sum_{j \in J} v_j\Big\| < \epsilon$$

for any finite set $J \subseteq \mathbb{Z}_{>0}$ for which $J \cap I = \emptyset$. Let $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a bijection and choose $N \in \mathbb{Z}_{>0}$ sufficiently large that $I \subseteq \{\phi(1), \dots, \phi(N)\}$. Then, for $k, l \geq N$ with $l > k$ the set $\{\phi(k+1), \dots, \phi(l)\}$ does not intersect I. Thus

$$\Big\| \sum_{j=k+1}^{l} v_{\phi(j)} \Big\| < \epsilon,$$

showing that the lth partial sum minus the kth partial sum is bounded above in norm by ϵ for any $k, l \geq N$. Thus $\sum_{j=1}^{\infty} v_{\phi(j)}$ is Cauchy.

(ii) $\Longrightarrow$ (i) Suppose that (ii) does not hold. Then there exists $\epsilon \in \mathbb{R}_{>0}$ such that, for any finite set $I \subseteq \mathbb{Z}_{>0}$, there exists a finite set $J \subseteq \mathbb{Z}_{>0}$ with $J \cap I = \emptyset$ and such that

$$\Big\| \sum_{j \in J} v_j \Big\| > \epsilon.$$

Now let $I_1 \subseteq \mathbb{Z}_{>0}$ be finite and let $J_1 \subseteq \mathbb{Z}_{>0}$ be finite with $J_1 \cap I_1 = \emptyset$ and with

$$\Big\| \sum_{j \in J_1} v_j \Big\| > \epsilon.$$

Note that $I_2 = I_1 \cup J_1$ is finite. Thus there exists a finite set $J_2 \subseteq \mathbb{Z}_{>0}$ such that $J_2 \cap I_2 = \emptyset$ and such that

$$\Big\| \sum_{j \in J_2} v_j \Big\| > \epsilon.$$

We can continue in this way to define a sequence $(J_k)_{k \in \mathbb{Z}_{>0}}$ of finite pairwise disjoint subsets of $\mathbb{Z}_{>0}$ with the property that

$$\Big\| \sum_{j \in J_k} v_j \Big\| > \epsilon, \qquad k \in \mathbb{Z}_{>0}.$$

Let us denote $\min J_k = m_k$ and $\max J_k = M_k$. Also denote $J_k = \{j_{k,1}, \dots, j_{k,r_k}\}$. Now let $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a bijection such that

$$\phi(\{m_k, \dots, M_k\}) \subseteq \{m_k, \dots, M_k\}$$

and such that

$$\phi(m_k) = j_{k,1}, \dots, \phi(m_k + r_k - 1) = j_{k,r_k}$$

for each $k \in \mathbb{Z}_{>0}$. Then, for any $k \in \mathbb{Z}_{>0}$ we have

$$\Big\| \sum_{j=m_k}^{m_k+r_k-1} v_{\phi(j)} \Big\| = \Big\| \sum_{j \in J_k} v_j \Big\| > \epsilon.$$

Therefore, no matter how large we choose $N \in \mathbb{Z}_{>0}$, there exists $k, l \geq N$ such that the lth partial sum minus the kth partial sum for the series $\sum_{j=1}^{\infty} v_{\phi(j)}$ is bounded below in norm by ϵ. Thus the series is not Cauchy.

(iii) $\implies$ (iv) Suppose that $\sum_{j\in\mathbb{Z}_{>0}} v_j$ converges to v_0 in the sense of Definition 3.4.16 to v_0. Let $\epsilon \in \mathbb{R}_{>0}$ and let $I \subseteq \mathbb{Z}_{>0}$ be a finite set such that

$$\Big\|\sum_{j\in J} v_j - v_0\Big\| < \epsilon$$

for any finite subset $J \subseteq \mathbb{Z}_{>0}$ for which $I \subseteq J$. Let $\phi\colon \mathbb{Z}_{>0} \to \mathbb{Z}_{>0}$ be a bijection. Choose $N \in \mathbb{Z}_{>0}$ sufficiently large that $I \subseteq \{\phi(1), \dots, \phi(N)\}$ and note that, for $k \geq N$ we have

$$\Big\|\sum_{j=1}^{k} v_{\phi(j)} - v_0\Big\| < \epsilon$$

since $S \subseteq \{\phi(1), \dots, \phi(k)\}$. Thus $\sum_{j=1}^{\infty} v_{\phi(j)}$ converges to v_0.

(iv) $\implies$ (iii) Now suppose that $\sum_{j=1}^{\infty} v_j$ converges unconditionally to v_0. Then $\sum_{j=1}^{\infty} v_j$ is unconditionally Cauchy and so Cauchy in the sense of Definition 3.4.16 by the implication (ii) $\implies$ (i). Let $\epsilon \in \mathbb{R}_{>0}$ and let $I' \subseteq \mathbb{Z}_{>0}$ be a finite subset such that

$$\Big\|\sum_{j\in J'} v_j\Big\| < \frac{\epsilon}{2}$$

for every finite subset $J' \subseteq \mathbb{Z}_{>0}$ for which $J' \cap I' = \varnothing$. Let $N \in \mathbb{Z}_{>0}$ be such that

$$\Big\|\sum_{j=1}^{k} v_j - v_0\Big\| < \frac{\epsilon}{2}$$

for every $k \geq N$ and such that $I' \subseteq N$. Define $I = \{1, \dots, N\}$ and let $J \subseteq \mathbb{Z}_{>0}$ be a finite set such that $I \subseteq J$. Write $J = I \cup J'$ with $J' \cap I = \varnothing$. Note that $J' \cap I' = \varnothing$. Therefore,

$$\Big\|\sum_{j\in J} v_j - v_0\Big\| = \Big\|\sum_{j=1}^{N} v_j - v_0 + \sum_{j\in J'} v_j\Big\| \leq \Big\|\sum_{j=1}^{N} v_j - v_0\Big\| + \Big\|\sum_{j\in J'} v_j\Big\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

Thus $\sum_{j\in\mathbb{Z}_{>0}} v_j$ converges to v_0 in the sense of Definition 3.4.16. ■

3.4.7 Notes

We saw in Section 3.4.5 that revised notions of convergence can be applied to divergent series. The classic book of Hardy [1949] discusses divergent series in detail.

Theorem 3.4.8 was first proved by Dvoretzky and Rogers [1950], and the proof we give follows the original proof in form.

Exercises

3.4.1 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Show that

$$\Big\|\sum_{j=1}^{m} v_j\Big\| \leq \sum_{j=1}^{m} \|v_j\|$$

for any finite family $(v_1, \dots, v_m)$ in V.

3.4.2 In Definition 3.4.1 we defined the notions of "convergent series," "Cauchy series," "unconditionally convergent series," and "unconditionally Cauchy series." We also defined the notion of "absolutely convergent series." Why did we not define the notion of "absolutely Cauchy series"?

Section 3.5

Continuous maps between normed vector spaces

As with so many areas of mathematics, for normed vector spaces it is interesting to study maps that preserve the structure, in this case the structure defined by the norm. Normed vector spaces have two facets to their structure: (1) the vector space structure and (2) the topology defined by the norm. Thus the interesting maps to consider are linear *and* continuous. We studied linear maps from an algebraic point of view in Sections I-5.1 and I-5.4, with particular emphasis on the finite-dimensional setting in Section I-5.8. Maps between topological spaces were the subject of Section 1.3. As we shall see, in combining these points of view, one ends up with some quite rich structure.

Do I need to read this section? Continuous linear maps are extremely important in applications. Indeed, the Fourier and Laplace transforms studied in Volume 3 are important examples of continuous linear maps. Therefore, the basic material in this section is important to understand. Some of the more detailed material, for example that in , can be skimmed at a first reading, and referred to as needed. • what?

3.5.1 General continuous maps between normed vector spaces

Most often we will be interested in continuous *linear* maps between normed vector spaces. However, there are also times when it will be helpful to have on hand the notion of continuity for general maps. Thus we present this first.

3.5.1 Definition (Continuous maps between normed vector spaces) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces. For open sets $S \subseteq \mathsf{U}$ and $T \subseteq \mathsf{V}$ and for $u_0 \in S$, a map $f\colon S \to T$ is:

(i) ***continuous at*** $\mathbf{u_0}$ if, for each $\epsilon \in \mathbb{R}_{>0}$ there exists $\delta \in \mathbb{R}_{>0}$ such that $\|f(u) - f(u_0)\|_{\mathsf{V}} < \epsilon$ whenever $u \in S$ satisfies $\|u - u_0\|_{\mathsf{U}} < \delta$;

(ii) ***continuous*** if it is continuous at each $u_0 \in S$;

(iii) ***uniformly continuous*** if, for each $\epsilon \in \mathbb{R}_{>0}$ there exists $\delta \in \mathbb{R}_{>0}$ such that $\|f(u_1) - f(u_2)\| < \epsilon$ for all $u_1, u_2 \in S$ satisfying $\|u_1 - u_2\| < \delta$;

(iv) ***discontinuous at*** $\mathbf{u_0}$ if it is not continuous at u_0;

(v) ***discontinuous*** if it is not continuous. •

We will give interesting examples of continuous *linear* maps in Example 3.5.10. Here let us record some alternative characterisations of continuity.

3.5.2 Theorem (Alternative characterisations of continuity) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ *and* $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ *be normed* $\mathbb{F}$*-vector spaces. For a map* $\mathrm{f}\colon \mathrm{S} \to \mathsf{V}$ *defined on an open subset* $\mathrm{S} \subseteq \mathsf{U}$ *and for* $\mathrm{u}_0 \in \mathrm{S}$*, the following statements are equivalent:*

(i) f *is continuous at* u_0*;*

(ii) for every neighbourhood B *of* $\mathrm{f}(\mathrm{u}_0)$ *there exists a neighbourhood* A *of* u_0 *in* S *such that* $\mathrm{f}(\mathrm{A}) \subseteq \mathrm{B}$*;*

(iii) $\lim_{\mathrm{u}\to\mathrm{u}_0} \mathrm{f}(\mathrm{u}) = \mathrm{f}(\mathrm{u}_0)$.

Proof In the proof we denote open balls in U and V by $\mathsf{B}_{\mathsf{U}}(r, u)$ and $\mathsf{B}_{\mathsf{V}}(r, v)$, respectively.

(i) $\implies$ (ii) Let $B \subseteq \mathsf{V}$ be a neighbourhood of $f(u_0)$. Let $\epsilon \in \mathbb{R}_{>0}$ be defined such that $\mathsf{B}_{\mathsf{V}}(\epsilon, f(u_0)) \subseteq B$, this being possible since B is open. Since f is continuous at u_0, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $u \in \mathsf{B}_{\mathsf{U}}(\delta, u_0) \cap S$, then we have $f(u) \in \mathsf{B}(\epsilon, f(u_0))$. This shows that, around the point u_0, we can find an open set A in S whose image lies in B.

(ii) $\implies$ (iii) Let $(u_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in S converging to u_0 and let $\epsilon \in \mathbb{R}_{>0}$. By hypothesis there exists a neighbourhood A of u_0 in S such that $f(A) \subseteq \mathsf{B}_{\mathsf{V}}(\epsilon, f(u_0))$. Thus there exists $\delta \in \mathbb{R}_{>0}$ such that $f(\mathsf{B}_{\mathsf{U}}(\delta, u_0) \cap S) \subseteq \mathsf{B}_{\mathsf{V}}(\epsilon, f(u_0))$ since A is open in S. Now choose $N \in \mathbb{Z}_{>0}$ sufficiently large that $|u_j - u_0| < \delta$ for $j \geq N$. It then follows that $|f(u_j) - f(u_0)| < \epsilon$ for $j \geq N$, so giving convergence of $(f(u_j))_{j\in\mathbb{Z}_{>0}}$ to $f(u_0)$, as desired, keeping in mind Notation 3.2.2.

(iii) $\implies$ (i) Let $\epsilon \in \mathbb{R}_{>0}$. Then, by definition of $\lim_{u\to u_0} f(u) = f(u_0)$ from Notation 3.2.2, there exists $\delta \in \mathbb{R}_{>0}$ such that, for $u \in \mathsf{B}_{\mathsf{U}}(\delta, u_0) \cap S$, $|f(u) - f(u_0)| < \epsilon$, which is exactly the definition of continuity of f at u_0. ■

As we have seen, different norms can really be different (i.e., not equivalent), and so, in particular, maps continuous in one norm may not be continuous in another. Moreover, even in finite-dimensions where all norms are equivalent, it is sometimes convenient to use one norm or another, and in this case one would like to ensure that one's conclusions concerning continuity are not dependent on norm. In some sense this is trivial, since equivalent norms define the same topology (Theorem 3.1.14), and it is the topology that determines continuity. However, it is instructive to verify independence of continuity on a choice of equivalent norm. Thus we state the result here, and leave the proof to the reader as Exercise 3.5.3. The result assumes the fact that open sets are the same for equivalent norms; this is exactly what Theorem 3.1.14 shows.

3.5.3 Proposition (Continuity is independent of equivalent norm) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* U *and* V *be* $\mathbb{F}$*-vector spaces, let* $\|\cdot\|_{1,\mathsf{U}}$ *and* $\|\cdot\|_{2,\mathsf{U}}$ *be equivalent norms on* U*, let* $\|\cdot\|_{1,\mathsf{V}}$ *and* $\|\cdot\|_{2,\mathsf{V}}$ *be equivalent norms on* V*, and let* $\mathrm{S} \subseteq \mathsf{U}$ *and* $\mathrm{T} \subseteq \mathsf{V}$ *be open sets. Then, for a map* $\mathrm{f}\colon \mathrm{S} \to \mathrm{T}$*, the following statements are equivalent:*

(i) f *is continuous relative to the norms* $\|\cdot\|_{1,\mathsf{U}}$ *on* U *and* $\|\cdot\|_{1,\mathsf{V}}$ *on* V*;*

(ii) f *is continuous relative to the norms* $\|\cdot\|_{1,\mathsf{U}}$ *on* U *and* $\|\cdot\|_{2,\mathsf{V}}$ *on* V*;*

(iii) f *is continuous relative to the norms* $\|\cdot\|_{2,\mathsf{U}}$ *on* U *and* $\|\cdot\|_{1,\mathsf{V}}$ *on* V*;*

(iv) f *is continuous relative to the norms* $\|\cdot\|_{2,\mathsf{U}}$ *on* U *and* $\|\cdot\|_{2,\mathsf{V}}$ *on* V*.*

With the definition of continuity, let us prove the continuity of some of the standard vector space operations relative to the norm.

3.5.4 Proposition (Continuity properties of operations on normed vector spaces) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. Then the following maps are continuous:*

(i) $\mathsf{V} \ni \mathrm{v} \mapsto \mathrm{v} + \mathrm{v}_0 \in \mathsf{V}$ *for* $\mathrm{v}_0 \in \mathsf{V}$*;*

(ii) $\mathsf{V} \oplus \mathsf{V} \ni (\mathrm{v}_1, \mathrm{v}_2) \mapsto \mathrm{v}_1 + \mathrm{v}_2 \in \mathsf{V}$*;*

(iii) $\mathsf{V} \ni \mathrm{v} \mapsto \mathrm{a}\mathrm{v} \in \mathsf{V}$ *for* $\mathrm{a} \in \mathbb{F}$*;*

(iv) $\mathbb{F} \oplus \mathsf{V} \ni (\mathrm{a}, \mathrm{v}) \mapsto \mathrm{a}\mathrm{v} \in \mathsf{V}$*;*

(v) $\mathsf{V} \ni \mathrm{v} \mapsto \|\mathrm{v}\| \in \mathbb{R}$.

Moreover, the maps in parts (i), (ii), (iii), and (v) are uniformly continuous.

Proof (i) For $\epsilon \in \mathbb{R}_{>0}$ let $\delta = \epsilon$. Let $v, v' \in \mathsf{V}$ satisfy $\|v' - v\| < \delta$. We then have

$$\|(v' + v_0) - (v - v_0)\| = \|v' - v\| < \delta = \epsilon,$$

giving uniform continuity of the stated map.

(ii) Let $\epsilon \in \mathbb{R}_{>0}$ and let $\delta = \epsilon$. Let $(u_1, u_2), (v_1, v_2) \in \mathsf{V} \oplus \mathsf{V}$ satisfy $\|(v_1, v_2) - (u_1, u_2)\| < \delta$, where, by abuse of notation, $\|\cdot\|$ denotes the norm on $\mathsf{V} \oplus \mathsf{V}$. Then we have

$$\|v_1 - v_2 - (u_1 + u_2)\| \le \|v_1 - u_1\| + \|v_2 - u_2\| = \|(v_1, v_2) - (u_1, u_2)\| < \epsilon,$$

giving uniform continuity of the stated map.

(iii) If $a = 0$ then the map is constant, and so certainly uniformly continuous. If $a \neq 0$, let $\epsilon \in \mathbb{R}_{>0}$ and define $\delta = \frac{\epsilon}{|a|}$. Then, if $\|v - v'\| < \delta$ we have

$$\|av - av'\| = |a|\|v - v'\| < \epsilon,$$

giving uniform continuity as desired.

(iv) Let $\epsilon \in \mathbb{R}_{>0}$ and let $(a_0, v_0) \in \mathbb{F} \oplus \mathsf{V}$. Define

$$\delta = \min\Big\{1, \frac{\epsilon}{2(|a_0| + 1)}, \frac{\epsilon}{2(\|v_0\| + 1)}\Big\}$$

and note that if $\|(a, v) - (a_0, v_0)\| < \delta$ (again we abuse notation and denote by $\|\cdot\|$ the norm on $\mathbb{F} \oplus \mathsf{V}$) then we have

$$|a - a_0| + \|v - v_0\| < \delta$$

which in turn implies that

$$\begin{aligned} |a - a_0| < 1 \quad &\Longrightarrow \quad |a| < |a_0| + 1, \\ |a - a_0| &< \frac{\epsilon}{2(|a_0| + 1)}, \\ \|v - v_0\| &< \frac{\epsilon}{2(\|v_0\| + 1)}. \end{aligned}$$

We then compute, for $\|(a, v) - (a_0, v_0)\| < \delta$,

$$\begin{aligned} \|av - a_0 v_0\| &= \|av - av_0 + av_0 - a_0 v_0\| = \|a(v - v_0) + (a - a_0)v_0\| \\ &\le |a|\|v - v_0\| + |a - a_0|\|v_0\| \\ &\le (|a_0| + 1)\frac{\epsilon}{2(|a_0| + 1)} + \frac{\epsilon}{2(\|v_0\| + 1)}(\|v_0\| + 1) = \epsilon. \end{aligned}$$

(v) For $\epsilon \in \mathbb{R}_{>0}$ define $\delta = \epsilon$. Then, if $v, v' \in \mathsf{V}$ satisfy $\|v - v'\| < \delta$, we have

$$|\|v\| - \|v'\|| \le \|v - v'\| < \delta = \epsilon,$$

giving uniform continuity of the norm. ■

Particularly interesting are continuous bijections with continuous inverses.

3.5.5 Definition (Homeomorphism) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces, and let $S \subseteq \mathsf{U}$ and $T \subseteq \mathsf{V}$ be open sets. A map $f: S \to T$ is a ***homeomorphism*** if f is a continuous bijection with a continuous inverse. •

Let us give some examples of homeomorphisms.

3.5.6 Examples (Homeomorphism)

1. The map $f: (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ defined by $f(x) = \tan(x)$ is a homeomorphism with inverse $f^{-1} = \arctan$.
2. Let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space and let $v_0 \in \mathsf{V}$. The map $v \mapsto v + v_0$ is a homeomorphism of V with itself, and has inverse $v \mapsto v - v_0$.
3. Let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space and let $a \in \mathbb{F} \setminus \{0\}$. The map $v \mapsto av$ is a homeomorphism of V with itself, and has inverse $v \mapsto a^{-1}v$. •

3.5.2 Continuous linear maps between normed vector spaces

For vector spaces the maps that preserve the structure are linear maps. For topological spaces the maps that preserve the structure are continuous maps. Thus is makes sense that for normed vector spaces, as they have both the structure of a vector space and a topological space, the most informative maps to consider are those that are linear and continuous. These have a surprisingly rich structure. In this section we give some of their more elementary properties.

Let us first give the notation we will use for continuous linear maps, along with some other useful concepts that can be attached to a linear map.

3.5.7 Definition (Continuous linear maps between normed vector spaces) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}; \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces. The set of continuous linear maps from U to V is denoted by $\mathrm{L}(\mathsf{U}; \mathsf{V})$. A linear map $\mathsf{L} \in \mathrm{Hom}_{\mathbb{F}}(\mathsf{U}; \mathsf{V})$ is:

(i) ***bounded*** if there exists $M \in \mathbb{R}_{>0}$ such that $\|\mathsf{L}(u)\|_{\mathsf{V}} \le M\|u\|_{\mathsf{U}}$ for every $u \in \mathsf{U}$;

(ii) ***unbounded*** if it is not bounded;

(iii) ***norm-preserving*** if $\|\mathsf{L}(u)\|_{\mathsf{V}} = \|u\|_{\mathsf{U}}$ for all $u \in \mathsf{U}$;

(iv) an ***isomorphism of normed vector spaces*** if it is an isomorphism of vector spaces and is norm-preserving. •

Note that a homeomorphism of normed vector spaces is not necessarily an isomorphism of normed vector spaces, as can be seen in Exercise 3.5.4.

The following result gives a collection of useful conditions that are equivalent to continuity.

3.5.8 Theorem (Characterisations of continuous linear maps) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}; \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces. For $\mathsf{L} \in \mathrm{Hom}_{\mathbb{F}}(\mathsf{U}; \mathsf{V})$ the following conditions are equivalent:*

(i) L *is continuous;*

(ii) L *is continuous at* 0_{U}*;*

(iii) L *is uniformly continuous;*

(iv) L *is bounded.*

Moreover, any of the preceding four conditions implies the following:

(v) $\ker(\mathsf{L})$ *is a closed subspace of* U*.*

Proof (i) $\Longrightarrow$ (ii) This is clear.

(ii) $\Longrightarrow$ (iii) Let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta \in \mathbb{R}_{>0}$ such that $\|\mathsf{L}(u)\|_{\mathsf{V}} < \epsilon$ if $\|u\|_{\mathsf{U}} < \delta$; this is possible since L is linear at 0_{U}. Now let $u_0 \in \mathsf{U}$ and suppose that $\|u - u_0\|_{\mathsf{U}} < \delta$. Then

$$\|\mathsf{L}(u) - \mathsf{L}(u_0)\| = \|\mathsf{L}(u - u_0)\| < \epsilon,$$

which gives uniform continuity, as desired.

(iii) $\Longrightarrow$ (iv) Since L is uniformly continuous, it is continuous at 0_{U}. Let $M \in \mathbb{R}_{>0}$ be such that if $\|u\|_{\mathsf{U}} < \frac{2}{M}$ then $\|\mathsf{L}(u)\|_{\mathsf{V}} < 1$. Let $u \in \mathsf{U}$ and note that

$$\Big\|\frac{u}{M\|u\|_{\mathsf{U}}}\Big\|_{\mathsf{U}} < \frac{2}{M} \quad \Longrightarrow \quad \Big\|\frac{\mathsf{L}(u)}{M\|u\|_{\mathsf{U}}}\Big\|_{\mathsf{V}} < 1 \quad \Longrightarrow \quad \|\mathsf{L}(u)\|_{\mathsf{V}} < M\|u\|_{\mathsf{U}}.$$

Thus L is bounded.

(iv) $\Longrightarrow$ (i) Let $M \in \mathbb{R}_{>0}$ be such that $\|\mathsf{L}(u)\|_{\mathsf{V}} < M\|u\|_{\mathsf{U}}$ for all $u \in \mathsf{U}$. For $\epsilon \in \mathbb{R}_{>0}$ let $\delta = \frac{\epsilon}{M}$. If $u_0 \in \mathsf{U}$ and if $\|u - u_0\|_{\mathsf{U}} < \delta$ we have

$$\|\mathsf{L}(u) - \mathsf{L}(u_0)\|_{\mathsf{V}} = \|\mathsf{L}(u - u_0)\|_{\mathsf{V}} \leq M\|u - u_0\|_{\mathsf{U}} < \epsilon.$$

This gives continuity of L.

(iv) $\Longrightarrow$ (v) Let $(u_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in $\ker(\mathsf{L})$ converging to $v \in \mathsf{V}$. Then, since L is bounded,

$$\|\mathsf{L}(v) - \mathsf{L}(u_j)\|_{\mathsf{V}} = \|\mathsf{L}(v - u_j)\|_{\mathsf{V}} \leq M\|v - u_j\|_{\mathsf{U}}.$$

Therefore, if $\epsilon \in \mathbb{R}_{>0}$ we can take $N \in \mathbb{Z}_{>0}$ sufficiently large that $\|v - u_j\|_{\mathsf{U}} < \frac{\epsilon}{M}$, and for $j \geq N$ we have

$$\|\mathsf{L}(v)\|_{\mathsf{V}} = \|\mathsf{L}(v) - \mathsf{L}(u_j)\|_{\mathsf{V}} < \epsilon.$$

Thus $\mathsf{L}(v) = 0_{\mathsf{U}}$ and so $v \in \ker(\mathsf{L})$. Thus $\ker(\mathsf{L})$ is closed by Proposition 3.6.8 below. ■

In finite-dimensions, as is so often the case, things simplify.

3.5.9 Theorem (Linear maps from finite-dimensional spaces are continuous) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces. If U is finite-dimensional then* $\mathrm{L}(\mathsf{U}; \mathsf{V}) = \mathrm{Hom}_{\mathbb{F}}(\mathsf{U}; \mathsf{V})$.

Proof Let $\{e_1, \dots, n\}$ be a basis for U and denote

$$M' = \max\{\|\mathsf{L}(e_1)\|_{\mathsf{V}}, \dots, \|\mathsf{L}(e_n)\|_{\mathsf{V}}\}.$$

Define a norm $\|\cdot\|_{1,\mathsf{U}}$ on U by

$$\|u_1 e_1 + \dots + u_n e_n\| = |u_1| + \dots + |u_n|.$$

By Theorem 3.1.15 there exists $C \in \mathbb{R}_{>0}$ such that $\|u\|_{1,\mathsf{U}} \leq C\|u\|_{\mathsf{U}}$ for all $u \in \mathsf{U}$. Take $M = CM'$. Then, for $u = u_1 e_1 + \dots + u_n e_n \in \mathsf{U}$,

$$\begin{aligned}\|\mathsf{L}(u)\|_{\mathsf{V}} &= \|\mathsf{L}(u_1 e_1 + \dots + u_n e_n)\|_{\mathsf{V}} \\ &\leq |u_1|\|\mathsf{L}(e_1)\|_{\mathsf{V}} + \dots + |u_n|\|\mathsf{L}(e_n)\|_{\mathsf{V}} \\ &\leq M'\|u\|_{1,\mathsf{U}} \leq M\|u\|_{\mathsf{U}},\end{aligned}$$

showing that L is bounded, and so continuous. ■

Let us give some examples of continuous and discontinuous linear maps, noting that the only interesting examples are infinite-dimensional.

3.5.10 Examples (Continuous linear maps)

1. On any normed vector space $(\mathsf{V}, \|\cdot\|)$ the linear operator $(\mathrm{id}_{\mathsf{V}}, \mathsf{V})$ is continuous and invertible.
2. Let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an inner product space with $(e_j)_{j\in\mathbb{Z}_{>0}}$ a complete orthonormal family in V. For $v \in \mathsf{V}$ let $a_j(v)$ be the components of v in the complete orthonormal family so that

$$v = \sum_{j=1}^{\infty} a_j(v) e_j.$$

   We define the ***shift operator*** on V to be the linear operator $(\mathsf{L}, \mathrm{dom}(\mathsf{L}) = \mathsf{V})$ defined by

$$\mathsf{L}(v) = \sum_{j=1}^{\infty} a_j(v) e_{j+1}.$$

   By Parseval's inequality we have $\|\mathsf{L}(v)\| \leq \|v\|$, thus L is continuous.
3. We take the normed $\mathbb{F}$-vector space $\mathsf{C}^0([a,b];\mathbb{F})$ of continuous $\mathbb{F}$-valued functions on $[a,b]$ equipped with the norm $\|\cdot\|_\infty$ as in Example 3.1.3–10. Define $\mathsf{L}\colon \mathsf{C}^0([a,b];\mathbb{F}) \to \mathsf{C}^0([a,b];\mathbb{F})$ by

$$\mathsf{L}(f)(x) = \int_a^x f(\xi)\,\mathrm{d}\xi.$$

   It is easy to show that L is linear, using linearity of the integral. We claim that L is also continuous. To prove this, it suffices to prove that L is continuous at zero. Let $\epsilon \in \mathbb{R}_{>0}$ and let $\delta = \frac{\epsilon}{b-a}$. Then, if $\|f\|_\infty < \delta$,

$$\begin{aligned} \|\mathsf{L}(f)\|_\infty &= \sup\{|\mathsf{L}(f)(x)| \mid x \in [a,b]\} \\ &= \sup\left\{\left|\int_a^x f(\xi)\,\mathrm{d}\xi\right| \;\middle|\; x \in [a,b]\right\} \\ &\leq \sup\left\{\int_a^x |f(\xi)|\,\mathrm{d}\xi \;\middle|\; x \in [a,b]\right\} \\ &\leq \delta(b-a) = \epsilon, \end{aligned}$$

as desired.

4. Let $\mathsf{C}^1([0,1];\mathbb{R})$ be the $\mathbb{R}$-vector space of continuously differentiable $\mathbb{R}$-valued functions on $[0,1]$. Define $\mathsf{L}\colon \mathsf{C}^1([0,1];\mathbb{R}) \to \mathsf{C}^0([0,1];\mathbb{R})$ by $\mathsf{L}(f) = f'$. By linearity of the derivative, L is linear. We claim that L is not continuous if we use the norm $\|\cdot\|_\infty$ on both $\mathsf{C}^1([0,1];\mathbb{R})$ and $\mathsf{C}^0([0,1];\mathbb{R})$. To show this we shall use the following lemma that is useful in its own right.

1 Lemma *Let $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$, let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces, and let $\mathsf{L} \in \mathrm{Hom}_{\mathbb{F}}(\mathsf{U};\mathsf{V})$. Then L is discontinuous if and only if there exists a sequence $(\mathsf{u}_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{B}_{\mathsf{U}}(1, 0_{\mathsf{U}})$ such that the sequence $(\|\mathsf{L}(\mathsf{u}_j)\|_{\mathsf{V}})_{j\in\mathbb{Z}_{>0}}$ diverges.*

Proof Suppose that L is continuous. Then there exists $M \in \mathbb{R}_{>0}$ such that $\mathsf{L}(\mathsf{B}_{\mathsf{U}}(1, 0_{\mathsf{U}})) \subseteq \mathsf{B}_{\mathsf{V}}(M, 0_{\mathsf{V}})$ by boundedness of L. Thus, there can exist no sequence $(u_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{B}_{\mathsf{U}}(1, 0_{\mathsf{U}})$ such that the sequence $(\|\mathsf{L}(u_j)\|_{\mathsf{V}})_{j\in\mathbb{Z}_{>0}}$ is unbounded.
No suppose that there is a sequence $(u_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{B}_{\mathsf{U}}(1, 0_{\mathsf{U}})$ such that the sequence $(\|\mathsf{L}(u_j)\|_{\mathsf{V}})_{j\in\mathbb{Z}_{>0}}$ diverges. Then, for any $M \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{Z}_{>0}$ such that

$$\|\mathsf{L}(u_j)\|_{\mathsf{V}} \geq M \geq M\|u_j\|_{\mathsf{U}}, \qquad j \geq N.$$

Thus L is unbounded, and so not continuous. ▼

Now consider the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{C}^1([0,1];\mathbb{R})$ given by $f_j(x) = x^j$. This sequence satisfies $\|f_j\|_\infty = 1$. But $\mathsf{L}(f_j)(x) = jx^{j-1}$, and so $\|\mathsf{L}(f_j)\|_\infty = j$, showing that the sequence $(\|\mathsf{L}(f_j)\|_\infty)_{j\in\mathbb{Z}_{>0}}$ diverges. By the lemma it follows that L is discontinuous.

5. •

As a final basic result, let us show that continuous linear maps extend uniquely to the closure. We have not yet defined closure for normed vector spaces, so if you feel like you need to be reminded about what it is, you may refer ahead to Definition 3.6.7.

3.5.11 Proposition (Extension of continuous linear maps to the closure) *Let $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$, let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces with V complete, and let $\mathsf{W} \subseteq \mathsf{U}$ be a subspace for which $\mathrm{cl}(\mathsf{W}) = \mathsf{U}$. Then, for $\mathsf{L} \in \mathrm{L}(\mathsf{W};\mathsf{V})$ there exists a unique $\bar{\mathsf{L}} \in \mathrm{L}(\mathsf{U};\mathsf{V})$ such that $\bar{\mathsf{L}}(\mathrm{w}) = \mathsf{L}(\mathrm{w})$ for all $\mathrm{w} \in \mathsf{W}$.*

Proof We let $u \in \mathsf{U}$ and let $(w_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence with the property that $\lim_{j\to\infty}\|u - w_j\|_{\mathsf{U}} = 0$. We first claim that $(\mathsf{L}(w_j))_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence. Let $M \in \mathbb{R}_{>0}$ be such that $\|\mathsf{L}(w)\|_{\mathsf{V}} \le M\|w\|_{\mathsf{U}}$ for all $w \in \mathsf{W}$. Then

$$\|\mathsf{L}(w_j) - \mathsf{L}(w_k)\|_{\mathsf{V}} = \|\mathsf{L}(w_j - w_k)\|_{\mathsf{V}} \le M\|w_j - w_k\|_{\mathsf{U}}.$$

Since $(w_j)_{j\in\mathbb{Z}_{>0}}$ converges it is a Cauchy sequence, and so it follows that there exists $N \in \mathbb{Z}_{>0}$ for which $\|w_j - w_k\|_{\mathsf{U}} < \frac{\epsilon}{M}$ for $j,k \ge N$. This gives $\|\mathsf{L}(w_j) - \mathsf{L}(w_k)\|_{\mathsf{V}} < \epsilon$ for $j,k \ge N$, so showing that $(\mathsf{L}(w_j))_{j\in\mathbb{Z}_{>0}}$ is indeed a Cauchy sequence. Since $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ is complete, there exists $\bar{\mathsf{L}}(u) \in \mathsf{V}$ which is the limit of the sequence $(\mathsf{L}(w_j))_{j\in\mathbb{Z}_{>0}}$. Next we claim that this limit is independent of the sequence $(w_j)_{j\in\mathbb{Z}_{>0}}$ in W that converges to $u \in \mathsf{U}$. Thus let $(\tilde{w}_j)_{j\in\mathbb{Z}_{>0}}$ be another sequence in W converging to u. We denote by $\tilde{\mathsf{L}}(u)$ the limit in V of the Cauchy sequence $(\mathsf{L}(\tilde{w}_j))_{j\in\mathbb{Z}_{>0}}$. For $j \in \mathbb{Z}_{>0}$ we have

$$\|w_j - \tilde{w}_j\|_{\mathsf{U}} \le \|w_j - u\|_{\mathsf{U}} + \|\tilde{w}_j - u\|_{\mathsf{U}},$$

implying that $\lim_{j\to\infty}\|w_j - \tilde{w}_j\|_{\mathsf{U}} = 0$. Therefore

$$\|\bar{\mathsf{L}}(u) - \tilde{\mathsf{L}}(u)\|_{\mathsf{V}} \le \|\bar{\mathsf{L}}(u) - \mathsf{L}(w_j)\|_{\mathsf{V}} + \|\tilde{\mathsf{L}}(u) - \mathsf{L}(\tilde{w}_j)\|_{\mathsf{V}} + \|\mathsf{L}(\tilde{w}_j) - \mathsf{L}(w_j)\|_{\mathsf{V}}.$$

Taking the limit as $j \to \infty$ we see that $\|\bar{\mathsf{L}}(u) - \tilde{\mathsf{L}}(u)\|_{\mathsf{V}}$ can be made smaller than any positive number, and so must be zero.

This then gives us a well-defined element $\bar{\mathsf{L}}(u)$ associated to each $u \in \mathsf{U}$. We next claim that the assignment $u \mapsto \bar{\mathsf{L}}(u)$ is linear. For $u, \tilde{u} \in \mathsf{U}$ let $(w_j)_{j\in\mathbb{Z}_{>0}}$ and $(\tilde{w}_j)_{j\in\mathbb{Z}_{>0}}$ be sequences in W converging to u and $\tilde{u}$, respectively. Then $(w_j + \tilde{w}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $u + \tilde{u}$ by Proposition 3.2.6. Similarly, $(aw_j)_{j\in\mathbb{Z}_{>0}}$ converges to au for $a \in \mathbb{F}$. Therefore

$$\begin{aligned}\|\bar{\mathsf{L}}(u) + \bar{\mathsf{L}}(\tilde{u}) - \bar{\mathsf{L}}(u + \tilde{u})\|_{\mathsf{V}} &\le \|\bar{\mathsf{L}}(u) + \bar{\mathsf{L}}(\tilde{u}) - \mathsf{L}(w_j) - \mathsf{L}(\tilde{w}_j)\|_{\mathsf{V}} \\ &\quad + \|\bar{\mathsf{L}}(u + \tilde{u}) - \mathsf{L}(w_j + \tilde{w}_j)\|_{\mathsf{V}} \\ &\le \|\bar{\mathsf{L}}(u) - \mathsf{L}(w_j)\|_{\mathsf{V}} + \|\bar{\mathsf{L}}(\tilde{u}) - \mathsf{L}(\tilde{w}_j)\|_{\mathsf{V}} \\ &\quad + \|\bar{\mathsf{L}}(u + \tilde{u}) - \mathsf{L}(w_j + \tilde{w}_j)\|_{\mathsf{V}}.\end{aligned}$$

Taking the limit as $j \to \infty$ shows that the left hand side must be zero, giving $\bar{\mathsf{L}}(u + \tilde{u}) = \bar{\mathsf{L}}(u) + \bar{\mathsf{L}}(u)$. In an entirely similar way we have

$$\|\bar{\mathsf{L}}(au) - a\bar{\mathsf{L}}(u)\|_{\mathsf{V}} \le \|\bar{\mathsf{L}}(au) - \mathsf{L}(aw_j)\|_{\mathsf{V}} + \|a\bar{\mathsf{L}}(u) - a\mathsf{L}(w_j)\|_{\mathsf{V}},$$

and taking the limit $j \to \infty$ gives $\bar{\mathsf{L}}(au) - a\bar{\mathsf{L}}(u)$.

Let us now demonstrate the uniqueness of the extension $\bar{\mathsf{L}}$. Suppose that $\tilde{\mathsf{L}} \in \mathsf{L}(\mathsf{U}; \mathsf{V})$ is another continuous linear map with the property that it agrees with L on W. For $u \in \mathsf{U}$ let $(w_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in W converging to u. Then

$$\tilde{\mathsf{L}}(u) = \lim_{j\to\infty} \tilde{\mathsf{L}}(w_j) = \lim_{j\to\infty} \mathsf{L}(w_j) = \bar{\mathsf{L}}(w_j)$$

Move

by continuity of $\tilde{\mathsf{L}}$.

Finally we show that the operator norm of $\bar{\mathsf{L}}$ is the same as that of L. Since $\bar{\mathsf{L}}$ and L agree on W we have

$$\|\bar{\mathsf{L}}\|_{\mathsf{U},\mathsf{V}} = \sup_{\|u\|_{\mathsf{U}}=1} \|\mathsf{L}(u)\|_{\mathsf{V}} \ge \sup_{\|w\|_{\mathsf{U}}=1} \|\mathsf{L}(w)\|_{\mathsf{V}} = \|\mathsf{L}\|_{\mathsf{W},\mathsf{V}}.$$

Now we prove the opposite inequality. Let $u \in \mathsf{U}$ and let $(w_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in W converging to u. We then have

$$\|\bar{\mathsf{L}}(u)\|_{\mathsf{V}} = \lim_{j\to\infty}\|\mathsf{L}(w_j)\| \le \lim_{j\to\infty}\|\mathsf{L}\|_{\mathsf{W},\mathsf{V}}\|w_j\|_{\mathsf{U}} = \|\mathsf{L}\|_{\mathsf{W},\mathsf{V}}\|u\|_{\mathsf{U}}.$$

This gives the desired inequality since this must hold for all $u \in \mathsf{U}$, and so concludes the proof. ■

We also have the following related result.

3.5.12 Proposition (Extension of isomorphisms from dense subspaces) *Let* $(\mathsf{V}, \|\cdot\|)$ *be a Banach space with* W *a dense subspace. Suppose that* $\mathsf{L} \in \mathrm{L}_c(\mathsf{V};\mathsf{V})$ *is a continuous linear map with the property that* $\mathsf{L}|\mathsf{W}$ *is a continuous norm preserving bijection from* W *to itself with* $(\mathsf{L}|\mathsf{W})^{-1}$ *being continuous.*[2] *Then* L *is an isomorphism, and* L^{-1} *is the extension, as defined by Proposition 3.5.11, of* $(\mathsf{L}|\mathsf{W})^{-1}$ *to* V.

Proof First we note that by Proposition 3.5.11, $\|\mathsf{L}\|_{\mathsf{V},\mathsf{V}} = \|\mathsf{L}|\mathsf{W}\|_{\mathsf{W},\mathsf{W}}$. We claim that this implies that L is norm-preserving. Indeed, let $v \in \mathsf{V}$ and let $(w_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in W converging to v. Then

$$\|\mathsf{L}(u)\| = \lim_{j\to\infty}\|\mathsf{L}(w_j)\| = \lim_{j\to\infty}\|w_j\| = \|u\|,$$

as desired. We next claim that this implies injectivity of L. Indeed, if $\mathsf{L}(v) = 0$ for $v \in \mathsf{V}$ we must then have $\|v\| = \|\mathsf{L}(v)\| = 0$, giving $v = 0$. Thus L is injective. We also claim that $\operatorname{image}(\mathsf{L})$ is a closed subspace. Let $(\mathsf{L}(v_j))_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\operatorname{image}(\mathsf{L})$ converging to $u \in \mathsf{V}$. Then since $\|\mathsf{L}(v_j) - \mathsf{L}(v_k)\| = \|v_j - v_k\|$ it follows that $(v_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence. Let $v \in \mathsf{V}$ denote the limit of this sequence. We need to show that $\mathsf{L}(v) = u$. Indeed,

$$\|\mathsf{L}(v) - u\| \le \|\mathsf{L}(v) - \mathsf{L}(v_j)\| + \|u - \mathsf{L}(v_j)\|,$$

and taking the limit as $j \to \infty$ gives $\|\mathsf{L}(v) - u\| = 0$, so showing that $\operatorname{image}(\mathsf{L})$ is closed. Since $\mathsf{W} \subseteq \operatorname{image}(\mathsf{L})$ and since $\operatorname{cl}(\mathsf{W}) = \mathsf{V}$ we must have $\operatorname{cl}(\operatorname{image}(\mathsf{L})) = \operatorname{image}(\mathsf{L}) = \mathsf{V}$, thus showing surjectivity of L.

Finally we must show that L^{-1} is the unique continuous extension of $(\mathsf{L}|\mathsf{W})^{-1}$ to V. Let M denote the unique continuous extension of $(\mathsf{L}|\mathsf{W})^{-1}$ to V. Just as L is a continuous bijection, so too is M. Let $v \in \mathsf{V}$ and let $(w_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in W converging to $\mathsf{L}(v)$. Then

$$\mathsf{M} \circ \mathsf{L}(v) = \lim_{j\to\infty}(\mathsf{L}|\mathsf{W})^{-1}(w_j).$$

There then exists a sequence $(u_j)_{j\in\mathbb{Z}_{>0}}$ so that $\mathsf{L}(u_j) = w_j$, $j \in \mathbb{Z}_{>0}$. We then have

$$\mathsf{M} \circ \mathsf{L}(v) = \lim_{j\to\infty}(\mathsf{L}|\mathsf{W})^{-1} \circ \mathsf{L}(u_j) = \lim_{j\to\infty} u_j.$$

We claim that $\lim_{j\to\infty} u_j = v$. Since L is continuous and injective, this is equivalent to showing that $\lim_{j\to\infty} \mathsf{L}(u_j) = \mathsf{L}(v)$. However, this follows directly from the definition of

[2]The assumption that $(\mathsf{L}|\mathsf{W})^{-1}$ be continuous is actually superfluous by the ***Banach Isomorphism Theorem***.

the sequence $(u_j)_{j\in\mathbb{Z}_{>0}}$. Next let $v \in \mathsf{V}$ and let $(w_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in W converging to $\mathsf{M}(v)$. Then

$$\mathsf{L}\circ\mathsf{M}(v) = \lim_{j\to\infty} \mathsf{L}(w_j).$$

Let $(u_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in W with the property that $(\mathsf{L}|\mathsf{W})^{-1}(u_j) = w_j$, $j \in \mathbb{Z}_{>0}$. Then we have

$$\mathsf{L}\circ\mathsf{M}(v) = \lim_{j\to\infty} \mathsf{L}\circ(\mathsf{L}|\mathsf{W})^{-1}(u_j) = \lim_{j\to\infty} u_j.$$

We must show that $\lim_{j\to\infty} u_j = v$. Since M is continuous and injective this is equivalent to showing that $\lim_{j\to\infty} \mathsf{M}(u_j) = \mathsf{M}(v)$. This follows, however, from the definition of the sequence $(u_j)_{j\in\mathbb{Z}_{>0}}$. Thus we have shown that $\mathsf{M}\circ\mathsf{L}(v) = \mathsf{L}\circ\mathsf{M}(v) = v$ for all $v \in \mathsf{V}$. Thus $\mathsf{M} = \mathsf{L}^{-1}$. ■

3.5.3 Induced topologies on continuous linear maps

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces. In Corollary I-5.4.17 we showed that $\mathrm{Hom}_{\mathbb{F}}(\mathsf{U};\mathsf{V})$ is an $\mathbb{F}$-vector space. This is a purely algebraic observation. Now we wish to study the structure of the continuous linear maps. As we shall see, this is itself a normed vector space.

First we should establish that the set of continuous linear maps form a vector space.

3.5.13 Proposition (L(U; V) is a subspace of Hom$_{\mathbb{F}}$(U; V)) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ are normed $\mathbb{F}$-vector spaces, then $\mathrm{L}(\mathsf{U};\mathsf{V})$ is a subspace of $\mathrm{Hom}_{\mathbb{F}}(\mathsf{U};\mathsf{V})$.*

Proof Let $\mathsf{L}_1, \mathsf{L}_2 \in \mathrm{L}(\mathsf{U};\mathsf{V})$. For $\epsilon \in \mathbb{R}_{>0}$ let $\delta \in \mathbb{R}_{>0}$ be such that $\|\mathsf{L}_1(u)\|_{\mathsf{V}} < \frac{\epsilon}{2}$ and $\|\mathsf{L}_2(u)\|_{\mathsf{V}} < \frac{\epsilon}{2}$ for $\|u\|_{\mathsf{U}} < \delta$. Then compute

$$\|(\mathsf{L}_1 + \mathsf{L}_2)(u)\|_{\mathsf{V}} \le \|\mathsf{L}_1(u)\|_{\mathsf{V}} + \|\mathsf{L}_2(u)\|_{\mathsf{V}} < \epsilon,$$

showing that $\mathsf{L}_1 + \mathsf{L}_2$ is continuous at 0_{U}, and so continuous. Also let $a \in \mathbb{F}$ and $\mathsf{L} \in \mathrm{L}(\mathsf{U};\mathsf{V})$. If $a = 0$ it is clear that $a\mathsf{L}$ is continuous. So suppose that $a \neq 0$, let $\epsilon \in \mathbb{R}_{>0}$, and let $\delta \in \mathbb{R}_{>0}$ be such that if $\|u\|_{\mathsf{U}} < \delta$ then $\|\mathsf{L}(u)\|_{\mathsf{V}} < \frac{\epsilon}{|a|}$. For $\|u\|_{\mathsf{U}} < \delta$ we then have

$$\|(a\mathsf{L})(u)\|_{\mathsf{V}} = |a|\|\mathsf{L}(u)\|_{\mathsf{V}} < \epsilon,$$

giving continuity of $a\mathsf{L}$. ■

This shows that $\mathrm{L}(\mathsf{U};\mathsf{V})$ is indeed an $\mathbb{F}$-vector space. It is moreover true that it is a *normed* vector space.

3.5.14 Theorem (L(U; V) is a normed vector space) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces. For $\mathsf{L} \in \mathrm{L}(\mathsf{U};\mathsf{V})$ define*

$$\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} = \inf\{\mathrm{M} \in \mathbb{R}_{>0} \mid \|\mathsf{L}(\mathrm{u})\|_{\mathsf{V}} \le \mathrm{M}\|\mathrm{u}\|_{\mathsf{U}},\ \mathrm{u} \in \mathsf{U}\}.$$

Then $\|\cdot\|_{\mathsf{U},\mathsf{V}}$ is a norm on $\mathrm{L}(\mathsf{U};\mathsf{V})$. Moreover,

(i) $\|\mathsf{L}(\mathrm{u})\|_{\mathsf{V}} \le \|\mathsf{L}\|_{\mathsf{U},\mathsf{V}}\|\mathrm{u}\|_{\mathsf{U}}$ *for all* $\mathrm{u} \in \mathsf{U}$,

(ii) $\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} = \sup\left\{\frac{\|\mathsf{L}(\mathrm{u})\|_{\mathsf{V}}}{\|\mathrm{u}\|_{\mathsf{U}}} \;\middle|\; \mathrm{u} \in \mathsf{U} \setminus \{0_{\mathsf{V}}\}\right\}$,

(iii) $\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} = \sup\{\|\mathsf{L}(\mathrm{u})\|_{\mathsf{V}} \mid \|\mathrm{u}\|_{\mathsf{U}} = 1\}$, *and*

(iv) $\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} = \sup\{\|\mathsf{L}(\mathrm{u})\|_{\mathsf{V}} \mid \|\mathrm{u}\|_{\mathsf{U}} \le 1\}$, *and*

(v) *if* $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ *is complete then so is* $(\mathrm{L}(\mathsf{U};\mathsf{V}), \|\cdot\|_{\mathsf{U},\mathsf{V}})$.

Proof Let us first verify (i), disregarding whether or not $\|\cdot\|_{\mathsf{U},\mathsf{V}}$ is a norm. Suppose that (i) does not hold. Then there exists $u \in \mathsf{U}$ such that $\|\mathsf{L}(u)\|_{\mathsf{V}} > \|\mathsf{L}\|_{\mathsf{U},\mathsf{V}}\|u\|_{\mathsf{U}}$. Thus there exists $\epsilon \in \mathbb{R}_{>0}$ such that

$$\|\mathsf{L}(u)\|_{\mathsf{V}} > (\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} - \epsilon)\|u\|_{\mathsf{U}},$$

and this contradicts the definition of $\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}}$.

We next note that $\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} \in \mathbb{R}_{>0}$ for every $\mathsf{L} \in \mathrm{L}(\mathsf{U};\mathsf{V})$. Moreover, $\|0_{\mathrm{L}(\mathsf{U};\mathsf{V})}\|_{\mathsf{U},\mathsf{V}} = 0$. Now suppose that $\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} = 0$. Then

$$\|\mathsf{L}(u)\|_{\mathsf{V}} \le \|\mathsf{L}\|_{\mathsf{U},\mathsf{V}}\|u\|_{\mathsf{U}} = 0, \qquad u \in \mathsf{U}.$$

Thus $\mathsf{L} = 0_{\mathrm{L}(\mathsf{U};\mathsf{V})}$. Clearly we have $\|0\mathsf{L}\|_{\mathsf{U},\mathsf{V}} = |0|\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}}$. If $a \in \mathbb{F} \setminus \{0\}$ then we compute

$$\begin{aligned}\|a\mathsf{L}\|_{\mathsf{U},\mathsf{V}} &= \inf\{M \in \mathbb{R}_{>0} \mid \|a\mathsf{L}(u)\|_{\mathsf{V}} \le M\|u\|_{\mathsf{U}}, u \in \mathsf{U}\}\\ &= \inf\{M \in \mathbb{R}_{>0} \mid |a|\|\mathsf{L}(u)\|_{\mathsf{V}} \le M\|u\|_{\mathsf{U}}, u \in \mathsf{U}\}\\ &= \inf\left\{M \in \mathbb{R}_{>0} \;\middle|\; \|\mathsf{L}(u)\|_{\mathsf{V}} \le \frac{M}{|a|}\|u\|_{\mathsf{U}}, u \in \mathsf{U}\right\}\\ &= \inf\{|a|M' \in \mathbb{R}_{>0} \mid \|\mathsf{L}(u)\|_{\mathsf{V}} \le M'\|u\|_{\mathsf{U}}, u \in \mathsf{U}\} = |a|\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}},\end{aligned}$$

using Proposition I-2.2.28. Finally, if $\mathsf{L}_1, \mathsf{L}_2 \in \mathrm{L}(\mathsf{U};\mathsf{V})$ then

$$\begin{aligned}\|\mathsf{L}_1 + \mathsf{L}_2\|_{\mathsf{U},\mathsf{V}} &= \inf\{M \in \mathbb{R}_{>0} \mid \|(\mathsf{L}_1 + \mathsf{L}_2)(u)\|_{\mathsf{V}} \le M\|u\|_{\mathsf{U}},\ u \in \mathsf{U}\}\\ &\le \inf\{M \in \mathbb{R}_{>0} \mid \|\mathsf{L}_1(u)\|_{\mathsf{V}} + \|\mathsf{L}_2(u)\|_{\mathsf{V}} \le M\|u\|_{\mathsf{U}},\ u \in \mathsf{U}\}\\ &= \inf\{M_1 + M_2 \in \mathbb{R}_{>0} \mid \|\mathsf{L}_1(u)\|_{\mathsf{V}} \le M_1\|u\|_{\mathsf{U}},\\ &\quad \|\mathsf{L}_2(u)\|_{\mathsf{V}} \le M_2\|u\|_{\mathsf{U}},\ u \in \mathsf{U}\}\\ &= \inf\{M \in \mathbb{R}_{>0} \mid \|\mathsf{L}_1(u)\|_{\mathsf{V}} \le M\|u\|_{\mathsf{U}},\ u \in \mathsf{U}\}\\ &\quad + \inf\{M \in \mathbb{R}_{>0} \mid \|\mathsf{L}_2(u)\|_{\mathsf{V}} \le M\|u\|_{\mathsf{U}},\ u \in \mathsf{U}\}\\ &= \|\mathsf{L}_1\|_{\mathsf{U},\mathsf{V}} + \|\mathsf{L}_2\|_{\mathsf{U},\mathsf{V}},\end{aligned}$$

where we have used Proposition I-2.2.28. This verifies that $\|\cdot\|_{\mathsf{U},\mathsf{V}}$ has the properties demanded of a norm.

(ii) First note that the equality is trivial when $\mathsf{L} = 0_{\mathrm{L}(\mathsf{U};\mathsf{V})}$, so we suppose this is not the case. In this case, $\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} > 0$ and so

$$\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} = \inf\{M \in \mathbb{R}_{>0} \mid \|\mathsf{L}(u)\|_{\mathsf{V}} \le M\|u\|_{\mathsf{U}},\ u \in \mathsf{U} \setminus \{0_{\mathsf{V}}\}\}$$

and so

$$\begin{aligned}\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} &= \inf\{M \in \mathbb{R}_{>0} \mid \|\mathsf{L}(u)\|_{\mathsf{V}} \le M\|u\|_{\mathsf{U}},\ u \in \mathsf{U} \setminus \{0_{\mathsf{V}}\}\}\\ &= \inf\left\{M \in \mathbb{R}_{>0} \;\middle|\; \frac{\|\mathsf{L}(u)\|_{\mathsf{V}}}{\|u\|_{\mathsf{U}}} \le M,\ u \in \mathsf{U} \setminus \{0_{\mathsf{U}}\}\right\}\\ &= \sup\left\{\frac{\|\mathsf{L}(u)\|_{\mathsf{V}}}{\|u\|_{\mathsf{U}}} \;\middle|\; u \in \mathsf{U} \setminus \{0_{\mathsf{U}}\}\right\}.\end{aligned}$$

(iii) Carrying on from part (ii) we have

$$\begin{aligned}\|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} &= \sup\Big\{\frac{\|\mathsf{L}(u)\|_{\mathsf{V}}}{\|u\|_{\mathsf{U}}}\ \Big|\ u \in \mathsf{U}\setminus\{0_{\mathsf{U}}\}\Big\}\\ &= \sup\Big\{\Big\|\mathsf{L}\Big(\frac{u}{\|u\|_{\mathsf{U}}}\Big)\Big\|\ \Big|\ u\in\mathsf{U}\setminus\{0_{\mathsf{U}}\}\Big\}\\ &= \sup\{\|\mathsf{L}(u)\|_{\mathsf{V}} \mid \|u\|_{\mathsf{U}} = 1\}.\end{aligned}$$

(iv) It is evident that

$$\sup\{\|\mathsf{L}(u)\|_{\mathsf{V}} \mid \|u\|_{\mathsf{U}} \le 1\} \ge \sup\{\|\mathsf{L}(u)\|_{\mathsf{V}} \mid \|u\|_{\mathsf{U}} = 1\},$$

the supremum on the left being taken over a larger set. On the other hand,

$$\begin{aligned}\sup\{\|\mathsf{L}(u)\|_{\mathsf{V}} \mid \|u\|_{\mathsf{U}} \le 1\} &= \sup\{\|\mathsf{L}(\lambda u)\|_{\mathsf{V}} \mid \lambda\in[0,1],\ \|u\|_{\mathsf{U}}=1\}\\ &= \sup\{\lambda\|\mathsf{L}(u)\|_{\mathsf{V}} \mid \lambda\in[0,1],\ \|u\|_{\mathsf{U}}=1\}\\ &\le \sup\{\|\mathsf{L}(u)\|_{\mathsf{V}} \mid \|u\|_{\mathsf{U}}=1\},\end{aligned}$$

giving the result.

(v) Let $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathrm{L}(\mathsf{U};\mathsf{V})$. We claim that $(\mathsf{L}_j(u))_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in V. This is clear if $u = 0_{\mathsf{U}}$, so let us suppose otherwise. Let $\epsilon\in\mathbb{R}_{>0}$ and let $N\in\mathbb{Z}_{>0}$ be sufficiently large that $\|\mathsf{L}_j-\mathsf{L}_k\|_{\mathsf{U},\mathsf{V}} < \frac{\epsilon}{\|u\|_{\mathsf{U}}}$ for $j,k\ge N$. Then

$$\|\mathsf{L}_j(u)-\mathsf{L}_k(u)\|_{\mathsf{V}} \le \|\mathsf{L}_j-\mathsf{L}_k\|_{\mathsf{U},\mathsf{V}}\|u\|_{\mathsf{V}} < \epsilon$$

for $j,k\ge N$. Thus the sequence $(\mathsf{L}_j(u))_{j\in\mathbb{Z}_{>0}}$ converges to an element in V which we denote by $\mathsf{L}(u)$. One may easily show that the assignment $u\mapsto\mathsf{L}(u)$ is well-defined and linear, cf. the proof of Proposition 3.5.11. Thus this defines $\mathsf{L}\in\mathrm{Hom}_{\mathbb{F}}(\mathsf{U};\mathsf{V})$.

We now show that L is continuous. Let $\epsilon\in\mathbb{R}_{>0}$ and let $N\in\mathbb{Z}_{>0}$ be such that $\|\mathsf{L}_j-\mathsf{L}_k\|_{\mathsf{U},\mathsf{V}}<\epsilon$ for $j,k\ge N$. Then, if $\|u\|_{\mathsf{U}}\le 1$,

$$\|(\mathsf{L}_j-\mathsf{L}_k)(u)\|_{\mathsf{V}} \le \|\mathsf{L}_j-\mathsf{L}_k\|_{\mathsf{U},\mathsf{V}}\|u\|_{\mathsf{U}} < \epsilon.$$

Using continuity of the norm and Theorem 3.5.2 we have, for fixed $j\ge N$,

$$\lim_{k\to\infty}\|(\mathsf{L}_j-\mathsf{L}_k)(u)\|_{\mathsf{V}} = \Big\|(\mathsf{L}_j-\lim_{k\to\infty}\mathsf{L}_k)(u)\Big\|_{\mathsf{V}} = \|(\mathsf{L}_j-\mathsf{L})(u)\|_{\mathsf{V}} < \epsilon.$$

Therefore, for any $u\in\mathsf{U}$ we have

$$\|(\mathsf{L}_j-\mathsf{L})(u)\|_{\mathsf{V}} < \epsilon\|u\|_{\mathsf{U}},$$

implying that $\mathsf{L}_j-\mathsf{L}$ is bounded and so $\mathsf{L}_j-\mathsf{L}\in\mathrm{L}(\mathsf{U};\mathsf{V})$. Since $\mathsf{L}_j\in\mathrm{L}(\mathsf{U};\mathsf{V})$ and since $\mathrm{L}(\mathsf{U};\mathsf{V})$ is a subspace it follows that $\mathsf{L}\in\mathrm{L}(\mathsf{U};\mathsf{V})$.

Moreover, our computations also show that, for any $\epsilon\in\mathbb{R}_{>0}$ there exists $N\in\mathbb{Z}_{>0}$ such that $\|\mathsf{L}_j-\mathsf{L}\|_{\mathsf{U},\mathsf{V}}<\epsilon$ for $j\ge N$. Thus $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ converges to L. ■

Let us attach some terminology to our norm on $\mathrm{L}(\mathsf{U};\mathsf{V})$.

3.5.15 Definition (Induced norm, operator norm, convergence in norm) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces.

(i) The norm $\|\cdot\|_{\mathsf{U},\mathsf{V}}$ is the ***induced norm*** or the ***operator norm*** on $\mathrm{L}(\mathsf{U}; \mathsf{V})$.

(ii) A sequence $(\mathsf{L}_j)_{j \in \mathbb{Z}_{>0}}$ ***converges in norm*** if it converges in the normed $\mathbb{F}$-vector space $(\mathrm{L}(\mathsf{U}; \mathsf{V}), \|\cdot\|_{\mathsf{U},\mathsf{V}})$. •

Let us explicitly compute some operator norms.

3.5.16 Example (Example 3.5.10 cont'd)

1. We first consider the case of a linear map $\mathbf{L}\colon \mathbb{R}^m \to \mathbb{R}^n$. We claim that $\|\mathbf{L}\|_{\mathbb{R}^m,\mathbb{R}^n}$ is equal to the largest eigenvalue of the matrix $\mathbf{L}^T\mathbf{L}$. First note that $\mathbf{L}^T\mathbf{L}$ is a symmetric matrix, so its eigenvalues are all real. Furthermore, its eigenvalues are nonnegative since $\mathbf{L}^T\mathbf{L}$ is positive-semidefinite. Let $\boldsymbol{x} \in \mathbb{R}^m$ be an eigenvector for the largest eigenvalue λ of $\mathbf{L}^T\mathbf{L}$. We then compute

$$\|\mathbf{L}\boldsymbol{x}\|^2 = \boldsymbol{x}^T\mathbf{L}^T\mathbf{L}\boldsymbol{x} = \lambda^2\boldsymbol{x}^T\boldsymbol{x} = \lambda^2\|\boldsymbol{x}\|^2.$$

This shows that $\|\mathbf{L}\|_{\mathbb{R}^m,\mathbb{R}^n} \geq \lambda$. Now note that since $\mathbf{L}^T\mathbf{L}$ is symmetric we may find an orthonormal basis $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_m\}$ for $\mathbb{R}^m$ comprised of eigenvectors of $\mathbf{L}^T\mathbf{L}$. We may then write

$$\boldsymbol{x} = (\boldsymbol{x} \cdot \boldsymbol{v}_1)\boldsymbol{v}_1 + \dots + (\boldsymbol{x} \cdot \boldsymbol{v}_m)\boldsymbol{v}_m$$

for any $\boldsymbol{x} \in \mathbb{R}^m$. We then have

$$\begin{aligned}\|\mathbf{L}\boldsymbol{x}\|^2 &= \boldsymbol{x}^T\mathbf{L}^T\mathbf{L}\boldsymbol{x} \\ &= \sum_{i,j=1}^m (\boldsymbol{x} \cdot \boldsymbol{v}_i)(\boldsymbol{x} \cdot \boldsymbol{v}_j)\boldsymbol{v}_i^T\mathbf{L}^T\mathbf{L}\boldsymbol{v}_j \\ &= \sum_{i,j=1}^m (\boldsymbol{x} \cdot \boldsymbol{v}_i)(\boldsymbol{x} \cdot \boldsymbol{v}_j)\lambda_i\lambda_j\boldsymbol{v}_i^T\boldsymbol{v}_i \\ &= \sum_{i=1}^m \lambda_i^2(\boldsymbol{x} \cdot \boldsymbol{v}_i)^2 \leq \lambda^2\sum_{i=1}^m (\boldsymbol{x} \cdot \boldsymbol{v}_i)^2 = \lambda^2\|\boldsymbol{x}\|^2,\end{aligned}$$

thus showing that $\|\mathbf{L}\|_{\mathbb{R}^m,\mathbb{R}^n} \leq \lambda$.

2. For the linear map $\mathsf{L}(f)(t) = \int_0^t f(\tau)\,\mathrm{d}\tau$ defined on $(\mathsf{C}^0([0,1];\mathbb{R}), \|\cdot\|_\infty)$ we claim that the operator norm is 1. In Example 3.5.10 we showed that the operator norm is at least 1. To show that it is at most 1, consider the function $f(t) = c$ for some nonzero constant c. We then have $\|\mathsf{L}(f)\|_\infty = c$, giving our assertion. •

The induced norm also satisfies nice properties with respect to composition.

3.5.17 Proposition (Induced norm and composition) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$, $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$, *and* $(\mathsf{W}, \|\cdot\|_{\mathsf{W}})$ *be normed* $\mathbb{F}$*-vector spaces. If* $\mathsf{L} \in \mathrm{L}(\mathsf{U}; \mathsf{V})$ *and* $\mathsf{K} \in \mathrm{L}(\mathsf{V}; \mathsf{W})$ *then*

$$\|\mathsf{K} \circ \mathsf{L}\|_{\mathsf{U},\mathsf{W}} \leq \|\mathsf{K}\|_{\mathsf{V},\mathsf{W}} \|\mathsf{L}\|_{\mathsf{U},\mathsf{V}}.$$

In particular, $\mathsf{K} \circ \mathsf{L} \in \mathrm{L}(\mathsf{U}; \mathsf{W})$.

Proof For $u \in \mathsf{U}$ we compute

$$\|\mathsf{K} \circ \mathsf{L}(u)\|_{\mathsf{W}} \leq \|\mathsf{K}\|_{\mathsf{V},\mathsf{W}} \mathsf{L}(u) \leq \|\mathsf{K}\|_{\mathsf{V},\mathsf{W}} \|\mathsf{L}\|_{\mathsf{U},\mathsf{V}} \|u\|_{\mathsf{U}},$$

as desired. ■

As suggested by the terminology "converges in norm," we wish to allow other versions of convergence of sequences of continuous linear maps. The principal such notion is the following.

3.5.18 Definition (Weak convergence) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces. A sequence $(\mathsf{L}_j)_{j \in \mathbb{Z}_{>0}}$ in $\mathrm{L}(\mathsf{U}; \mathsf{V})$ ***converges weakly*** to $\mathsf{L} \in \mathrm{L}(\mathsf{U}; \mathsf{V})$ if, for each $u \in \mathsf{U}$, the sequence $(\mathsf{L}_j(u))_{j \in \mathbb{Z}_{>0}}$ converges. •

Let us explore weak convergence by providing its relationship with convergence in norm.

3.5.19 Proposition (Convergence in norm implies weak convergence) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ *and* $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ *be normed* $\mathbb{F}$*-vector spaces. A sequence* $(\mathsf{L}_{\mathsf{j}})_{\mathsf{j} \in \mathbb{Z}_{>0}}$ *in* $\mathrm{L}(\mathsf{U}; \mathsf{V})$ *converges weakly if it converges in norm.*

Proof This is Exercise 3.5.5. ■

It is not generally true that weak convergence implies convergence in norm. The following example relies on the reader knowing about Banach spaces of sequences as discussed in Section 3.8.2.

3.5.20 Example (Weak convergence may not imply norm convergence) We consider the $\mathbb{F}$-Banach space $\ell^2(\mathbb{F})$ of sequences $(a_j)_{j \in \mathbb{Z}_{>0}}$ in $\mathbb{F}$ for which $\sum_{j=1}^{\infty} |a_j|^2 < \infty$. This is a Banach space with norm

$$\|(a_j)_{j \in \mathbb{Z}_{>0}}\|_2 = \Big(\sum_{j=1}^{\infty} |a_j|^2\Big)^{1/2}.$$

For $k \in \mathbb{Z}_{>0}$ define $\mathsf{L}_k \in \mathrm{L}(\ell^2(\mathbb{R}); \mathbb{F})$ by $\mathsf{L}_k((a_j)_{j \in \mathbb{Z}_{>0}}) = a_k$ (it is clear that L_k is linear and bounded). Now note that

$$(\mathsf{L}_k - \mathsf{L}_l)((a_j)_{j \in \mathbb{Z}_{>0}}) = a_k - a_l$$

so that

$$|(\mathsf{L}_k - \mathsf{L}_l)((a_j)_{j \in \mathbb{Z}_{>0}})| \leq |a_k| + |a_l| \leq \sqrt{2}(|a_k|^2 + |a_l|^2)^{1/2} \leq \sqrt{2}\|(a_j)_{j \in \mathbb{Z}_{>0}}\|_2,$$

where we have used Proposition II-1.1.11. Thus $\|\mathsf{L}_k - \mathsf{L}_l\|_{\ell^2(\mathbb{F}),\mathbb{F}} \le \sqrt{2}$. However, taking the particular sequence

$$a_j = \begin{cases} 1, & j = k, \\ -1, & j = l, \\ 0, & \text{otherwise}, \end{cases}$$

we have

$$|(\mathsf{L}_k - \mathsf{L}_l)((a_j)_{j\in\mathbb{Z}_{>0}})| = \sqrt{2}\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_2,$$

showing that $\|\mathsf{L}_k - \mathsf{L}_l\|_{\ell^2(\mathbb{F}),\mathbb{F}} \le \sqrt{2}$. In particular, the sequence $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ is not Cauchy, and so does not converge in norm. We claim that it does, however, converge weakly. Indeed, if $(a_j)_{j\in\mathbb{Z}_{>0}} \in \ell^2(\mathbb{F})$ then we have $\lim_{j\to\infty}|a_j|^2 = 0$ by Proposition I-2.4.7. Therefore,

$$\lim_{k\to\infty} \mathsf{L}_k((a_j)_{j\in\mathbb{Z}_{>0}}) = \lim_{k\to\infty} a_k = 0,$$

showing that the sequence $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ converges weakly to the zero linear map. •

The preceding example notwithstanding, the reader may not be surprised to learn that weak and norm convergence agree in finite-dimensions.

3.5.21 Proposition (Equivalence of weak and norm convergence in finite-dimensions) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be finite-dimensional normed $\mathbb{F}$-vector spaces. A sequence $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathrm{L}(\mathsf{U};\mathsf{V})$ converges weakly if and only if it converges in norm.*

Proof Let $\{e_1, \dots, e_n\}$ be a basis for U. We claim that

$$|||\mathsf{L}||| = \max\{\|\mathsf{L}(e_1)\|_{\mathsf{V}}, \dots, \|\mathsf{L}(e_n)\|_{\mathsf{V}}\}$$

is a norm on $\mathrm{L}(\mathsf{U};\mathsf{V})$. The only possibly nontrivial fact to verify is the triangle inequality. For this we have

$$\begin{aligned} |||\mathsf{L}_1 + \mathsf{L}_2||| &= \max\{\|(\mathsf{L}_1 + \mathsf{L}_2)(e_1)\|_{\mathsf{V}}, \dots, \|(\mathsf{L}_1 + \mathsf{L}_2)(e_n)\|_{\mathsf{V}}\} \\ &\le \max\{\|\mathsf{L}_1(e_1)\|_{\mathsf{V}} + \|\mathsf{L}_2(e_1)\|_{\mathsf{V}}, \dots, \|\mathsf{L}_1(e_n)\|_{\mathsf{V}} + \|\mathsf{L}_2(e_n)\|_{\mathsf{V}}\} \\ &= \max\{\|\mathsf{L}_1(e_1)\|_{\mathsf{V}}, \dots, \|\mathsf{L}_1(e_n)\|_{\mathsf{V}}\} + \max\{\|\mathsf{L}_2(e_1)\|_{\mathsf{V}}, \dots, \|\mathsf{L}_2(e_n)\|_{\mathsf{V}}\} \\ &= |||\mathsf{L}_1||| + |||\mathsf{L}_2|||, \end{aligned}$$

as desired.

Now we claim that a sequence $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathrm{L}(\mathsf{U};\mathsf{V})$ converges weakly to L if and only if it converges to L in the norm $|||\cdot|||$. Indeed, weak convergence implies immediately that $\lim_{j\to\infty} \mathsf{L}_j(e_k) = \mathsf{L}(e_k)$ for each $k \in \{1, \dots, n\}$. This in turn implies convergence in the norm $|||\cdot|||$. Conversely, if a sequence $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ converges in the norm $|||\cdot|||$ then, for each $k \in \{1, \dots, n\}$, $(\mathsf{L}_j(e_k))_{j\in\mathbb{Z}_{>0}}$ converges in V to $\mathsf{L}(e_k)$. Thus, if $u = u_1e_1 + \dots + u_ne_n \in \mathsf{U}$ we have,

$$\lim_{j\to\infty} \mathsf{L}_j(u_1e_1 + \dots + u_ne_n) = \sum_{k=1}^{n} u_k \lim_{j\to\infty} \mathsf{L}(e_k) = \mathsf{L}(u_1e_1 + \dots + u_ne_n).$$

Thus $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ converges to L weakly.

The result follows from this since the norms $|||\cdot|||$ and $\|\cdot\|_{\mathsf{U},\mathsf{V}}$ are equivalent by virtue of $\mathrm{L}(\mathsf{U};\mathsf{V})$ being finite-dimensional (see Exercise I-5.4.8). ■

We close this section by indicating that weak convergence is, in fact, convergence in a suitable topology. The material here relies on an understanding of topics covered in . It is not necessary to understand this to understand weak convergence.

what?

3.5.22 Definition (Weak operator topology) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces. The ***weak operator topology*** is the topology for which sets of the form

$$\cap_{k=1}^m \{\mathsf{L} \in \mathrm{L}(\mathsf{U};\mathsf{V}) \mid \|\mathsf{L}(u_k) - \mathsf{L}_0(u_k)\| < \epsilon_k\}, \quad u_1, \dots, u_m \in \mathsf{U},\ \epsilon_1, \dots, \epsilon_m \in \mathbb{R}_{>0},$$

are a neighbourhood basis about L_0. •

what?

That this does indeed define a topology on $\mathrm{L}(\mathsf{U};\mathsf{V})$ follows from .

The following result connects the weak operator topology with the notion of weak convergence.

3.5.23 Theorem (Weak convergence is convergence in the weak operator topology) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces. Then a sequence $(\mathsf{L}_{\mathsf{j}})_{\mathsf{j}\in\mathbb{Z}_{>0}}$ in $\mathrm{L}(\mathsf{U};\mathsf{V})$ converges weakly to L_0 if and only if it converges to L_0 in the weak operator topology.*

Proof First suppose that $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ converges weakly to L_0. Let $S \subseteq \mathrm{L}(\mathsf{U};\mathsf{V})$ be a neighbourhood of L_0 in the weak operator topology and let $\epsilon_1, \dots, \epsilon_k \in \mathbb{R}_{>0}$ and $u_1, \dots, u_k \in \mathsf{U}$ be such that

$$\cap_{k=1}^m \{\mathsf{L} \in \mathrm{L}(\mathsf{U};\mathsf{V}) \mid \|\mathsf{L}(u_k) - \mathsf{L}_0(u_k)\|_{\mathsf{V}} < \epsilon_k\} \subseteq S.$$

For $k \in \{1, \dots, m\}$ let $N_k \in \mathbb{Z}_{>0}$ be sufficiently large that $\|\mathsf{L}_j(u_k) - \mathsf{L}_0(u_k)\|_{\mathsf{V}} < \epsilon_k$ for $j \geq N_k$ and let $N = \max\{N_1, \dots, N_m\}$. Then, for $j \geq N$ and for $k \in \{1, \dots, m\}$,

$$\|\mathsf{L}_j(u_k) - \mathsf{L}_0(u_k)\|_{\mathsf{V}} < \epsilon_k$$

so that

$$\mathsf{L}_j \in \cap_{k=1}^m \{\mathsf{L} \in \mathrm{L}(\mathsf{U};\mathsf{V}) \mid \|\mathsf{L}(u_k) - \mathsf{L}_0(u_k)\|_{\mathsf{V}} < \epsilon_k\}.$$

Thus $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ converges in the weak operator topology.

Now suppose that $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ converges to L_0 in the weak operator topology. For $\epsilon \in \mathbb{R}_{>0}$ and $u \in \mathsf{U}$ note that

$$S(\mathsf{L}_0, u, \epsilon) \triangleq \{\mathsf{L} \in \mathrm{L}(\mathsf{U};\mathsf{V}) \mid \|\mathsf{L}(u) - \mathsf{L}_0(u)\|_{\mathsf{V}} < \epsilon\}$$

is a neighbourhood of L_0 in the weak operator topology. Thus, for $\epsilon \in \mathbb{R}_{>0}$ and $u \in \mathsf{U}$, there exists $N \in \mathbb{Z}_{>0}$ such that $\mathsf{L}_j \in S(\mathsf{L}_0, u, \epsilon)$ for $j \geq N$. That is, for each $\epsilon \in \mathbb{R}_{>0}$ and for each $u \in \mathsf{U}$, there exists $N \in \mathbb{Z}_{>0}$ such that $\|\mathsf{L}(u) - \mathsf{L}_0(u)\|_{\mathsf{V}} < \epsilon$ showing that $(\mathsf{L}_j(u))_{j\in\mathbb{Z}_{>0}}$ converges to $\mathsf{L}_0(u)$. This is exactly weak convergence of $(\mathsf{L}_j)_{j\in\mathbb{Z}_{>0}}$ to L_0. ■

3.5.24 Remark (The weak operator topology is locally convex) As a glimpse ahead to Chapter 6 we make the observation that the weak operator topology is the locally convex topology defined by the family of seminorms $(p_u)_{u \in \mathsf{U}}$ where $p_u(\mathsf{L}) = \|\mathsf{L}(u)\|_{\mathsf{V}}$. • normable? metrisable?

3.5.4 Invertibility of continuous linear maps

The notion of invertibility of a linear map $\mathsf{L}\colon \mathbb{R}^n \to \mathbb{R}^n$ is well understood, and is equivalent to the condition that if we think of L as an $n \times n$ matrix then $\det(\mathsf{L}) \neq 0$. As expected, for linear operators defined on infinite-dimensional normed vector spaces, the issues are more complicated. Indeed, as we shall see, there are various ways in which a linear operator can be singular, and only some of the possibilities will be of interest to us.

Let us first consider injective linear operators. In the following discussion we let $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ be a linear operator on V. The following result has likely been encountered in a basic linear algebra course.

3.5.25 Lemma *A linear operator* $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ *on* V *is injective if and only if* $\ker(\mathsf{L}) = \{0\}$.

Proof First suppose that L is injective and that $\mathsf{L}(v) = 0$. Since $\mathsf{L}(0) = 0$ this implies that $v = 0$. Next suppose that $\ker(\mathsf{L}) = \{0\}$ and that $\mathsf{L}(v_1) = \mathsf{L}(v_2)$. Then $\mathsf{L}(v_1 - v_2) = 0$ by linearity, implying that $v_1 = v_2$. ■

If $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ is injective then $\mathsf{L}\colon \mathrm{dom}(\mathsf{L}) \to \mathrm{image}(\mathsf{L})$ is necessarily an isomorphism. In this case we define a linear operator $(\mathsf{L}^{-1}, \mathrm{image}(\mathsf{L}))$ where $\mathsf{L}\colon \mathrm{image}(\mathsf{L}) \subseteq \mathsf{V} \to \mathrm{dom}(\mathsf{L}) \subseteq \mathsf{U}$. Note that L^{-1} defined in this manner is *not* defined on all of V, only on $\mathrm{image}(\mathsf{L})$. We shall say that $(\mathsf{L}^{-1}, \mathrm{image}(\mathsf{L}))$ is the ***inverse*** of L, and so say that $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ is ***invertible***.

3.5.26 Definition Let $(\mathsf{V}, \|\cdot\|)$ be a normed vector space and let $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ be a linear operator on V.

(i) $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ is ***essentially regular*** if
  (a) L is injective,
  (b) $(\mathsf{L}^{-1}, \mathrm{image}(\mathsf{L}))$ is continuous, and
  (c) $\mathrm{cl}(\mathrm{image}(\mathsf{L})) = \mathsf{V}$.

(ii) $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ is ***regular*** if it is essentially regular and if $\mathrm{image}(\mathsf{L}) = \mathsf{V}$.

(iii) $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ is ***singular*** if it is neither regular nor essentially regular. •

In a manner resembling closed and closable linear operators, one can go from an essentially regular linear operator to a regular linear operator in a natural way.

3.5.27 Proposition *Let* $(\mathsf{V}, \|\cdot\|)$ *be a Banach space and let* $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ *be a linear operator on* V. *If* $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ *is essentially regular then there exists a regular linear operator* $(\bar{\mathsf{L}}, \mathrm{dom}(\bar{\mathsf{L}}))$ *on* V *which is an extension of* $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$. $(\bar{\mathsf{L}}, \mathrm{dom}(\bar{\mathsf{L}}))$ *is called the* ***regularisation*** *of* $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$.

Proof We proceed by defining $\overline{\mathsf{L}}^{-1}$. For any $v_0 \in \mathsf{V}$ there exists a Cauchy sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ in image(L) and converging to v_0. Since L^{-1} is continuous the sequence $(\mathsf{L}^{-1}(v_j))_{j\in\mathbb{Z}_{>0}}$ converges to $u_0 \in \mathsf{V}$. We define $\overline{\mathsf{L}}^{-1}(v_0) = u_0$. One shows that the collection of $u \in \mathsf{V}$ that are images under L^{-1} of Cauchy sequences in image(L) form a subspace of V, and we denote this subspace by dom($\overline{\mathsf{L}}$). One then defines $\overline{\mathsf{L}} = (\overline{\mathsf{L}}^{-1})^{-1}$. ■

Often one is interested in solutions of equations of the form $\mathsf{L}(v) = u$. To ensure the existence to such an equation, one wants $u \in \text{image}(\mathsf{L})$; to guarantee uniqueness of the solution, one wants L to be injective; and to ensure that the solutions of the equation do not vary wildly as one varies u, one wants the inverse to be continuous. This motivates our interest in regular linear operators.

classification of invertibility for continuous linear maps

3.5.28 Examples

1. On any normed vector space $(\mathsf{V}, \|\cdot\|)$ the linear operator $(\mathrm{id}_{\mathsf{V}}, \mathsf{V})$ is continuous and invertible. Furthermore, $\text{image}(\mathrm{id}_{\mathsf{V}}) = \mathsf{V}$, so this linear operator is regular.
2. On any normed vector space $(\mathsf{V}, \|\cdot\|)$ the linear operator (L, V) defined by $\mathsf{L}(v) = 0$ is continuous. It is certainly not invertible, however, so it represents an example of case of the theorem. what?
3. On $\mathsf{V} = (\mathsf{L}^2([0,1];\mathbb{F})$ consider the linear operator $(\mathsf{L}, \text{dom}(\mathsf{L}))$ given by $\text{dom}(\mathsf{L}) = \mathsf{V}$ and $\mathsf{L}(f)(t) = tf(t)$. It is clear that L is continuous since we have
$$\|\mathsf{L}(f)\|_2^2 = \int_0^1 |tf(t)|^2\,\mathrm{d}t \le \int_0^1 |f(t)|^2\,\mathrm{d}t = \|f\|_2^2.$$
It is also evident that L is invertible since $\mathsf{L}(f) = 0$ obviously implies that $f = 0$ a.e. We note that $\text{image}(\mathsf{L}) \subset \mathsf{V}$ since the function $f(t) = 1$ is not in the image of L. Indeed, if this function *were* in image(L) then there would be a function $f \in \mathsf{L}^2([0,1];\mathbb{F})$ so that $tf(t) = 1$. Thus $f(t) = \frac{1}{t}$, but this function is not in $\mathsf{L}^2([0,1];\mathbb{F})$. However, if we define
$$\mathsf{S} = \{f \in \mathsf{L}^2([0,1];\mathbb{F}) \mid \text{ there exists a neighbourhood of 0 on which } f \text{ vanishes}\},$$
then clearly $\mathsf{S} \subseteq \text{image}(\mathsf{L})$. Furthermore, one easily sees that S is dense in V, showing that image(L) is dense in V. This shows that $(\mathsf{L}, \text{dom}(\mathsf{L}))$ belongs to the functions of case of the theorem. what?
4. Let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an inner product space with $(e_j)_{j\in\mathbb{Z}_{>0}}$ a complete orthonormal family in V. For $v \in \mathsf{V}$ let $a_j(v)$ be the components of v in the complete orthonormal family so that
$$v = \sum_{j=1}^{\infty} a_j(v)e_j.$$
We define the ***shift operator*** on V to be the linear operator $(\mathsf{L}, \text{dom}(\mathsf{L}) = \mathsf{V})$ defined by
$$\mathsf{L}(v) = \sum_{j=1}^{\infty} a_j(v)e_{j+1}.$$

By Parseval's inequality we have $\|\mathsf{L}(v)\| \leq \|v\|$, thus L is continuous.

It is also evident that $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ is invertible. Indeed, suppose that $\mathsf{L}(v) = 0$. Then

$$0 = \sum_{j=1}^{\infty} a_j(v) e_{j+1} = 0 e_1 + \sum_{j=1}^{\infty} a_j(v) e_{j+1}.$$

By Proposition 4.4.23 it follows that $a_j(v) = 0$, $j \in \mathbb{Z}_{>0}$, or that $v = 0$.

We also claim that $\mathrm{image}(\mathsf{L})$ is not dense in V. Indeed, it is clear that the function e_1 is orthogonal to $\mathrm{image}(\mathsf{L})$, which prohibits $\mathrm{image}(\mathsf{L})$ from being dense.

All this shows that L belongs to the class of linear operator described by case of the theorem. • what?

3.5.5 Spectral properties for continuous operators on Banach spaces

3.5.29 Definition Let $(\mathsf{V}, \|\cdot\|)$ be a Banach space with $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ a linear operator for which $\mathrm{dom}(\mathsf{L})$ is dense in V.

(i) $\lambda \in \mathbb{C}$ is in the ***resolvent set*** for $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ if $(\mathsf{L}_\lambda, \mathrm{dom}(\mathsf{L}_\lambda))$ is regular.

(ii) $\lambda \in \mathbb{C}$ is in the ***spectrum*** for $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ if $(\mathsf{L}_\lambda, \mathrm{dom}(\mathsf{L}_\lambda))$ is singular. We denote the spectrum of $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ by $\mathrm{spec}(\mathsf{L})$.

(iii) If $(\mathsf{L}_\lambda, \mathrm{dom}(\mathsf{L}_\lambda))$ is not invertible then λ is an ***eigenvalue*** for $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ and nonzero vectors in $\ker(\mathsf{L}_\lambda)$ are ***eigenvectors*** for $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ corresponding to the eigenvalue λ. The dimension of $\ker(\mathsf{L}_\lambda)$ is the ***multiplicity*** of λ. The collection of eigenvalues is the ***point spectrum*** of $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ which we denote by $\mathrm{spec}_0(\mathsf{L})$.

(iv) If $(\mathsf{L}_\lambda, \mathrm{dom}(\mathsf{L}_\lambda))$ is invertible but

(a) L_λ^{-1} is unbounded,

(b) $\mathrm{image}(\mathsf{L}_\lambda) \subset \mathsf{V}$, and

(c) $\mathrm{cl}(\mathrm{image}(\mathsf{L}_\lambda)) = \mathsf{V}$

then λ is in the ***continuous spectrum*** of $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$. The continuous spectrum of $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ is denoted $\mathrm{spec}_1(\mathsf{L})$.

(v) If $(\mathsf{L}_\lambda, \mathrm{dom}(\mathsf{L}_\lambda))$ is invertible but $\mathrm{cl}(\mathrm{image}(\mathsf{L}_\lambda)) \subset \mathsf{V}$ then λ is in the ***residual spectrum*** of $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$. The residual spectrum of $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ is denoted $\mathrm{spec}_{-1}(\mathsf{L})$. The dimension of $\mathsf{V}/\mathrm{cl}(\mathrm{image}(\mathsf{L}_\lambda))$ is the ***deficiency*** of λ. •

Note that our definition of deficiency in part (v) requires the notion of a quotient V/U of a vector space V by a subspace U. Readers unfamiliar with the notion of a quotient space need not despair since, as we shall shortly see, the linear operators of interest to us have empty residual spectrum. Let us give an example to illustrate our notions of spectrum.

3.5.30 Example On $(\mathsf{V} = \mathsf{L}^2([0,1];\mathbb{F}), \|\cdot\|_2)$ we consider the linear operator $(\mathsf{L}, \operatorname{dom}(\mathsf{L}) = \mathsf{V})$ defined by

$$\mathsf{L}(f)(t) = \int_0^t f(\xi)\,\mathrm{d}\xi. \tag{3.6}$$

To examine the spectrum of L we need to consider the operator $\mathsf{L}_\lambda = \mathsf{L} - \lambda\,\mathrm{id}_{\mathsf{V}}$. First we take $\lambda = 0$ where $\mathsf{L}_\lambda = \mathsf{L}$. We note that $\operatorname{image}(\mathsf{L})$ consists of functions which vanish at $t = 0$ and which possess an L^2-derivative. The collection of all such functions is dense in $\mathsf{L}^2([0,1];\mathbb{F})$. Indeed, note that the differentiable functions vanishing at $t = 0$ are dense in $\operatorname{image}(\mathsf{L})$ by Theorem IV-1.3.11(i). One can also easily see that the differentiable functions vanishing at $t = 0$ are dense in the set of all differentiable functions. Thus $\operatorname{image}(\mathsf{L})$ is dense in V by Theorem IV-1.3.11(i). We claim that $(\mathsf{L}, \operatorname{dom}(\mathsf{L}))$ is invertible and that $(\mathsf{L}^{-1}, \operatorname{image}(\mathsf{L}))$ is unbounded. That L is invertible is clear since

$$\int_0^t f(\xi)\,\mathrm{d}\xi = 0 \quad\Longrightarrow\quad f(\xi) = 0 \text{ a.e.}$$

The unboundedness of L^{-1} follows since if f possess an L^2-derivative and vanishes at $t = 0$ then $\mathsf{L}^{-1}(f) = f'$. We have seen in that this map in unbounded. This shows that $\operatorname{image}(\mathsf{L}_0) \subset \mathsf{V}$, $\operatorname{cl}(\operatorname{image}(\mathsf{L}_0)) = \mathsf{V}$, and that L_0^{-1} is unbounded. Thus $0 \in \operatorname{spec}_1(\mathsf{L})$.

what?

Now we consider $\lambda \neq 0$. Here we claim that $(\mathsf{L}_\lambda, \operatorname{dom}(\mathsf{L}_\lambda))$ is regular. First let us show that it is invertible. Let $\mathsf{L}_\lambda(f) = 0$. Thus

$$\int_0^t f(\xi)\,\mathrm{d}\xi - \lambda f(t) = 0 \quad\Longrightarrow\quad f'(t) - \frac{1}{\lambda} f(t) = 0.$$

The solution to this ordinary differential equation is $f(t) = Ce^{t/\lambda}$. Using the initial condition $f(0) = 0$ we see that $C = 0$, thus showing that L_λ is invertible. Next we show that $\operatorname{image}(\mathsf{L}_\lambda) = \mathsf{V}$. For $f \in \mathsf{V}$ we must find $g \in \operatorname{dom}(\mathsf{L})$ so that $\mathsf{L}_\lambda(g) = f$, or equivalently

$$\int_0^t g(\xi)\,\mathrm{d}\xi - \lambda g(t) = f(t).$$

Define $h(t) = f(t) + \lambda g(t)$ so that

$$h(t) = \int_0^t f(\xi)\,\mathrm{d}\xi.$$

In particular, it follows that h possesses an L^2-derivative and that $h(0) = 0$. Therefore

$$g(t) = h'(r) \quad\Longrightarrow\quad h'(t) - \frac{1}{\lambda} h(t) = -\frac{1}{\lambda} f(t).$$

This equations can now be solved using an integrating factor, as you learned when you were a child, and the solution is

$$h(t) = -\frac{f(t)}{\lambda} - \frac{e^{t/\lambda}}{\lambda} \int_0^t f(\xi)e^{-\xi/\lambda}\,\mathrm{d}\xi,$$

using the fact that $h(0) = 0$. Differentiating this then gives a function g satisfying $\mathsf{L}_\lambda(g) = f$:

$$g(t) = -\frac{f(t)}{\lambda} - \frac{e^{t/\lambda}}{\lambda^2} \int_0^t f(\xi)e^{-\xi/\lambda}\,\mathrm{d}\xi. \tag{3.7}$$

Thus $\mathrm{image}(\mathsf{L}_\lambda) = \mathsf{V}$. Finally, we show that L_λ^{-1} is continuous. We compute, using (3.7),

$$\begin{aligned}|\mathsf{L}_\lambda^{-1}(f)(t)|^2 &= \left|\frac{f(t)}{\lambda} + \frac{e^{t/\lambda}}{\lambda^2}\int_0^t f(\xi)e^{-\xi/\lambda}\,\mathrm{d}\xi\right|^2 \\ &\leq \frac{1}{|\lambda|^2}|f(t)|^2 + \frac{2M}{|\lambda^3|}|f(t)|\left|\int_0^t f(\xi)e^{-\xi/\lambda}\,\mathrm{d}\xi\right| \\ &\quad + \frac{M}{|\lambda|^4}\left|\int_0^t f(\xi)e^{-\xi/\lambda}\,\mathrm{d}\xi\right|^2\end{aligned}$$

where

$$M = \sup_{t\in[0,1]}\{|e^{t/\lambda}|\}.$$

Now we use the Cauchy-Schwarz-Bunyakovsky inequality to further compute

$$\begin{aligned}|\mathsf{L}_\lambda^{-1}(f)(t)|^2 &\leq \frac{1}{|\lambda|^2}|f(t)|^2 + \frac{2M}{|\lambda^3|}|f(t)|\Big(\int_0^t |f(\xi)|^2\,\mathrm{d}\xi\Big)\Big(\int_0^t |e^{-\xi/\lambda}|\,\mathrm{d}\xi\Big) \\ &\quad + \frac{M}{|\lambda|^4}\Big(\int_0^t |f(\xi)|^2\,\mathrm{d}\xi\Big)^2\Big(\int_0^t |e^{-\xi/\lambda}|\,\mathrm{d}\xi\Big)^2 \\ &\leq a|f(t)|^2 + b|f(t)|\|f\| + c\|f\|^2,\end{aligned}$$

where $a, b, c > 0$ are messy constants that are independent of f. Then we compute, again using the Cauchy-Schwarz-Bunyakovsky inequality,

$$\begin{aligned}\|\mathsf{L}_\lambda^{-1}(f)\|^2 &= \int_0^1 |\mathsf{L}_\lambda^{-1}(f)(t)|^2\,\mathrm{d}t \\ &\leq \int_0^1 \Big(a|f(t)|^2 + b|f(t)|\|f\| + c\|f\|^2\Big)\,\mathrm{d}t \\ &= a\|f\|^2 + b\|f\|\Big(\int_0^1 \mathrm{d}t\Big)^{1/2}\Big(\int_0^1 |f(t)|^2\,\mathrm{d}t\Big)^{1/2} + c\|f\|^2 \\ &= (a+b+c)\|f\|_2.\end{aligned}$$

This shows that L_λ^{-1} is continuous for $\lambda \neq 0$.

Thus, all of the above shows the following: On $(\mathsf{L}^2([0,1];\mathbb{F}), \langle\cdot,\cdot\rangle_2)$ the linear L given in (3.6) satisfies

1. $\mathrm{spec}_0(\mathsf{L}) = \varnothing$,
2. $\mathrm{spec}_1(\mathsf{L}) = \{0\}$, and
3. $\mathrm{spec}_{-1}(\mathsf{L}) = \varnothing$.

While some of the computations used to deduce these conclusions may be tedious, they are not essentially difficult. •

3.5.6 The Open Mapping Theorem and Closed Graph Theorem

still two?

how many?

In the preceding two sections we studied some of the more basic characterisations of continuous linear maps between normed vector spaces. In the next sections we give some deeper results which provide some very useful structure for Banach spaces.

3.5.31 Theorem (Banach–Schauder Open Mapping Theorem) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be $\mathbb{F}$-Banach spaces. If $\mathsf{L} \in \mathrm{L}(\mathsf{U};\mathsf{V})$ is surjective then it is open, i.e., $\mathsf{L}(\mathsf{S})$ is open for every open subset $\mathsf{S} \subseteq \mathsf{U}$.*

Proof ■

It is worth reflecting on whether it is necessary that U and V be Banach spaces in order for the result to hold. It turns out that these assumptions are necessary.

3.5.32 Examples (Open Mapping Theorem fails for normed vector spaces)

1. Consider the following data:
   (a) $\mathsf{U} = \mathsf{C}^0([0,1],\mathbb{R})$;
   (b) $\|\cdot\|_{\mathsf{U}} = \|\cdot\|_\infty$;
   (c) V is the subspace of functions f in $\mathsf{C}^0([0,1];\mathbb{R})$ that are continuously differentiable and that satisfy $f(0) = 0$;
   (d) $\|\cdot\|_{\mathsf{V}} = \|\cdot\|_\infty$;
   (e) $\mathsf{L} \in \mathrm{Hom}_{\mathbb{R}}(\mathsf{U};\mathsf{V})$ is defined by

$$\mathsf{L}(f)(x) = \int_0^x f(\xi)\,\mathrm{d}\xi.$$

   Note that V is not complete; we invite the reader to adapt Example 3.6.25–2 to provide a Cauchy sequence in V that does not converge.

   We claim that L is a continuous bijection but its inverse is not continuous.

2. Let $(\mathsf{V}, \|\cdot\|)$ be a Banach space of infinite-dimension and let $\{e_i\}_{i\in I}$ be a basis for V, and suppose without loss of generality that $\|e_i\| = 1$ for each $i \in I$. As in the proof of Proposition 3.1.4 define a norm $\|\cdot\|_1$ on V by

$$\Big\|\sum_{i\in I} c_i e_i\Big\|_1 = \sum_{i\in I} |c_i|,$$

this definition making sense since the sum is finite. As in , $(\mathsf{V}, \|\cdot\|_1)$ is incomplete. what?
We claim that the identity map on V, thought of as a linear map from the normed vector space $(\mathsf{V}, \|\cdot\|_1)$ to the Banach space $(\mathsf{V}, \|\cdot\|)$, is a continuous bijection but has an inverse that is not continuous.

Exercises

3.5.1 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ and $(\mathsf{V}, \|\cdot\|_{\mathsf{V}})$ be normed $\mathbb{F}$-vector spaces, and let $\mathsf{L} \in \mathrm{L}(\mathsf{U}; \mathsf{V})$. Show that L is norm-preserving if and only if it is an isometry of the metric spaces associated with the norms (cf. Proposition 3.1.7).

3.5.2 Show that the operator norm $\|\cdot\|_{\mathbb{R}^m,\mathbb{R}^n}$ defined in Example 3.5.16–1 is not derived from an inner product on $\mathrm{L}(\mathbb{R}^m; \mathbb{R}^n)$.

3.5.3 Prove Proposition 3.5.3.

3.5.4 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. On $\mathbb{F}^n$ consider the two norms $\|\cdot\|_1$ and $\|\cdot\|_2$ as in Example 3.1.3. Show that $\mathrm{id}_{\mathbb{F}^n}$ is a homeomorphism of the normed vector spaces $(\mathbb{F}^n, \|\cdot\|_1)$ and $(\mathbb{F}^n, \|\cdot\|_2)$ but is not an isomorphism of normed vector spaces.

3.5.5 Prove Proposition 3.5.19.

Section 3.6

Topology of normed vector spaces

Since a (semi)normed vector space is a metric space, and so a topological space, one has all of the usual notions associated with topological spaces: interior, closure, boundary, compactness, etc. These notions inherit all of the attributes from general topological spaces as discussed in detail in Chapter 1. We would like, however, for the reader to be able to at least read the results in this section without having first read Chapter 1. Therefore, we adopt the following approach for presentation. All definitions and theorems are stated so that they can be read independently of having read Chapter 1. When it is easily done, proofs are given in a way that does not rely on understanding general notions from topology. However, we also do not shy away from using some general ideas from Chapter 1 in a proof when doing so avoids duplication. The bottom line is this: A reader should be able to understand the flow of ideas without having read Chapter 1, but understanding all proofs may require understanding some parts of Chapter 1.

It is also the case that, like quite a few of the results in this chapter, the statements and proofs bear a strong resemblance to those for real numbers; the reader should thus compare what we say here with what has been said already in Section I-2.5. The similarities and the differences together will help reader understand normed vector spaces.

Say more here

what?

Do I need to read this section? Readers already familiar with topology can forgo the basic definitions and theorems. The notion of a Schauder basis in Section 3.6.5 will come up in . •

3.6.1 Properties of balls in normed vector spaces

In this section we give some fairly easy and pretty "obvious" results concerning the character of open and closed balls in normed vector spaces. These results will be used constantly in our description of the topology of normed vector spaces.

We know that, by definition, the open balls in a normed vector space form a basis for the norm topology; every open set is by definition a union of open balls. This description can be refined a little to show that it is really open balls about 0_V that are important.

3.6.1 Proposition (Balls about the origin are sufficient to describe the norm topology) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(V, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. For any open set $U \subseteq V$ there exists an index set I, positive numbers $(r_i)_{i\in I}$, vectors $(v_i)_{i\in I}$ such that*

$$U = \cup_{i\in I}\{v_i + B(r_i, 0_V)\},$$

where

$$v_i + B(r_i, 0_V) = \{v + v_i \mid v \in B(r_i, 0_V)\}.$$

Proof This follows since $\mathsf{B}(r, v)$ is the translation by v of $\mathsf{B}(r, 0_{\mathsf{V}})$, cf. the proof of Proposition 3.1.12. ■

Let us next give some fairly elementary properties of open and closed balls.

3.6.2 Proposition (Properties of open and closed balls) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $(\mathsf{V}, \|\cdot\|)$ *be a normed* $\mathbb{F}$*-vector space, and let* $\mathrm{r} \in \mathbb{R}_{>0}$ *and* $\mathrm{v}_0 \in \mathsf{V}$*. Then the following statements hold:*

(i) $\mathsf{B}(\mathrm{r}, \mathrm{v}_0)$ *is open;*

(ii) $\overline{\mathsf{B}}(\mathrm{r}, \mathrm{v}_0)$ *is closed and bounded;*

(iii) $\overline{\mathsf{B}}(\mathrm{r}, \mathrm{v}_0)$ *is compact if and only if* $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ *is compact.*

Proof (i) This is Exercise 3.1.1.

(ii) If $M = \|v\| + r$ and if $v \in \overline{\mathsf{B}}(r, v)$ then

$$\|v\| = \|v - v_0 + v_0\| \leq \|v - v_0\| + \|v_0\| \leq M,$$

showing that $\overline{\mathsf{B}}(r, v_0) \subseteq \overline{\mathsf{B}}(M, 0_{\mathsf{V}})$ and so $\overline{\mathsf{B}}(r, v_0)$ is bounded. Define $f\colon \mathsf{V} \to \mathbb{R}$ by $f(v) = \|v\|$ and note that $\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) = f^{-1}([0, 1])$. Since f is continuous by Proposition 3.5.4 and since $[0, 1]$ is closed, it follows that $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ is closed by Proposition 1.3.1. Now define $f_r, f_{v_0}\colon \mathsf{V} \to \mathsf{V}$ by $f_r(v) = rv$ and $f_{v_0}(v) = v + v_0$. By Proposition 3.5.4 these maps are homeomorphisms. Therefore, $f_{v_0} \circ f_r$ is continuous. Since $\overline{\mathsf{B}}(r, v_0) = f_{v_0} \circ f_r(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}))$ and since the homeomorphic image of a closed set is closed (Corollary 1.3.2), it follows that $\overline{\mathsf{B}}(r, v_0)$ is closed.

(iii) As in the preceding part of the proof, $\overline{\mathsf{B}}(r, v_0) = f_{v_0} \circ f_r(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}))$, and since the continuous image of compact sets is compact (Proposition 1.6.5), the result follows. ■

3.6.2 Interior, closure, boundary, etc.

The definitions and results here are similar to those for $\mathbb{R}$ given in Section I-2.5.3, so we will go through them quickly. Examples, discussion, and motivation can be found in Section I-2.5.3.

3.6.3 Definition (Neighbourhood) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. For $v \in \mathsf{V}$, a ***neighbourhood*** of v is an open set U for which $v \in U$. •

3.6.4 Definition (Accumulation point, cluster point, limit point) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. For a subset $A \subseteq \mathsf{V}$, a point $v \in \mathsf{V}$ is:

(i) an ***accumulation point*** for A if, for every neighbourhood U of v, the set $A \cap (U \setminus \{v\})$ is nonempty;

(ii) a ***cluster point*** for A if, for every neighbourhood U of v, the set $A \cap U$ is infinite;

(iii) a ***limit point*** of A if there exists a sequence $(v_j)_{j \in \mathbb{Z}_{>0}}$ in A converging to v.

The set of accumulation points of A is called the ***derived set*** of A, and is denoted by $\mathrm{der}(A)$. •

In Remark I-2.5.12 we made some comments about conventions concerning the words "accumulation point," "cluster point," and "limit point." Those remarks apply equally here.

3.6.5 Proposition ("Accumulation point" equals "cluster point") *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. For a set $A \subseteq \mathsf{V}$, $v \in \mathsf{V}$ is an accumulation point for A if and only if it is a cluster point for A.*

Proof It is clear that a cluster point for A is an accumulation point for A. Suppose that v is not a cluster point. Then there exists a neighbourhood U of v for which the set $A \cap U$ is finite. If $A \cap U = \{v\}$, then clearly v is not an accumulation point. If $A \cap U \neq \{v\}$, then $A \cap (U \setminus \{v\}) \supseteq \{v_1, \dots, v_k\}$ where the points $v_1, \dots, v_k$ are distinct from v. Now let

$$\epsilon = \tfrac{1}{2} \min\{\|v_1 - v\|, \dots, \|v_k - v\|\}.$$

Clearly $A \cap (\mathsf{B}(\epsilon, v) \setminus \{v\})$ is then empty, and so v is not an accumulation point for A. ■

3.6.6 Proposition (Properties of the derived set) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. For $A, B \subseteq \mathsf{V}$ and for a family of subsets $(A_i)_{i \in I}$ of V, the following statements hold:*

(i) $\mathrm{der}(\varnothing) = \varnothing$;

(ii) $\mathrm{der}(\mathsf{V}) = \mathsf{V}$;

(iii) $\mathrm{der}(\mathrm{der}(A)) = \mathrm{der}(A)$;

(iv) *if* $A \subseteq B$ *then* $\mathrm{der}(A) \subseteq \mathrm{der}(B)$;

(v) $\mathrm{der}(A \cup B) = \mathrm{der}(A) \cup \mathrm{der}(B)$;

(vi) $\mathrm{der}(A \cap B) \subseteq \mathrm{der}(A) \cap \mathrm{der}(B)$.

Proof Parts (i) and (ii) follow directly from the definition of the derived set.

True or not? (iii)

(iv) Let $v \in \mathrm{der}(A)$ and let U be a neighbourhood of v. Then the set $A \cap (U \setminus \{v\})$ is nonempty, implying that the set $B \cap (U \setminus \{v\})$ is also nonempty. Thus $v \in \mathrm{der}(B)$.

(v) Let $v \in \mathrm{der}(A \cup B)$ and let U be a neighbourhood of v. Then the set $U \cap ((A \cup B) \setminus \{v\})$ is nonempty. But

$$\begin{aligned} U \cap ((A \cup B) \setminus \{v\}) &= U \cap ((A \setminus \{v\}) \cup (B \setminus \{v\})) \\ &= (U \cap (A \setminus \{v\})) \cup (U \cap (B \setminus \{v\})). \end{aligned} \quad (3.8)$$

Thus it cannot be that both $U \cap (A \setminus \{v\})$ and $U \cap (B \setminus \{v\})$ are empty. Thus x is an element of either $\mathrm{der}(A)$ or $\mathrm{der}(B)$.

Now let $v \in \mathrm{der}(A) \cup \mathrm{der}(A)$. Then, using (3.8), $U \cap ((A \cup B) \setminus \{v\})$ is nonempty, and so $v \in \mathrm{der}(A \cup B)$.

(vi) Let $x \in \mathrm{der}(A \cap B)$ and let U be a neighbourhood of v. Then $U \cap ((A \cap B) \setminus \{v\}) \neq \varnothing$. We have

$$U \cap ((A \cap B) \setminus \{v\}) = U \cap ((A \setminus \{v\}) \cap (B \setminus \{v\}))$$

Thus the sets $U \cap (A \setminus \{v\})$ and $U \cap (B \setminus \{v\})$ are both nonempty, showing that $v \in \mathrm{der}(A) \cap \mathrm{der}(B)$. ■

3.6.7 Definition (Interior, closure, and boundary) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. Let $A \subseteq \mathsf{V}$.

(i) The ***interior*** of A is the set

$$\mathrm{int}(A) = \cup\{U \mid U \subseteq A,\ U \text{ open}\}.$$

(ii) The ***closure*** of A is the set

$$\mathrm{cl}(A) = \cap\{C \mid A \subseteq C,\ C \text{ closed}\}.$$

(iii) The ***boundary*** of A is the set $\mathrm{bd}(A) = \mathrm{cl}(A) \cap \mathrm{cl}(\mathsf{V} \setminus A)$. •

3.6.8 Proposition (Characterisation of interior, closure, and boundary) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \|\cdot\|)$ *be a normed* $\mathbb{F}$*-vector space. For* $A \subseteq \mathsf{V}$*, the following statements hold:*

(i) $v \in \mathrm{int}(A)$ *if and only if there exists a neighbourhood* U *of* v *such that* $U \subseteq A$*;*

(ii) $v \in \mathrm{cl}(A)$ *if and only if, for each neighbourhood* U *of* v*, the set* $U \cap A$ *is nonempty;*

(iii) $v \in \mathrm{bd}(A)$ *if and only if, for each neighbourhood* U *of* v*, the sets* $U \cap A$ *and* $U \cap (\mathsf{V} \setminus A)$ *are nonempty.*

Proof (i) Suppose that $v \in \mathrm{int}(A)$. Since $\mathrm{int}(A)$ is open, there exists a neighbourhood U of v contained in $\mathrm{int}(A)$. Since $\mathrm{int}(A) \subseteq A$, $U \subseteq A$.

Next suppose that $v \notin \mathrm{int}(A)$. Then, by definition of interior, for any open set U for which $U \subseteq A$, $v \notin U$.

(ii) Suppose that there exists a neighbourhood U of v such that $U \cap A = \varnothing$. Then $\mathsf{V} \setminus U$ is a closed set containing A. Thus $\mathrm{cl}(A) \subseteq \mathsf{V} \setminus U$. Since $v \notin \mathsf{V} \setminus U$, it follows that $v \notin \mathrm{cl}(A)$.

Suppose that $v \notin \mathrm{cl}(A)$. Then v is an element of the open set $\mathsf{V} \setminus \mathrm{cl}(A)$. Thus there exists a neighbourhood U of v such that $U \subseteq \mathsf{V} \setminus \mathrm{cl}(A)$. In particular, $U \cap A = \varnothing$.

(iii) This follows directly from part (ii) and the definition of boundary. ■

3.6.9 Proposition (Properties of interior) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \|\cdot\|)$ *be a normed* $\mathbb{F}$*-vector space. For* $A, B \subseteq \mathsf{V}$ *and for a family of subsets* $(A_i)_{i \in I}$ *of* V*, the following statements hold:*

(i) $\mathrm{int}(\varnothing) = \varnothing$*;*

(ii) $\mathrm{int}(\mathsf{V}) = \mathsf{V}$*;*

(iii) $\mathrm{int}(\mathrm{int}(A)) = \mathrm{int}(A)$*;*

(iv) if $A \subseteq B$ *then* $\mathrm{int}(A) \subseteq \mathrm{int}(B)$*;*

(v) $\mathrm{int}(A \cup B) \supseteq \mathrm{int}(A) \cup \mathrm{int}(B)$*;*

(vi) $\mathrm{int}(A \cap B) = \mathrm{int}(A) \cap \mathrm{int}(B)$*;*

(vii) $\mathrm{int}(\cup_{i \in I} A_i) \supseteq \cup_{i \in I} \mathrm{int}(A_i)$*;*

(viii) $\mathrm{int}(\cap_{i \in I} A_i) \subseteq \cap_{i \in I} \mathrm{int}(A_i)$.

Moreover, a set $A \subseteq V$ *is open if and only if* $\mathrm{int}(A) = A$.

Proof Parts (i) and (ii) are clear by definition of interior. Part (v) follows from part (vii), so we will only prove the latter.

(iii) This follows since the interior of an open set is the set itself.

(iv) Let $v \in \mathrm{int}(A)$. Then there exists a neighbourhood U of v such that $U \subseteq A$. Thus $U \subseteq B$, and the result follows from Proposition 3.6.8.

(vi) Let $v \in \mathrm{int}(A) \cap \mathrm{int}(B)$. Since $\mathrm{int}(A) \cap \mathrm{int}(B)$ is open by Exercise I-2.5.1, there exists a neighbourhood U of v such that $U \subseteq \mathrm{int}(A) \cap \mathrm{int}(B)$. Thus $U \subseteq A \cap B$. This shows that $v \in \mathrm{int}(A \cap B)$. This part of the result follows from part (viii).

(vii) Let $v \in \cup_{i \in I} \mathrm{int}(A_i)$. By Exercise I-2.5.1 the set $\cup_{i \in I} \mathrm{int}(A_i)$ is open. Thus there exists a neighbourhood U of v such that $U \subseteq \cup_{i \in I} \mathrm{int}(A_i)$. Thus $U \subseteq \cup_{i \in I} A_i$, from which we conclude that $v \in \mathrm{int}(\cup_{i \in I} A_i)$.

(viii) Let $v \in \mathrm{int}(\cap_{i \in I} A_i)$. Then there exists a neighbourhood U of v such that $U \subseteq \cap_{i \in I} A_i$. It therefore follows that $U \subseteq A_i$ for each $i \in I$, and so that $v \in \mathrm{int}(A_i)$ for each $i \in I$.

The final assertion follows directly from Proposition 3.6.8. ■

3.6.10 Proposition (Properties of closure) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(V, \|\cdot\|)$ *be a normed* $\mathbb{F}$*-vector space. For* $A, B \subseteq V$ *and for a family of subsets* $(A_i)_{i \in I}$ *of* V*, the following statements hold:*

(i) $\mathrm{cl}(\varnothing) = \varnothing$;

(ii) $\mathrm{cl}(V) = V$;

(iii) $\mathrm{cl}(\mathrm{cl}(A)) = \mathrm{cl}(A)$;

(iv) if $A \subseteq B$ *then* $\mathrm{cl}(A) \subseteq \mathrm{cl}(B)$;

(v) $\mathrm{cl}(A \cup B) = \mathrm{cl}(A) \cup \mathrm{cl}(B)$;

(vi) $\mathrm{cl}(A \cap B) \subseteq \mathrm{cl}(A) \cap \mathrm{cl}(B)$;

(vii) $\mathrm{cl}(\cup_{i \in I} A_i) \supseteq \cup_{i \in I} \mathrm{cl}(A_i)$;

(viii) $\mathrm{cl}(\cap_{i \in I} A_i) \subseteq \cap_{i \in I} \mathrm{cl}(A_i)$.

Moreover, a set $A \subseteq V$ *is closed if and only if* $\mathrm{cl}(A) = A$.

Proof Parts (i) and (ii) follow immediately from the definition of closure. Part (vi) follows from part (viii), so we will only prove the latter.

(iii) This follows since the closure of a closed set is the set itself.

(iv) Suppose that $v \in \mathrm{cl}(A)$. Then, for any neighbourhood U of v, the set $U \cap A$ is nonempty, by Proposition 3.6.8. Since $A \subseteq B$, it follows that $U \cap B$ is also nonempty, and so $v \in \mathrm{cl}(B)$.

(v) Let $v \in \mathrm{cl}(A \cup B)$. Then, for any neighbourhood U of v, the set $U \cap (A \cup B)$ is nonempty by Proposition 3.6.8. By Proposition I-1.1.4, $U \cap (A \cup B) = (U \cap A) \cup (U \cap B)$. Thus the sets $U \cap A$ and $U \cap B$ are not both nonempty, and so $v \in \mathrm{cl}(A) \cup \mathrm{cl}(B)$. That $\mathrm{cl}(A) \cup \mathrm{cl}(B) \subseteq \mathrm{cl}(A \cup B)$ follows from part (vii).

(vi) Let $v \in \mathrm{cl}(A \cap B)$. Then, for any neighbourhood U of v, the set $U \cap (A \cap B)$ is nonempty. Thus the sets $U \cap A$ and $U \cap B$ are nonempty, and so $v \in \mathrm{cl}(A) \cap \mathrm{cl}(B)$.

(vii) Let $v \in \cup_{i\in I}\operatorname{cl}(A_i)$ and let U be a neighbourhood of v. Then, for each $i \in I$, $U \cap A_i \neq \emptyset$. Therefore, $\cup_{i\in I}(U \cap A_i) \neq \emptyset$. By Proposition I-1.1.7, $\cup_{i\in I}(U \cap A_i) = U \cap (\cup_{i\in I}A_i)$, showing that $U \cap (\cup_{i\in I}A_i) \neq \emptyset$. Thus $v \in \operatorname{cl}(\cup_{i\in I}A_i)$.

(viii) Let $v \in \operatorname{cl}(\cap_{i\in I}A_i)$ and let U be a neighbourhood of v. Then the set $U \cap (\cap_{i\in I}A_i)$ is nonempty. This means that, for each $i \in I$, the set $U \cap A_i$ is nonempty. Thus $v \in \operatorname{cl}(A_i)$ for each $i \in I$, giving the result. ■

3.6.11 Proposition (Joint properties of interior, closure, boundary, and derived set) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. For $\mathrm{A} \subseteq \mathsf{V}$, the following statements hold:*

(i) $\mathsf{V} \setminus \operatorname{int}(\mathrm{A}) = \operatorname{cl}(\mathsf{V} \setminus \mathrm{A})$*;*

(ii) $\mathsf{V} \setminus \operatorname{cl}(\mathrm{A}) = \operatorname{int}(\mathsf{V} \setminus \mathrm{A})$.

(iii) $\operatorname{cl}(\mathrm{A}) = \mathrm{A} \cup \operatorname{bd}(\mathrm{A})$*;*

(iv) $\operatorname{int}(\mathrm{A}) = \mathrm{A} - \operatorname{bd}(\mathrm{A})$*;*

(v) $\operatorname{cl}(\mathrm{A}) = \operatorname{int}(\mathrm{A}) \cup \operatorname{bd}(\mathrm{A})$*;*

(vi) $\operatorname{cl}(\mathrm{A}) = \mathrm{A} \cup \operatorname{der}(\mathrm{A})$*;*

(vii) $\mathsf{V} = \operatorname{int}(\mathrm{A}) \cup \operatorname{bd}(\mathrm{A}) \cup \operatorname{int}(\mathsf{V} \setminus \mathrm{A})$.

Proof (i) Let $v \in \mathsf{V} \setminus \operatorname{int}(A)$. Since $v \notin \operatorname{int}(A)$, for every neighbourhood U of v it holds that $U \not\subseteq A$. Thus, for any neighbourhood U of v, we have $U \cap (\mathsf{V} \setminus A) \neq \emptyset$, showing that $v \in \operatorname{cl}(\mathsf{V} \setminus A)$.

Now let $v \in \operatorname{cl}(\mathsf{V} \setminus A)$. Then for any neighbourhood U of v we have $U \cap (\mathsf{V} \setminus A) \neq \emptyset$. Thus $v \notin \operatorname{int}(A)$, so $v \in \mathsf{V} \setminus A$.

(ii) The proof here strongly resembles that for part (i), and we encourage the reader to provide the explicit arguments.

(iii) This follows from part (v).

(iv) Clearly $\int(A) \subseteq A$. Suppose that $v \in A \cap \operatorname{bd}(A)$. Then, for any neighbourhood U of v, the set $U \cap (\mathsf{V} \setminus A)$ is nonempty. Therefore, no neighbourhood of v is a subset of A, and so $v \notin \operatorname{int}(A)$. Conversely, if $v \in \operatorname{int}(A)$ then there is a neighbourhood U of v such that $U \subseteq A$. The precludes the set $U \cap (\mathsf{V} \setminus A)$ from being nonempty, and so we must have $v \notin \operatorname{bd}(A)$.

(v) Let $v \in \operatorname{cl}(A)$. For a neighbourhood U of v it then holds that $U \cap A \neq \emptyset$. If there exists a neighbourhood V of v such that $V \subseteq A$, then $v \in \operatorname{int}(A)$. If there exists *no* neighbourhood V of v such that $V \subseteq A$, then for every neighbourhood V of v we have $V \cap (\mathsf{V} \setminus A) \neq \emptyset$, and so $v \in \operatorname{bd}(A)$.

Now let $v \in \operatorname{int}(A) \cup \operatorname{bd}(A)$. If $v \in \operatorname{int}(A)$ then $v \in A$ and so $v \in\subseteq \operatorname{cl}(A)$. If $v \in \operatorname{bd}(A)$ then it follows immediately from Proposition 3.6.8 that $v \in \operatorname{cl}(A)$.

(vi) Let $v \in \operatorname{cl}(A)$. If $v \notin A$ then, for every neighbourhood U of v, $U \cap A = U \cap (A \setminus \{v\}) \neq \emptyset$, and so $v \in \operatorname{der}(A)$.

If $v \in A \cup \operatorname{der}(A)$ then either $v \in A \subseteq \operatorname{cl}(A)$, or $v \notin A$. In this latter case, $v \in \operatorname{der}(A)$ and so the set $U \cap (A \setminus \{v\})$ is nonempty for each neighbourhood U of v, and we again conclude that $v \in \operatorname{cl}(A)$.

(vii) Clearly $\operatorname{int}(A) \cap \operatorname{int}(\mathsf{V} \setminus A) = \emptyset$ since $A \cap (\mathsf{V} \setminus A) = \emptyset$. Now let $v \in \mathsf{V} \setminus (\operatorname{int}(A) \cup \operatorname{int}(\mathsf{V} \setminus A))$. Then, for any neighbourhood U of v, we have $U \not\subseteq A$ and $U \not\subseteq (\mathsf{V} \setminus A)$. Thus the sets $U \cap (\mathsf{V} \setminus A)$ and $U \cap A$ must both be nonempty, from which we conclude that $v \in \mathrm{bd}(A)$. ■

what? Let us close this section with a discussion of some notions not present in Section I-2.5, but which are important for normed vector spaces. General topological versions of these ideas have been discussed in .

3.6.12 Definition (Dense, nowhere dense, separable) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. Let $A, B \subseteq \mathsf{V}$ with $A \subseteq B$.

(i) The set A is ***dense*** in B if $\mathrm{cl}(A) = B$.

(ii) The set A is ***nowhere dense*** in B if $B \setminus \mathrm{cl}(A)$ is dense in B.

(iii) The set A is ***separable*** if there exists a countable dense subset of A. •

what? We refer to for simple examples that illustrate these definitions. Generally speaking, it is not uncommon to see the requirement that a Banach space be separable, although there are important examples of nonseparable Banach spaces, as we shall see in Section 3.8.

3.6.3 Compactness

As we shall shortly see, the discussion of compactness for normed vector spaces has a different flavour than that for compact subsets of $\mathbb{R}$. This is because compactness in infinite-dimensional normed vector spaces is quite a strict notion, for example more strict than closed and bounded. However, the initial definitions proceed just as for $\mathbb{R}$.

We begin with simple definitions concerning open covers.

3.6.13 Definition (Open cover) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space, and let $A \subseteq \mathsf{V}$.

(i) An ***open cover*** for A is a family $(U_i)_{i\in I}$ of open subsets of V having the property that $A \subseteq \cup_{i\in I} U_i$.

(ii) A ***subcover*** of an open cover $(U_i)_{i\in I}$ of A is an open cover $(V_j)_{j\in J}$ of A having the property that $(V_j)_{j\in J} \subseteq (U_i)_{i\in I}$. •

We may now define compactness and other related properties of a subset of a normed vector space.

3.6.14 Definition (Bounded, compact, totally bounded) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. A subset $A \subseteq \mathsf{V}$ is:

(i) ***bounded*** if there exists $M \in \mathbb{R}_{>0}$ such that $A \subseteq \overline{\mathsf{B}}(M, 0)$;

(ii) ***compact*** if every open cover $(U_i)_{i\in I}$ of A possesses a finite subcover;

(iii) ***precompact***[3] if $\mathrm{cl}(A)$ is compact;

[3]What we call "precompact" is very often called "relatively compact." However, we shall use the term "relatively compact" for something different.

(iv) ***totally bounded*** if, for every $\epsilon \in \mathbb{R}_{>0}$ there exists $v_1, \dots, v_k \in \mathsf{V}$ such that $A \subseteq \cup_{j=1}^k \mathsf{B}(\epsilon, v_j)$. •

3.6.15 Theorem (Compactness and dimension) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. Then the following statements are equivalent:*

(i) V is finite-dimensional;

(ii) the closed unit ball $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ is compact;

(iii) a subset $\mathrm{K} \subseteq \mathsf{V}$ is compact if and only if it is closed and bounded;

(iv) V with the norm topology is locally compact.

Proof (i) $\implies$ (ii) By Proposition 3.6.2 $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ is closed and bounded. Now, if $\{e_1, \dots, e_n\}$ is a basis for V, we have a map $\iota\colon \mathsf{V} \to \mathbb{F}^n$ defined by

$$\iota(v_1 e_1 + \cdots + v_n e_n) = (v_1, \dots, e_n)$$

which induces a norm on $\mathbb{F}^n$ (cf. the proof of Proposition 3.1.4), which we also denote by $\|\cdot\|$. Since ι is a homeomorphism of the normed vector spaces $(\mathsf{V}, \|\cdot\|)$ and $(\mathbb{F}^n, \|\cdot\|)$, it follows from the Heine–Borel Theorem that $\overline{\mathsf{B}}(1, 0_{\mathbb{F}^n})$ is compact. Since the image of compact sets under continuous maps is compact (Proposition 1.6.5), we conclude that $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ is compact.

(ii) $\implies$ (iii) Suppose that $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ is compact and let $K \subseteq \mathsf{V}$ be compact. By Proposition 1.6.6 we immediately have that K is closed. Let $\epsilon \in \mathbb{R}_{>0}$ and note that $(\mathsf{B}(\epsilon, v))_{v \in K}$ is an open cover of K. Then there exists a finite subset $v_1, \dots, v_k \in K$ such that

$$K \subseteq \mathsf{B}(\epsilon, v_1) \cup \cdots \cup \mathsf{B}(\epsilon, v_k).$$

We claim that $\cup_{j=1}^k \mathsf{B}(\epsilon, v_k)$ is bounded. Let

$$M = \max\{\|v_j\| \mid j, l \in \{1, \dots, k\}\} + \epsilon.$$

For $j \in \{1, \dots, k\}$ and $v \in \mathsf{B}(\epsilon, v_j)$ we compute

$$\|v\| = \|v - v_j + v_j\| \leq \|v - v_j\| + \|v_j\| \leq M.$$

Thus $\cup_{j=1}^k \mathsf{B}(\epsilon, v_k) \subseteq \overline{\mathsf{B}}(M, 0_{\mathsf{V}})$. Thus K is bounded as well as being closed.

Now suppose that $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ is compact and let $K \subseteq \mathsf{V}$ be closed and bounded. Since K is bounded $K \subseteq \overline{\mathsf{B}}(M, 0_{\mathsf{V}})$ for some $M \in \mathbb{R}_{>0}$. By Proposition 3.6.2, $\overline{\mathsf{B}}(M, 0_{\mathsf{V}})$ is compact. Then K is a closed subset of a compact set, and so is compact by Proposition 1.6.4.

(iii) $\implies$ (iv) Since V is a metric space it is Hausdorff by Proposition 1.8.1. Thus we need only show that $v \in \mathsf{V}$ possesses a precompact neighbourhood. However, for any $\epsilon \in \mathbb{R}_{>0}$, $\mathsf{B}(\epsilon, v)$ is a neighbourhood of v. We claim that $\overline{\mathsf{B}}(\epsilon, 0_{\mathsf{V}})$ is closed and bounded, and so compact by hypothesis. It is clearly bounded since $\overline{\mathsf{B}}(\epsilon, v) \subseteq \overline{\mathsf{B}}(M, 0_{\mathsf{V}})$ where $M = \|v\| + \epsilon$ (why?). It is moreover closed since, as we showed in the first part of the proof, it is the preimage of a closed set under a continuous map.

(iv) $\implies$ (i) Let us first show that, if $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ is compact, then V is finite-dimensional. Note that $(\mathsf{B}(\frac{1}{2}, v))_{v \in \overline{\mathsf{B}}(1,0_{\mathsf{V}})}$ is an open covering of $\overline{\mathsf{B}}(\frac{1}{2}, 0_{\mathsf{V}})$. Therefore, there exists $v_1, \dots, v_k \in \overline{\mathsf{B}}(\frac{1}{2}, 0_{\mathsf{V}})$ such that

$$\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \subseteq \mathsf{B}(\tfrac{1}{2}, v_1) \cup \dots \cup \mathsf{B}(\tfrac{1}{2}, v_k).$$

Let $\mathsf{U} = \operatorname{span}_{\mathbb{R}}(v_1, \dots, v_k)$, which is then a finite-dimensional subspace of V. Since U is complete by Theorem 3.3.3 it is closed by Proposition 1.1.33. We will show that $\mathsf{U} = \mathsf{V}$. Suppose this is not so and let $v_0 \in \mathsf{V} \setminus \mathsf{U}$. Since U is closed, $v_0 \notin \operatorname{cl}(\mathsf{U})$ and so by Proposition 3.6.8 the number

$$r = \inf\{\|u - v_0\| \mid u \in \mathsf{V}\}$$

is in $\mathbb{R}_{>0}$. Let $R \in \mathbb{R}_{>0}$ be such that $\overline{\mathsf{B}}(R, v_0) \cap \mathsf{U} \neq \emptyset$. Then $\overline{\mathsf{B}}(R, v_0) \cap \mathsf{U}$ is closed since it is the intersection of closed sets. The set $\overline{\mathsf{B}}(R, v_0) \cap \mathsf{U}$ is also clearly bounded. Since we have proved that (i) $\implies$ (iii) it follows that $\overline{\mathsf{B}}(R, v_0) \cap \mathsf{U}$ is compact. Define $f \colon \overline{\mathsf{B}}(R, v_0) \cap \mathsf{U} \to \mathbb{R}$ by $f(u) = \|u - v_0\|$. By Proposition 3.5.4 this function is continuous. By Theorem 1.6.7 it follows that f achieves its minimum on $\overline{\mathsf{B}}(R, v_0) \cap \mathsf{U}$. Since $R \geq r$ it follows that there exists $u_0 \in \overline{\mathsf{B}}(R, v_0) \cap \mathsf{U}$ such that $f(u_0 - v_0) = r$. Since $\frac{v_0 - u_0}{\|v_0 - u_0\|} \in \overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ there is some $j \in \{1, \dots, k\}$ such that $\frac{v_0 - u_0}{\|v_0 - u_0\|} \in \mathsf{B}(\frac{1}{2}, v_j)$. Therefore,

$$\left\| \frac{v_0 - u_0}{\|v_0 - u_0\|} - v_j \right\| \leq \frac{1}{2} \quad \implies \quad \left\| v_0 - u_0 - \|v_0 - u_0\| v_j \right\| \leq \tfrac{1}{2} \|v_0 - u_0\| = \tfrac{r}{2}.$$

But we also have $v_0 - u_0 - \|v_0 - u_0\| v_j \in \mathsf{U}$ and so

$$\left\| v_0 - u_0 - \|v_0 - u_0\| v_j \right\| \geq r,$$

giving a contradiction. Thus $\mathsf{U} = \mathsf{V}$ and so compactness of $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ implies finite-dimensionality of V.

Now suppose that V is locally compact. Then there exists a neighbourhood U of 0_{V} for which $\operatorname{cl}(U)$ is compact. Openness of U implies that there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}(\epsilon, 0_{\mathsf{V}}) \subseteq U$. Then $\overline{\mathsf{B}}(\epsilon, 0_{\mathsf{V}})$ is a closed subset of the compact set $\operatorname{cl}(U)$, implying by Proposition 1.6.4 that $\overline{\mathsf{B}}(\epsilon, 0_{\mathsf{V}})$ is compact. By Proposition 3.6.2 it follows that $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ is compact. Our argument above implies that V is finite-dimensional. ■

compact operators

The theorem is rather an important one, given that compact sets have important properties that one often makes use of in applications (see, for example,). Since, in infinite dimensions, one loses the convenient interpretation of compact sets as being equivalent to closed and bounded sets, it then becomes important to understand the nature of the compact sets in a given normed vector space. This can really only be done on a case-by-case basis. For example, we do this in for . The fact that the closed unit ball is not compact in infinite dimensions is also responsible for the sometimes nonintuitive distinctions between finite- and infinite-dimensional normed vector spaces. We shall try to point out specific instances of this as we go along.

where?

what?

3.6.4 Closed subspaces

Subspaces of normed vector spaces are again normed vector spaces by Proposition 3.1.7(iv). It is interesting to know what properties a subspace inherits from the space in which is sits. This is simple in finite-dimensions, but rather more complicated in infinite-dimensions. For reasons that are perhaps not *a priori* clear, closed subspaces play an important rôle in Banach space theory and practice, and for this reason we here study closed subspaces in a little detail.

First we characterise the closed subspaces of a Banach space in a manner completely analogous to Proposition 1.1.33 for metric spaces.

3.6.16 Proposition (Characterisations of closed subspaces) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. For a subspace $\mathsf{U} \subseteq \mathsf{V}$ and with $\|\cdot\|_{\mathsf{U}}$ the restriction of $\|\cdot\|$ to U, the following statements hold:*

(i) if V is a Banach space and if U is closed, then $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ is a Banach space;

(ii) if $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$ is a Banach space then U is closed.

Proof (i) If $(u_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in U, then this is also a Cauchy sequence V. Thus the sequence converges to some $v \in \mathsf{V}$. By Proposition 3.6.8(ii) it follows that $v \in \mathrm{cl}(\mathsf{U}) = \mathsf{U}$, and so U is complete.

(ii) Let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in U converging to $v \in \mathsf{V}$. This is a Cauchy sequence in V and so is also a Cauchy sequence in U, by definition of $\|\cdot\|_{\mathsf{U}}$. Therefore, $v \in \mathsf{U}$ since U is complete. By Proposition 3.6.8(ii) it follows that U is closed. ■

The result has the following useful corollaries. The first is simply a useful rewording of Proposition 3.6.16. But the result is nice, because it says that closed subspaces of Banach spaces are Banach spaces, and so closed subspaces are the proper notion of "subobject" when dealing with Banach spaces.

3.6.17 Corollary (Subspaces of Banach spaces are closed if and only if they are complete) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if $(\mathsf{V}, \|\cdot\|)$ is a $\mathbb{F}$-Banach space, a subspace $\mathsf{U} \subseteq \mathsf{V}$ is closed if and only if it is complete.*

The next corollary provides some insight into how one should view the completion of a normed vector space.

3.6.18 Corollary (The closure of a subspace is a completion of the subspace) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. For a subspace $\mathsf{U} \subseteq \mathsf{V}$ denote by $\|\cdot\|_{\mathsf{U}}$ and $\|\cdot\|_{\mathrm{cl}(\mathsf{U})}$ the restriction of $\|\cdot\|$ to U and $\mathrm{cl}(\mathsf{U})$, respectively. Then $(\mathrm{cl}(\mathsf{U}), \|\cdot\|_{\mathrm{cl}(\mathsf{U})})$ is a completion of $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$.*

Proof It is clear that the inclusion map of U into $\mathrm{cl}(\mathsf{U})$ preserves the norm, i.e., that $\|u\|_{\mathsf{U}} = \|u\|_{\mathrm{cl}(\mathsf{U})}$. Moreover, by Proposition 3.6.8(ii) it follows that, given $v \in \mathrm{cl}(\mathsf{U})$, there exists a sequence $(u_j)_{j\in\mathbb{Z}_{>0}}$ converging to v. Thus $(\mathrm{cl}(\mathsf{U}), \|\cdot\|_{\mathrm{cl}(\mathsf{U})})$ is indeed a completion of $(\mathsf{U}, \|\cdot\|_{\mathsf{U}})$. ■

The next two corollaries concern finite-dimensional cases.

3.6.19 Corollary (Finite-dimensional subspaces are closed) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \|\cdot\|)$ *be a normed* $\mathbb{F}$*-vector space. If* $\mathsf{U} \subseteq \mathsf{V}$ *is a finite-dimensional subspace then* U *is closed.*

Proof By Theorem 3.3.3, U is complete, and so is closed by part (ii) of Proposition 3.6.16. ■

3.6.20 Corollary (Subspaces of finite-dimensional normed vector spaces are closed) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \|\cdot\|)$ *be a finite-dimensional normed* $\mathbb{F}$*-vector space. If* $\mathsf{U} \subseteq \mathsf{V}$ *is subspace then* U *is closed.*

Proof Subspaces of finite-dimensional vector spaces are finite-dimensional, and so closed by Corollary 3.6.19. ■

Let us record the topological properties of the basic subspace operations of sum and intersection. For intersections the story is fairly simple.

3.6.21 Proposition (Intersections of closed subspaces) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, if* $(\mathsf{V}, \|\cdot\|)$ *is a normed* $\mathbb{F}$*-vector space, and if* $(\mathsf{U}_{\mathrm{a}})_{\mathrm{a}\in\mathrm{A}}$ *is a family of closed subspaces of* V*, then* $\cap_{\mathrm{a}\in\mathrm{A}}$ *is a closed subspace of* V*.*

Proof The set $\cap_{a\in A}\mathsf{U}_a$ is a subspace by Proposition I-4.5.34 and is closed by Proposition 1.2.8. ■

For sums the story is significantly more complex. First we give a counterexample to the simplest statement one may wish to make.

3.6.22 Example (The sum of closed subspaces may not be closed) The example we use here begins with the Banach space $\ell^2(\mathbb{F})$ consisting of sequences $(a_j)_{j\in\mathbb{Z}_{>0}}$ for which $\sum_{j=1}^{\infty}|a_j|^2 < \infty$. The norm we use is

$$\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_2 = \Big(\sum_{j=1}^{\infty}|a_j|^2\Big)^{1/2}.$$

In Corollary 3.8.21 we show that this is a Banach space and is, moreover, the completion of $\mathbb{F}_0^\infty$ under the norm $\|\cdot\|_2$. We denote by $(\boldsymbol{e}_j)_{j\in\mathbb{Z}_{>0}}$ the standard basis for $\mathbb{F}_0^\infty$. Thus

$$\boldsymbol{e}_j(k) = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

For the purposes of this example we consider two subspaces of $\ell^2(\mathbb{F})$. We let

$$\begin{aligned} \mathsf{U} &= \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\boldsymbol{e}_{2j-1}|\ \ j \in \mathbb{Z}_{>0})), \\ \mathsf{V} &= \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\boldsymbol{e}_{2j-1} + \tfrac{1}{j}\boldsymbol{e}_{2j}|\ \ j \in \mathbb{Z}_{>0})) \end{aligned}$$

so that U and V are closed subspaces.

Let us establish some facts about these subspaces via a sequence of lemmata.

1 **Lemma** $\mathsf{U} \cap \mathsf{V} = \{0_{\ell^2(\mathbb{F})}\}$.

Proof Let us denote

$$\mathsf{U}' = \operatorname{span}_{\mathbb{F}}(\boldsymbol{e}_{2j-1} |\ j \in \mathbb{Z}_{>0}), \quad \mathsf{V}' = \operatorname{span}_{\mathbb{F}}(\boldsymbol{e}_{2j-1} + \tfrac{1}{j}\boldsymbol{e}_{2j} |\ j \in \mathbb{Z}_{>0}).$$

Let $(a_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{U} \cap \mathsf{V}$. By definition of U and V there exist sequences $((x_{jl})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$ and $((y_{jl})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$ in U' and V', respectively, such that

$$\lim_{l\to\infty}(x_{jl})_{j\in\mathbb{Z}_{>0}} = \lim_{l\to\infty}(y_{jl})_{j\in\mathbb{Z}_{>0}} = (a_j)_{j\in\mathbb{Z}_{>0}}.$$

Since $(x_{jl})_{j\in\mathbb{Z}_{>0}} \in \mathsf{U}'$ for each $l \in \mathbb{Z}_{>0}$ it follows that $a_{2j} = 0$ for $j \in \mathbb{Z}_{>0}$. Therefore, $\lim_{l\to\infty} y_{(2j)l} = 0$ for $j \in \mathbb{Z}_{>0}$. Since $y_{(2j-1)l} = jy_{(2j)l}$ for each $j, l \in \mathbb{Z}_{>0}$ it then follows that $\lim_{l\to\infty} y_{(2j-1)l} = 0$ for each $j \in \mathbb{Z}_{>0}$. Therefore, $a_j = 0$ for each $j \in \mathbb{Z}_{>0}$, giving the lemma. ▼

2 **Lemma** $\operatorname{cl}(\mathsf{U} + \mathsf{V}) = \ell^2(\mathbb{F})$.

Proof Let U' and V' be as in the proof of Lemma 1. We claim that $\mathsf{U}' + \mathsf{V}' = \mathbb{F}_0^\infty$. To see this, let $(a_j)_{j\in\mathbb{Z}_{>0}}$ and write $a_j = x_j + y_j$ where

$$x_j = \begin{cases} a_j - ja_{j+1}, & j \text{ odd}, \\ 0, & j \text{ even}, \end{cases} \qquad y_j = \begin{cases} ja_{j+1}, & j \text{ odd}, \\ a_j, & j \text{ even}. \end{cases}$$

Note that $(x_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{U}'$ and $(y_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{V}'$. Thus $\mathsf{U}' + \mathsf{V}' = \mathbb{F}_0^\infty$, as desired. Therefore,

$$\operatorname{cl}(\mathsf{U}' + \mathsf{V}') = \operatorname{cl}(\mathbb{F}_0^\infty) = \ell^2(\mathbb{F})$$

and so

$$\operatorname{cl}(\mathsf{U}' + \mathsf{V}') \subseteq \operatorname{cl}(\mathsf{U} + \mathsf{V}) = \ell^2(\mathbb{F}),$$

as desired. ▼

3 **Lemma** $\mathsf{U} + \mathsf{V} \subset \ell^2(\mathbb{F})$.

Proof Following the proof of Lemma 1, elements of U and V have the form

$$(x_1, 0, x_2, 0, x_3, 0, \dots), \qquad (y_1, y_1, y_2, \tfrac{1}{2}y_2, y_3, \tfrac{1}{3}y_3, \dots),$$

respectively, where

$$\sum_{j=1}^{\infty} |x_j|^2 < \infty, \quad \sum_{j=1}^{\infty} |y_j|^2 + \sum_{j=1}^{\infty} \frac{|y_j|^2}{j^2} < \infty. \tag{3.9}$$

Thus an element of $\mathsf{U} + \mathsf{V}$ has the form

$$(x_1 + y_1, y_1, x_2 + y_2, \tfrac{1}{2}y_2, x_3 + y_3, \tfrac{1}{3}y_3, \dots), \tag{3.10}$$

where the inequalities (3.9) hold. Now consider the sequence

$$(1, 1, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{3}, \tfrac{1}{3}, \dots) \in \ell^2(\mathbb{F}).$$

We claim that this sequence is not in $\mathsf{U} + \mathsf{V}$. Indeed, suppose that the sequence can be expressed in the form (3.10). Then we must have $x_j + y_j = \frac{1}{j}$ and $\frac{1}{j} y_j = \frac{1}{j}$ for each $j \in \mathbb{Z}_{>0}$. Thus $x_j = \frac{1}{j} - 1$ and $y_j = 1$. The inequalities (3.9) do not hold in this case, so the sequence cannot be in $\mathsf{U} + \mathsf{V}$. ▼

Now we make the following observation. The subspaces U and V are closed and complementary. The sum $\mathsf{U} + \mathsf{V}$ is a strict subspace of $\ell^2(\mathbb{F})$ but is dense in $\ell^2(\mathbb{F})$. Thus $\mathsf{U} + \mathsf{V} \subset \mathrm{cl}(\mathsf{U} + \mathsf{V})$ and so $\mathsf{U} + \mathsf{V}$ is not closed. That is, the sum of closed subspaces need not be closed. •

Now being deprived of access to the nicest result concerning sums of closed subspaces, we must now wonder what *is* true. It turns out that the story here is a little complicated, but it is worth understanding since it actually reveals something interesting about Banach space geometry. So let us spend a few moments understanding this. Suppose that we have a Banach space $(\mathsf{V}, \|\cdot\|)$ with two closed subspaces U_1 and U_2. Then define

$$\delta(\mathsf{U}_1, \mathsf{U}_2) = \sup\{\rho \in [0, 1] \mid \overline{\mathsf{B}}(\rho, 0_{\mathsf{V}}) \cap (\mathsf{U}_1 + \mathsf{U}_2) \subseteq (\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2\},$$

with the convention that if $A, B \subseteq \mathsf{V}$ then

$$A + B = \{u + v \mid u \in A, \ v \in B\}.$$

This is a definition with geometric character so let us examine it in a simple case so that we have a little insight into what it means.

3.6.23 Example ($\delta(\mathsf{U}_1, \mathsf{U}_2)$) Let $\mathsf{V} = \mathbb{R}^2$ and let

$$\mathsf{U}_1 = \mathrm{span}_{\mathbb{R}}((1, 0)), \quad \mathsf{U}_2 = \mathrm{span}_{\mathbb{R}}((1, 1)).$$

We use the standard norm on $\mathbb{R}^2$: $\|(x_1, x_2)\| = \sqrt{x_1^2 + x_2^2}$. The set $(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2$ is depicted on the left in Figure 3.2 and $\overline{\mathsf{B}}(\rho, 0_{\mathsf{V}}) \cap (\mathsf{U}_1 + \mathsf{U}_2)$ is shown on the right. The idea is that $(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2$ is obtained by translating the unit ball in U_1 by all vectors in U_2. Thus one "thickens" U_2 by the unit ball in U_1. Now one take balls of increasing radius in $\mathsf{U}_1 + \mathsf{U}_2$ until the ball is no longer contained in $(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2$. In this example one can see that $\delta(\mathsf{U}_1, \mathsf{U}_2) = 1$. •

In finite dimensions the constructions we give are not so insightful. For example, if V is finite-dimensional then $\delta(\mathsf{U}_1, \mathsf{U}_2) > 0$. However, in infinite dimensions it turns out that $\delta(\mathsf{U}_1, \mathsf{U}_2)$ measures when $\mathsf{U}_1 + \mathsf{U}_2$ is not closed.

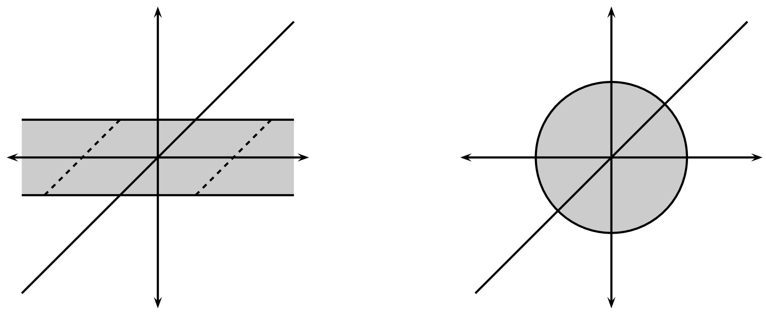


Figure 3.2 The definition of $\delta(\mathsf{U}_1, \mathsf{U}_2)$: $(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2$ on the left and $\overline{\mathsf{B}}(\rho, 0_{\mathsf{V}}) \cap (\mathsf{U}_1 + \mathsf{U}_2)$ on the right

3.6.24 Theorem (When is the sum of closed subspaces closed?) *If* $(\mathsf{V}, \|\cdot\|)$ *is a Banach space and if* U_1 *and* U_2 *are closed subspaces of* V*, then* $\mathsf{U}_1 + \mathsf{U}_2$ *is closed if and only if* $\delta(\mathsf{U}_1, \mathsf{U}_2) > 0$.

Proof Let us define

$$\begin{aligned}\alpha(\mathsf{U}_1, \mathsf{U}_2) = \ & \sup\{\rho \in [0,1] \mid \ \overline{\mathsf{B}}(\rho, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq \mathrm{cl}((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2)\},\\ \beta(\mathsf{U}_1, \mathsf{U}_2) = \ & \sup\{\rho \in [0,1] \mid \ \overline{\mathsf{B}}(\rho, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq (\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2\}.\end{aligned}$$

Both of these quantities are, in fact, equal to $\delta(\mathsf{U}_1, \mathsf{U}_2)$.

1 **Lemma** $\alpha(\mathsf{U}_1, \mathsf{U}_2) = \beta(\mathsf{U}_1, \mathsf{U}_2) = \delta(\mathsf{U}_1, \mathsf{U}_2)$.

Proof Let us abbreviate

$$\alpha = \alpha(\mathsf{U}_1, \mathsf{U}_2), \quad \beta = \beta(\mathsf{U}_1, \mathsf{U}_2), \quad \delta = \delta(\mathsf{U}_1, \mathsf{U}_2).$$

Let us first prove that $\alpha \leq \beta$. This is clearly true if $\alpha = 0$ so suppose that $\alpha > 0$. Since $\mathrm{cl}((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2)$ is closed we have

$$\overline{\mathsf{B}}(\alpha, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq \mathrm{cl}((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2).$$

Note that

$$(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2 = \cap_{r \in (0,1)}((\overline{\mathsf{B}}(\tfrac{1}{1-r}, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2).$$

Therefore, if

$$\overline{\mathsf{B}}(\alpha, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq (\overline{\mathsf{B}}(\tfrac{1}{1-r}, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2 \tag{3.11}$$

for every $r \in (0,1)$ then we have

$$\overline{\mathsf{B}}(\alpha, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq (\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2 \tag{3.12}$$

since $\overline{\mathsf{B}}(\alpha, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2)$ is closed. Moreover, if (3.11) holds then $\alpha \leq \beta$, and so it thus suffices for this part of the proof to show that (3.11) holds for every $r \in (0,1)$. By Proposition 3.6.8 we have

$$\mathrm{cl}((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2) \subseteq ((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2) + \overline{\mathsf{B}}(\alpha r, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2)$$

for every $r \in (0,1)$. By definition of α we then have

$$\overline{\mathsf{B}}(\alpha, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq \mathrm{cl}((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2) \\ \subseteq ((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2) + \overline{\mathsf{B}}(\alpha r, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \quad (3.13)$$

for every $r \in (0,1)$. Let

$$u_0 \in \overline{\mathsf{B}}(\alpha, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2).$$

By (3.13) there exists

$$u_1 \in \overline{\mathsf{B}}(\alpha r, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2), \quad v_0 \in (\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2$$

such that $u_0 = v_0 + u_1$. By definition of α we have

$$\overline{\mathsf{B}}(\alpha r, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq r\,\mathrm{cl}((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2),$$

where, if $a \in \mathbb{F}$ and $A \subseteq \mathsf{V}$, we denote $aA = \{av \mid v \in A\}$. Thus, again by (3.13), there exists

$$u_2 \in \overline{\mathsf{B}}(\alpha r^2, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2), \quad v_1 \in (\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2$$

such that $u_1 = rv_1 + u_2$. Continuing in this way, there exist sequences $(u_j)_{j\in\mathbb{Z}_{\geq 0}}$ and $(v_j)_{j\in\mathbb{Z}_{\geq 0}}$ such that

1. $u_j \in \overline{\mathsf{B}}(\alpha r^j, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq r^j\,\mathrm{cl}((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2)$,
2. $v_j \in (\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2$, and
3. $u_j = r^j v_j + v_{j+1}$

for each $j \in \mathbb{Z}_{>0}$. Clearly, then, $\lim_{j\to\infty} u_j = 0_{\mathsf{V}}$. Therefore,

$$u_0 - u_k = \sum_{j=0}^{k} q^j v_j \quad \Longrightarrow \quad u_0 = \lim_{k\to\infty}(u_0 - u_k) = \sum_{j=0}^{\infty} r^j v_j.$$

Also,

$$\|v_j\| = r^{-j}\|u_j - u_{j+1}\| \leq r^{-j}(\|u_j\| + u_{j+1}) \leq r^{-j}(\alpha r^j + \alpha r^{j+1}) = \alpha(1 + r).$$

Thus the sequence $(v_j)_{j\in\mathbb{Z}_{\geq 0}}$ is bounded. Now, for each $j \in \mathbb{Z}_{\geq 0}$ define $w_j \in \overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1$ and $z_j \in \mathsf{U}_2$ such that $v_j = w_j + z_j$. Then we have

$$\|z_j\| = \|v_j - w_j\| \leq \|v_j\| + \|w_j\| \leq \alpha(1 + r) + 1,$$

and so the sequence $(z_j)_{j\in\mathbb{Z}_{\geq 0}}$ is bounded. Therefore,

$$\Big\|\sum_{j=0}^{\infty} r^j w_j\Big\| \leq \sum_{j=0}^{\infty} r^j \|w_j\| \leq \frac{1}{1-r} \quad \Longrightarrow \quad \sum_{j=0}^{\infty} r^j w_j \in \overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1$$

since $\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1$ is closed. Similarly,

$$\Big\|\sum_{j=0}^{\infty} r^j z_j\Big\| \leq \sum_{j=0}^{\infty} r^j \|z_j\| \leq \frac{1}{1-r}(\alpha(1 + r) + 1) \quad \Longrightarrow \quad \sum_{j=0}^{\infty} r^j z_j \in \mathsf{U}_2$$

since U_2 is closed. Thus

$$u_0 = \sum_{j=0}^{\infty} r^j v_j = \sum_{j=0}^{\infty} r^j w_j + \sum_{j=0}^{\infty} r^j z_j \in \overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1 + \mathsf{U}_2.$$

Since u_0 was chosen arbitrarily from $\overline{\mathsf{B}}(\alpha, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2)$ and since the argument can be made for every $r \in (0, 1)$, we have shown that (3.11) holds for every $r \in (0, 1)$, giving $\alpha \leq \beta$.

That $\beta \leq \delta$ follows directly from the definitions.

To show that $\delta \leq \alpha$, and so to complete the proof, it suffices to show that

$$\mathrm{cl}(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap (\mathsf{U}_1 + \mathsf{U}_2)) = \overline{\mathsf{B}}(1, 0_{\mathsf{V}})\, \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2)$$

(why?). To show this, let $v \in \overline{\mathsf{B}}(1, 0_{\mathsf{V}})\, \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2)$. If $v = 0_{\mathsf{V}}$ then we obviously have $v \in \mathrm{cl}(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap (\mathsf{U}_1 + \mathsf{U}_2))$. Thus we can suppose that $v \neq 0_{\mathsf{V}}$. Let $(v_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in $\mathsf{U}_1 + \mathsf{U}_2$ converging to v. We can without loss of generality suppose that $v_j \neq 0_{\mathsf{V}}$ for each $j \in \mathbb{Z}_{>0}$. Then define $u_j = \frac{\|v\|}{\|v_j\|} v_j$ for each $j \in \mathbb{Z}_{>0}$, noting that $u_j \in \overline{\mathsf{B}}(1, 0_{\mathsf{V}})\, \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2)$. Moreover, $\lim_{j \to \infty} u_j = v$ and so $v \in \mathrm{cl}(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap (\mathsf{U}_1 + \mathsf{U}_2))$. This gives

$$\overline{\mathsf{B}}(1, 0_{\mathsf{V}})\, \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq \mathrm{cl}(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap (\mathsf{U}_1 + \mathsf{U}_2)).$$

By Proposition 3.6.10 we have

$$\mathrm{cl}(\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap (\mathsf{U}_1 + \mathsf{U}_2)) \subseteq \overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2).$$

This gives $\delta < \alpha$ be the definitions. ▼

Carrying on with the proof of the theorem, first suppose that $\mathsf{U}_1 + \mathsf{U}_2$ is closed. By Corollary 3.6.17 it follows that $\mathsf{U}_1 + \mathsf{U}_2$ is complete. We obviously have

$$\mathsf{U}_1 + \mathsf{U}_2 = \cup_{j=1}^{\infty} j((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2).$$

Therefore, by the Baire Category Theorem there exists at least one $j \in \mathbb{Z}_{>0}$ for which a corollary of it?

$$\mathrm{int}(\mathrm{cl}(j((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2))) \neq \emptyset.$$

Thus there exist $v \in \mathrm{cl}((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2)$ and $r \in \mathbb{R}_{>0}$ such that

$$\overline{\mathsf{B}}(r, v) \cap (\mathsf{U}_1 + \mathsf{U}_2) \subseteq j\, \mathrm{cl}((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2).$$

Therefore,

$$\overline{\mathsf{B}}(\tfrac{r}{j}, v) \cap (\mathsf{U}_1 + \mathsf{U}_2) \subseteq \mathrm{cl}((\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2),$$

giving $\alpha(\mathsf{U}_1, \mathsf{U}_1) > 0$ and so, by the lemma, $\delta(\mathsf{U}_1, \mathsf{U}_2) > 0$.

Conversely, suppose that $\delta(\mathsf{U}_1, \mathsf{U}_2) > 0$ and so, by the lemma, $\beta(\mathsf{U}_1, \mathsf{U}_2) > 0$. Let $\beta \in (0, \beta(\mathsf{U}_1, \mathsf{U}_2))$. We obviously have

$$\mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) = \cup_{j=1}^{\infty} j(\overline{\mathsf{B}}(\beta, 0_{\mathsf{V}})) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2).$$

By definition of $\beta(\mathsf{U}_1, \mathsf{U}_2)$ it holds that

$$\overline{\mathsf{B}}(\beta, 0_{\mathsf{V}}) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq (\overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U}_1) + \mathsf{U}_2.$$

Moreover, we then obviously have

$$\cup_{j=1}^{\infty} j(\overline{\mathsf{B}}(\beta, 0_{\mathsf{V}})) \cap \mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq \mathsf{U}_1 + \mathsf{U}_2.$$

This gives $\mathrm{cl}(\mathsf{U}_1 + \mathsf{U}_2) \subseteq \mathsf{U}_1 + \mathsf{U}_2$ which gives $\mathsf{U}_1 + \mathsf{U}_2$ as being closed, as desired. ■

Let us close our discussion of closed subspaces by considering some examples of closed and non-closed subspaces of normed vector spaces.

3.6.25 Examples (Closed subspace) Both examples we consider are subspaces of the normed vector space $(\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{R}), \|\cdot\|_\infty)$. By $\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$ we denote the set of bounded, continuous $\mathbb{R}$-valued functions on $\mathbb{R}$ and the norm $\|\cdot\|_\infty$ is defined thusly:

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in \mathbb{R}\}.$$

Note that convergence in the norm $\|\cdot\|_\infty$ is, by definition, uniform convergence. In Theorem I-3.6.8 we essentially showed that $(\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{R}), \|\cdot\|_\infty)$ is a Banach space, although we shall revisit this in Section 3.8.4.

1. Let $\mathsf{C}^0_0(\mathbb{R};\mathbb{R})$ be the subset of $\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$ consisting of those functions satisfying

$$\lim_{x\to-\infty} f(x) = 0, \quad \lim_{x\to\infty} f(x) = 0.$$

It is easy to verify (cf. Proposition I-2.3.23) that $\mathsf{C}^0_0(\mathbb{R};\mathbb{R})$ is a subspace of $\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$. We claim that it is a closed subspace. To show this, it suffices to show that, if $(f_j)_{j\in\mathbb{Z}_{>0}}$ is any sequence in $\mathsf{C}^0_0(\mathbb{R};\mathbb{R})$ converging in $\mathsf{C}_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$, then the limit function is in $\mathsf{C}^0_0(\mathbb{R};\mathbb{R})$. We shall prove this below as Theorem 3.8.40, but it is not too hard to imagine why it is true. Uniform convergence requires that the limit function be approximated uniformly over all of $\mathbb{R}$ by sufficiently large terms in the sequence. Since all functions in the sequence tend to zero at infinity, they will pull the limit function down to zero with them.

2. Let $\mathsf{C}^1_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$ be the subset of $\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$ consisting of those functions that are continuously differentiable. By Proposition I-3.2.10 it follows that $\mathsf{C}^1_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$ is a subspace of $\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$. We claim that it is not closed. To see this, define a sequence of functions $(f_j)_{j\in\mathbb{Z}_{>0}}$ as follows:

$$f_j(x) = \begin{cases} -1, & x \in (-\infty, -1-\frac{1}{j}), \\ \frac{1}{4}jx^2 + \frac{1}{2}(j+1)x + \frac{(j-1)^2}{4j}, & x \in [-1-\frac{1}{j}, -1+\frac{1}{j}], \\ x, & x \in (-1+\frac{1}{j}, 1-\frac{1}{j}), \\ -\frac{1}{4}jx^2 + \frac{1}{2}(j+1)x - \frac{(j-1)^2}{4j}, & x \in [1-\frac{1}{j}, 1+\frac{1}{j}], \\ 1, & x \in (1+\frac{1}{j}, \infty). \end{cases}$$

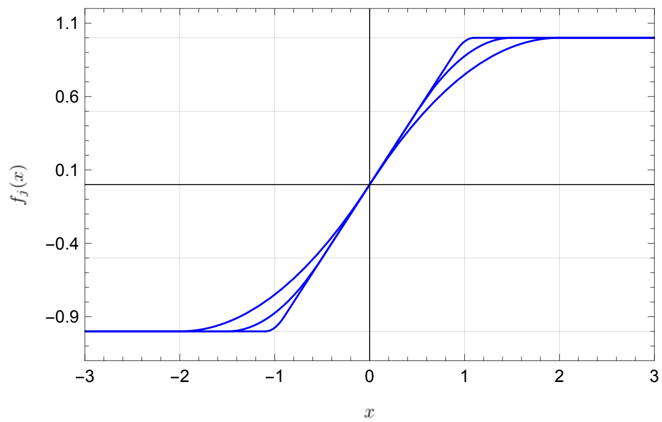


Figure 3.3 A sequence in $C^1_{bdd}(\mathbb{R};\mathbb{R})$ not converging in $C^1_{bdd}(\mathbb{R};\mathbb{R})$ (the terms f_1, f_2, and f_{10} are shown)

We depict this sequence in Figure 3.3. One can show by direct computation that f_j is differentiable for each $j \in \mathbb{Z}_{>0}$; one need only check that the left and right limits for the function and its derivative match at the points $-1-\frac{1}{j}$, $-1+\frac{1}{j}$, $1-\frac{1}{j}$, and $1+\frac{1}{j}$. A direct computation also shows that the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges pointwise to the function

$$f(x) = \begin{cases} -1, & x \in (-\infty,-1), \\ x, & x \in [-1,1], \\ 1, & x \in (1,\infty). \end{cases}$$

To show that $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges to f in $C^0_{bdd}(\mathbb{R};\mathbb{R})$ we need to show that the convergence is uniform.

This is sort of "obvious" from Figure 3.3, but let us go through the details anyway. The only possible problems can occur on the intervals $[-1-\frac{1}{j}, -1+\frac{1}{j}]$ and $[1-\frac{1}{j}, 1+\frac{1}{j}]$ since off these intervals f_j agrees with f. So let $\epsilon \in \mathbb{R}_{>0}$ and let N be sufficiently large that $\frac{1}{4N} < \epsilon$. On $[1-\frac{1}{j}, 1]$ the maximum deviation of f_j from f will occur at $x = 1$. Thus, for $x \in [1-\frac{1}{j}, 1]$ we have

$$\begin{aligned} |f_j(x) - f(x)| &\leq |f_j(1) - f(1)| \\ &= \left|-\tfrac{1}{4}j + \tfrac{1}{2}(j+1) - \tfrac{(j-1)^2}{4j} - 1\right| = \left|\tfrac{1}{4j}\right| < \epsilon \end{aligned}$$

for $j \geq N$. Similarly, on $[1, 1+\frac{1}{j}]$ the maximum deviation of f_j from f will occur at $x = 1$, and the same computation gives $|f_j(x) - f(x)| < \epsilon$ for $x \in [1, 1+\frac{1}{j}]$ for $j \geq N$. This gives $|f_j(x) - f(x)| < \epsilon$ for $x \in [1-\frac{1}{j}, 1+\frac{1}{j}]$. An entirely similar argument gives $|f_j(x) - f(x)| < \epsilon$ for $x \in [-1-\frac{1}{j}, -1+\frac{1}{j}]$.

The point is that the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{C}^1_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$ converges to $f \in \mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$. But since $f \notin \mathsf{C}^1_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$, it follows that $\mathsf{C}^1_{\mathrm{bdd}}(\mathbb{R};\mathbb{R})$ is not closed. The reason for this is fairly evident. The norm $\|\cdot\|_\infty$ does not know anything about the derivative of a function, and so it cannot be expected that the sequence of derivatives will converge to the derivative of the limit function, nor even that the limit function will indeed be even differentiable. •

3.6.5 Bases for normed vector spaces

In Section I-4.5.4 we discussed at length the notion of a basis for a vector space, sometimes called a Hamel basis. The fact that every vector space possesses a Hamel basis is of great use in algebra, but not great value in analysis. To exhibit the limitations of the effectiveness of Hamel bases, let us prove that certain vector spaces are incapable of supporting a norm for which the resulting normed vector space is complete (thus we are supposing here familiarity with completeness, a notion we discuss in detail in Section 3.3).

3.6.26 Theorem (Dimension of an infinite-dimensional Banach space) *If* $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ *and if* $(\mathsf{V},\langle\cdot,\cdot\rangle)$ *is an infinite-dimensional* $\mathbb{F}$*-Banach space, then* $\dim_{\mathbb{F}}(\mathsf{V}) \geq \mathrm{card}(\mathbb{R})$*. If* V *is separable then* $\dim_{\mathbb{F}}(\mathsf{V}) = \mathrm{card}(\mathbb{R})$*.*

Proof Let $v_1 \in \mathsf{V}\setminus\{0_{\mathsf{V}}\}$ be such that $\|v_1\| = 1$. Define $\hat{\alpha}_1\colon \mathrm{span}_{\mathbb{F}}(v_1) \to \mathbb{F}$ by $\hat{\alpha}_1(av_1) = a$. It is trivial to check that $\hat{\alpha}_1$ is a continuous linear function satisfying $\hat{\alpha}_1(v_1) = 1$. By the Hahn–Banach Theorem, Theorem 3.9.2, there exists $\alpha_1 \in \mathsf{V}^*$ such that $\alpha_1(v_1) = 1$. Next consider the closed subspace $\mathsf{V}_2 = \ker(\alpha_1)$ and let $v_2 \in \mathsf{V}_2$ so that $\alpha_1(v_2) = 0$. Also suppose that $\|v_2\| = 1$. Then define $\hat{\alpha}_2\colon \mathrm{span}_{\mathbb{F}}(v_1,v_2) \to \mathbb{F}$ by $\hat{\alpha}_2(a_1v_1 + a_2v_2) = a_2$. As above, use the Hahn–Banach Theorem to deduce the existence of $\alpha_2 \in \mathsf{V}^*$ such that $\alpha_2(a_1v_2 + a_2v_2) = a_2$ for every $a_1, a_2 \in \mathbb{F}$. We may continue inductively in this way to define sequences $(v_j)_{j\in\mathbb{Z}_{>0}}$ and $(\alpha_j)_{j\in\mathbb{Z}_{>0}}$ such that $\|v_j\| = 1$, $j \in \mathbb{Z}_{>0}$, and such that

$$\alpha_j(v_k) = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

We claim that the family $(v_j)_{j\in\mathbb{Z}_{>0}}$ is linearly independent. Indeed, suppose that

$$c_1v_{j_1} + \cdots + c_kv_{j_k} = 0$$

for some $c_1,\dots,c_k \in \mathbb{F}$ and $j_1,\dots,j_k \in \mathbb{Z}_{>0}$. For each $l \in \{1,\dots,k\}$, apply α_{j_l} to the preceding equality to get $c_l = 0$. This give the desired linear independence. We also claim that

$$v_k \notin \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(v_j|\ j \neq k))$$

for each $k \in \mathbb{Z}_{>0}$. Indeed, if $(w_l)_{l\in\mathbb{Z}_{>0}}$ is a convergent sequence in $\mathrm{span}_{\mathbb{F}}(v_j|\ j \neq k)$ then $\alpha_k(w_l) = 0$ for all $l \in \mathbb{Z}_{>0}$. Continuity of α_k and Theorem 3.5.2 ensure that

$$\alpha_k(\lim_{l\to\infty} w_l) = \lim_{l\to\infty} \alpha_k(w_l) = 0.$$

Thus $\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(v_j|\ j \neq k)) \subseteq \ker(\alpha_k)$. Since $\alpha_k(v_k) = 1$ our claim follows.

Now we use a lemma.

1 Lemma *If* S *is a countably infinite set then there exists a family* $(A_t)_{t\in[0,1]}$ *of infinite subsets of* S *such that* $A_{t_1} \cap A_{t_2}$ *is finite for* $t_1 \neq t_2$*.*

Proof For $\theta \in [0, \pi)$ denote

$$\Sigma_\theta = \{(x\cos\theta - y\sin\theta, x\sin\theta + y\cos\theta) \in \mathbb{R}^2 \mid x \in \mathbb{R},\ y \in [-1, 1]\}.$$

Thus Σ_θ is a bi-infinite strip of width 2 inclined at an angle θ to the x-axis in $\mathbb{R}^2$. For $\theta \in [0, \pi)$ define

$$\hat{A}_\theta = \{(x, y) \in \mathbb{Z}^2 \subseteq \mathbb{R}^2 \mid (x, y) \in \Sigma_\theta\}$$

as the points in $\mathbb{Z}^2$ lying in Σ_θ. Some elementary geometry can be used to verify the fact that if $\theta_1 \neq \theta_2$ then $\Sigma_{\theta_1} \cap \Sigma_{\theta_2}$ is compact. From this fact it follows that $\hat{A}_{\theta_1} \cap \hat{A}_{\theta_2}$ is finite for $\theta_1 \neq \theta_2$. Moreover, one can verify that $\hat{A}_\theta$ is infinite for every θ. To see this note that every ball of the form $\overline{\mathsf{B}}(1, (r\cos\theta, r\sin\theta))$ must contain a point with integer coordinates.

Since S and $\mathbb{Z}^2$ are both countable there exists a bijection $\phi\colon S \to \mathbb{Z}^2$. Since $[0, 1]$ and $[0, \pi)$ both have the cardinality of $\mathbb{R}$ (why?), there exists a bijection $\psi\colon [0, 1] \to [0, \pi)$. Then, for $t \in [0, 1]$, define

$$A_t = \{s \in S \mid \phi(s) \in \hat{A}_{\psi(t)}\}.$$

It then follows that A_t is infinite since $\hat{A}_{\psi(t)}$ is infinite and that $A_{t_1} \cap A_{t_2}$ is finite since $\hat{A}_{\psi(t_1)} \cap \hat{A}_{\psi(t_2)}$ is finite. ▼

Now, using the lemma, let $(A_t)_{t\in[0,1]}$ be a family of subsets of $\mathbb{Z}_{>0}$ such that $A_{t_1} \cap A_{t_2}$ is finite for $t_1 \neq t_2$. Then define

$$u_t = \sum_{j\in A_t} \frac{v_j}{2^j}, \qquad t \in [0, 1].$$

Note that

$$\|u_t\| = \Big\|\sum_{j\in A_t} \frac{v_j}{2^j}\Big\| \leq \sum_{j\in A_t} \frac{\|v_j\|}{2^j} < \infty$$

by Example I-2.4.2–4. Thus the series for u_t is absolutely convergent and so convergent by Theorem 3.4.6. We claim that the set $\{u_t\}_{t\in[0,1]}$ is linearly independent. For $l \in \{1, \dots, k\}$ and $m \in \mathbb{Z}_{>0}$ we have

$$\alpha_m(u_{t_l}) = \alpha_m\Big(\sum_{j\in A_t} \frac{v_j}{2^j}\Big) = \sum_{j\in A_t} \frac{\alpha_m(v_j)}{2^j},$$

using Theorem 3.5.2. Thus

$$\alpha_m(u_{t_l}) = \begin{cases} 2^{-m}, & m \in A_{t_l}, \\ 0, & m \notin A_{t_l}. \end{cases}$$

Now suppose that

$$c_1 u_{t_1} + \dots + c_k u_{t_k} = 0 \tag{3.14}$$

for $c_1,\dots,c_k \in \mathbb{F}$ and $t_1,\dots,t_k \in [0,1]$. Without loss of generality we may suppose that the numbers $t_1,\dots,t_k$ are distinct. Then $\cap_{l=1}^k A_{t_l}$ is finite; let us denote it by $\{m_1,\dots,m_r\}$. For $l \in \{1,\dots,k\}$ define $A'_l = A_{t_l} \setminus \{m_1,\dots,m_r\}$, noting that the sets A'_l, $l \in \{1,\dots,k\}$, are enumerable and disjoint. We can then rewrite (3.14) as

$$a_1 v_{m_1} + \cdots + a_r v_{m_r} + c_1 \sum_{j_1 \in A'_1} \frac{v_{j_1}}{2^{j_1}} + \cdots + c_k \sum_{j_k \in A'_k} \frac{v_{j_k}}{2^{j_k}}$$

for suitable constants $a_1,\dots,a_r \in \mathbb{F}$ that depend on the coefficients $c_1,\dots,c_k$ and factors of $\frac{1}{2}$; the precise form of these is immaterial to our computations. Indeed, for each $l \in \{1,\dots,k\}$ let $m_l \in A'_l$. Then, by the properties for $(v_j)_{j\in\mathbb{Z}_{>0}}$ and $(\alpha_j)_{j\in\mathbb{Z}_{>0}}$ given before the lemma,

$$0 = \alpha_{m_l}(c_1 u_{t_1} + \cdots + c_k u_{t_k}) = \frac{c_l}{2^{m_l}}.$$

Thus $c_l = 0$ for each $l \in \{1,\dots,k\}$, giving linear independence of $\{u_t\}_{t\in[0,1]}$. Since $\operatorname{card}([0,1]) = \operatorname{card}(\mathbb{R})$ the first assertion of the theorem follows.

For the final assertion of the theorem we shall prove that $\operatorname{card}(\mathsf{V}) = \operatorname{card}(\mathbb{R})$ if V is separable. It is clear that $\operatorname{card}(\mathsf{V}) \geq \operatorname{card}(\mathbb{R})$. For the opposite inequality, let $D \subseteq \mathsf{V}$ be a countable dense subset of V. For $v \in \mathsf{V}$ we can write $v = \lim_{j\to\infty} v_j$ for a sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ in D. Thus to every point in V we assign a sequence in the countable set D. The set of such sequences is $D^{\mathbb{Z}_{>0}}$, and so $\operatorname{card}(\mathsf{V}) \leq \operatorname{card}(D^{\mathbb{Z}_{>0}}) = \aleph_0^{\aleph_0}$. Now note that $2 \leq \aleph_0 \leq 2^{\aleph_0}$ by Example I-1.7.14–3 and Exercise I-1.7.4. Thus

$$2^{\aleph_0} \leq \aleph_0^{\aleph_0} \leq (2^{\aleph_0})^{\aleph_0} = 2^{\aleph_0\cdot\aleph_0} = 2^{\aleph_0}$$

by Theorem I-1.7.17. Thus $\aleph_0^{\aleph_0} = 2^{\aleph_0}$ and so $\operatorname{card}(\mathsf{V}) \leq 2^{\aleph_0} = \operatorname{card}(\mathbb{R})$ by Exercise I-1.7.5. ■

3.6.27 Corollary (There are no Banach spaces of countable dimension) *If $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ and if V is an $\mathbb{F}$-vector space with an enumerable Hamel basis, then there is no norm on V for which the resulting normed $\mathbb{F}$-vector space is complete.*

applied_analysis.ps.gz
for Schauder basis C0

3.6.6 Notes

Our approach to characterising the closedness of sums of closed subspaces follows Mennicken and Sagraloff [1979], who base their presentation on that of Kailath [1980]. Note that we also used this characterisation of sums of closed subspaces in our proofs of the Open Mapping Theorem and the Closed Graph Theorem. This idea is included in the paper of Mennicken and Sagraloff [1979].

The proof we give for Theorem 3.6.26 is due to Lacey [1973]. The proof of the lemma used in the proof of the theorem is from [Buddenhagen 1971]. An elementary proof of Corollary 3.6.27 can be found in [Bauer and Benner 1971].

Exercises

3.6.1 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a finite-dimensional normed $\mathbb{F}$-vector space. Show that a subspace $\mathsf{U} \subseteq \mathsf{V}$ is dense in V if and only if $\mathsf{U} = \mathsf{V}$. Point out which parts of your argument are not generally valid when V is infinite-dimensional.

3.6.2 Let $(\mathsf{V}, \|\cdot\|)$ be a normed vector space and let $A, B, C \subseteq \mathsf{V}$ be subsets with $A \subseteq B \subseteq C$. Show that if A is dense in B and if B is dense in C then A is dense in C.

3.6.3 Consider Example 3.6.22. On the subspace U (resp. V) denote the restriction of $\|\cdot\|_2$ by $\|\cdot\|_{\mathsf{U}}$ (resp. $\|\cdot\|_{\mathsf{V}}$). By Proposition 3.3.4 the normed vector space $\mathsf{U}_1 \oplus \mathsf{U}_2$ is complete. But in Example 3.6.22 we showed that $\mathsf{U}_1 \oplus \mathsf{U}_2$ is not a closed subspace of $\ell^2(\mathbb{F})$ and so is not complete by Corollary 3.6.17.

Why are these conclusions not in contradiction?

3.6.4 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. If $\mathsf{U} \subseteq \mathsf{V}$ is a subspace, show that $\mathrm{cl}(\mathsf{U})$ is a subspace.

Section 3.7

Split subspaces?

In this section we examine the problem of when all closed subspaces of a Banach space are split. This is an interesting and natural question, and to examine the question appropriately requires some effort and leads one to a deeper understanding of the geometry of Banach spaces. In particular, we shall delve into the topics of concentration of Gaussian measure and its relationship to the famous Dvoretzky Theorem.

Do I need to read this section? The main result we state in this section is interesting and worth knowing. It is stated as Theorem 3.7.5 below. The remainder of the material can be thought of as optional. •

3.7.1 Split subspaces

Now understanding, or at least pretending to understand, closed subspaces of normed vector spaces, let us examine when such a subspace possesses a closed complement. As with closed subspaces, it is perhaps not clear why one wishes to devote a lot of energy to this situation. Hopefully this will become clear as we go along.

3.7.1 **Definition (Split subspace)** Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed vector space. A subspace $\mathsf{U} \subseteq \mathsf{V}$ is ***split*** if there exists a closed subspace W of V such that $\mathsf{V} = \mathsf{U} \oplus \mathsf{W}$. •

Note that the immediate question that comes to mind in the definition of a split subspace is whether, if one complement is closed, all complements are closed. We resolve this as follows.

3.7.2 **Proposition (Characterisation of split subspaces)** *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. For a closed subspace $\mathsf{U} \subseteq \mathsf{V}$ the following statements are equivalent:*

(i) U *is split;*

(ii) *if* W *is a complement to* U *in* V *then* W *is closed;*

(iii) V/U *is a Banach space.*

Proof We first prove a lemma.

1 **Lemma** *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space, and let U be a closed subspace with $\|\cdot\|_{/\mathsf{U}}$ the norm on V/U as in Proposition 3.1.19. Let $\pi_{\mathsf{U}}\colon \mathsf{V} \to \mathsf{V}/\mathsf{U}$ be the canonical projection. If W is an (algebraic) complement to U and if $\|\cdot\|_{\mathsf{W}}$ denotes the restriction of $\|\cdot\|$ to W, then $\pi_{\mathsf{U}}|\mathsf{W}$ is a Banach space isomorphism from $(\mathsf{W}, \|\cdot\|_{\mathsf{W}})$ to $(\mathsf{V}/\mathsf{U}, \|\cdot\|_{/\mathsf{U}})$.*

Proof Let $v \in \mathsf{V}$ and write $v = u + w$ for $u \in \mathsf{U}$ and $w \in \mathsf{W}$. Then

$$\|v + \mathsf{U}\|_{/\mathsf{U}} = \|w + \mathsf{U}\|_{/\mathsf{U}} \leq \|w\| = \|w\|_{\mathsf{W}},$$

using the definition of $\|\cdot\|_{/\mathsf{U}}$.

For $j \in \mathbb{Z}_{>0}$ let $u_j \in \mathsf{U}$ have the property that

$$\|w + u_j\| - \|\pi_{\mathsf{U}}(w)\|_{/\mathsf{U}} < \tfrac{1}{j},$$

noting that the expression on the left of the inequality is always positive. ▼

(i) $\Longrightarrow$ (iii)
(iii) $\Longrightarrow$ (ii)
(ii) $\Longrightarrow$ (i) This is a tautology. ■

One reason why split subspaces are distinguished is the following.

3.7.3 Theorem (Algebraic direct sums of closed subspaces are topological direct sums) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \|\cdot\|)$ be a normed $\mathbb{F}$-vector space, and let U be a split subspace of V with closed complement W. On U and W let $\|\cdot\|_{\mathsf{U}}$ and $\|\cdot\|_{\mathsf{W}}$, respectively, be the restrictions of $\|\cdot\|$. Then the map* complete?

$$\mathsf{U} \oplus \mathsf{W} \ni (\mathrm{u}, \mathrm{w}) \mapsto \mathrm{u} + \mathrm{w} \in \mathsf{V}$$

is a homeomorphism.

Proof Let us first show that the map $(u, w) \mapsto u + w$ is continuous. Let us denote the direct sum norm on $\mathsf{U} \oplus \mathsf{W}$ by $\|\cdot\|_{\mathsf{U}\oplus\mathsf{W}}$, noting that, by definition,

$$\|(u, w)\|_{\mathsf{U}\oplus\mathsf{W}} = \|u\|_{\mathsf{U}} + \|w\|_{\mathsf{W}}.$$

For $\epsilon \in \mathbb{R}_{>0}$ and $(u_0, w_0) \in \mathsf{U} \oplus \mathsf{W}$ suppose that $(u, w) \in \mathsf{U} \oplus \mathsf{W}$ satisfy $\|(u, w) - (u_0, w_0)\|_{\mathsf{U}\oplus\mathsf{W}} < \epsilon$. Then

$$\begin{aligned} \|(u + w) - (u_0 + w_0)\| &\leq \|u - u_0\| + \|w - w_0\| = \|u - u_0\|_{\mathsf{U}} + \|w - w_0\|_{\mathsf{W}} \\ &= \|(u - u_0, w - w_0)\|_{\mathsf{U}\oplus\mathsf{W}} = \|(u, w) - (u_0, w_0)\|_{\mathsf{U}\oplus\mathsf{W}} < \epsilon. \end{aligned}$$

This gives (uniform) continuity of the map $(u, w) \mapsto u + w$. Since this map is clearly an isomorphism of vector spaces, it follows from Theorem 3.5.31 that it is a homeomorphism. ■

So the point is that a split subspace gives V not just as an *algebraic* direct sum, but as a topological direct sum. Thus the topology of V is determined from the topology of U and its closed complement W.

3.7.2 Questions and answers about closed and split subspaces

An obvious first question one asks about closed and split subspaces is whether all closed subspaces are split. They are not, as the following examples shows.

3.7.4 Examples (Closed subspaces that are not split)

1. We claim that $c_0(\mathbb{R})$ is a closed but not split subspace of $\ell^\infty(\mathbb{R})$.

Theorem 17.26 of rudin's real and complex analysis

Example 5.19 of Rudin's functional analysis

A subspace is split if and only if there is a continuous projection onto it (Corollary 2.2.18 in AMR)

A next natural question regarding closed and split subspaces is that of whether there is anything special about Banach spaces for which all closed subspaces are split. The following theorem indicates that the topology of such Banach spaces is that of a Hilbert space. (We shall engage in a detailed study of Hilbert spaces in Chapter 4.)

3.7.5 Theorem (The Split Subspace Theorem) *For a Banach space* $(\mathsf{V}, \|\cdot\|)$ *the following statements are equivalent:*

(i) *every closed subspace of* V *is split;*

(ii) *there exists an inner product on* V *whose induced norm is equivalent to* $\|\cdot\|$.

3.7.3 Projection constants

kadec-snobar.pdf.gz

3.7.4 The proof of the Split Subspace Theorem

complemented-subspace-problem.pdf.gz

3.7.5 Notes

[Lindenstrauss and Tzafriri 1971] [Joichi 1966] [Davis, Dean, and Singer 1968]

Section 3.8

Examples of Banach spaces

In this section we consider some of the common Banach spaces we will encounter in these volumes. As has already been mentioned, these examples serve as more than just an illustration of the concept of a Banach space; the examples are of great interest *per se*. Many of the examples are interconnected in that there is a very general example that contains simpler ones as a subcase. Logically, the proper way to present such examples is to give the most general construction first, and then provide the particular situations as following from the general. However, this method of presentation has serious defect that we are often most interested in the simpler situation, and a purely logical presentation would require the reader to understand some unnecessary abstraction. Therefore, we present our examples in order from the most particular to the most general. This has the drawback of being repetitive, but the advantage that a reader will not have to absorb a degree of abstraction that is not needed in the simpler examples.

Do I need to read this section? As we have said, some of the examples in this section are crucial in understanding a lot of the applied material that will follow. As the very least the reader should understand the spaces $\mathsf{L}^p(I;\mathbb{F})$ and $\ell^p(\mathbb{F})$. Some of the other examples can perhaps be omitted on a first reading, and covered when needed. •

3.8.1 The p-norms on $\mathbb{F}^n$

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let us begin our presentation with the simplest situation of a class of norms on a finite-dimensional $\mathbb{F}$-vector space. We are interested in a concrete collection of norms on the vector space $\mathbb{F}^n$. Specifically, for $p \in [1,\infty]$ we define a norm $\|\cdot\|_p$ on $\mathbb{F}^n$ by

$$\|(v_1,\dots,v_n)\|_p = \begin{cases} \left(\sum_{j=1}^n |v_j|^p\right)^{1/p}, & p \in [1,\infty), \\ \max\{|v_1|,\dots,|v_n|\}, & p = \infty. \end{cases}$$

That this is a norm for $p \in \{1,\infty\}$ has already been shown in Examples 3.1.3–3 and 3.1.3–4. In order to show that $\|\cdot\|_p$ is a norm for $p \in [1,\infty)$, the only nontrivial verification is of the triangle inequality. We verify this by using the following lemma.

3.8.1 Lemma (Hölder's inequality) *If* $\mathrm{a}_1,\dots,\mathrm{a}_\mathrm{n},\mathrm{b}_1,\dots,\mathrm{b}_\mathrm{n} \in \mathbb{R}_{\geq 0}$ *and if* $\mathrm{p} \in (1,\infty)$ *then*

$$\sum_{j=1}^{n} a_j b_j \leq \left(\sum_{j=1}^{n} a_j^p\right)^{1/p} \left(\sum_{j=1}^{n} b_j^{p'}\right)^{1/p'},$$

where $\frac{1}{\mathrm{p}} + \frac{1}{\mathrm{p}'} = 1$*. Moreover, equality holds if and only if* $(\mathrm{a}_1^{\mathrm{p}}, \dots, \mathrm{a}_{\mathrm{n}}^{\mathrm{p}})$ *and* $(\mathrm{b}_1^{\mathrm{p}'}, \dots, \mathrm{b}_{\mathrm{n}}^{\mathrm{p}'})$ *are collinear.*

Proof We first prove a lemma.

1 Sublemma *If* $\mathrm{a}, \mathrm{b} \in \mathbb{R}_{\geq 0}$ *and if* $\alpha \in (0, 1)$ *then*

$$\mathrm{a}^{\alpha}\mathrm{b}^{1-\alpha} \leq \alpha\mathrm{a} + (1-\alpha)\mathrm{b},$$

and equality holds if and only if $\mathrm{a} = \mathrm{b}$.

Proof If $a = b$ then both sides of the inequality are equal to a, and so the result holds in this case. Thus we consider the case when $a \neq b$. Since the desired inequality is symmetric with respect to a and b we can assume that $b > a$ without loss of generality. Consider the function $f\colon [a, b] \to \mathbb{R}$ defined by $f(x) = x^{1-\alpha}$. By the Mean Value Theorem there exists $c \in (a, b)$ such that

$$f'(c) = (1-\alpha)c^{-\alpha} = \frac{f(b) - f(a)}{b - a} = \frac{b^{1-\alpha} - a^{1-\alpha}}{b - a}.$$

Thus $b^{1-\alpha} - a^{1-\alpha} = (b-a)(1-\alpha)c^{-\alpha}$. Since $\alpha \in (0, 1)$ and since $c > a$ it follows that $c^{-\alpha} < a^{-\alpha}$. Therefore,

$$\begin{aligned} & b^{1-\alpha} - a^{1-\alpha} < (b-a)(1-\alpha)a^{-\alpha}, \\ \implies \quad & a^{\alpha}b^{1-\alpha} - a < (b-a)(1-\alpha) \\ \implies \quad & a^{\alpha}b^{1-\alpha} < \alpha a + (1-\alpha)b. \end{aligned}$$

Since this inequality is strict for $b > a$ the result follows. ▼

Let us denote $\alpha = \frac{1}{p}$ and $\beta = \frac{1}{p'} = 1 - \alpha$. Define $a'_j = a_j^{1/\alpha}$ and $b'_j = b_j^{1/\beta}$ and suppose initially that $\sum_{j=1}^n a'_j = 1$ and $\sum_{j=1}^n b'_j = 1$. By Sublemma 1 we have

$$\begin{aligned} & (a'_j)^{\alpha}(b'_j)^{\beta} \leq \alpha a'_j + \beta b'_j, \qquad j \in \{1, \dots, n\}, \\ \implies \quad & \sum_{j=1}^n ((a'_j)^{\alpha}(b'_j)^{\beta}) \leq \sum_{j=1}^n (\alpha a'_j + \beta b'_j) = \alpha + \beta = 1 = \left(\sum_{j=1}^n a'_j\right)^{\alpha} \left(\sum_{j=1}^n b'_j\right)^{\beta}, \\ \implies \quad & \sum_{j=1}^n a_j b_j \leq \left(\sum_{j=1}^n a_j^p\right)^{1/p} \left(\sum_{j=1}^n b_j^{p'}\right)^{1/p'}, \end{aligned}$$

with equality holding if and only if $a'_j = b'_j$, $j \in \{1, \dots, n\}$. This gives inequality in the sublemma when $\sum_{j=1}^n a'_j = 1$ and $\sum_{j=1}^n b'_j = 1$. If these relations do not hold then we have $\sum_{j=1}^n a'_j = \lambda$ and $\sum_{j=1}^n b'_j = \mu$ for some $\lambda, \mu \in \mathbb{R}_{\geq 0}$. Since the inequality is clearly equality if either $\lambda = 0$ or $\mu = 0$, we can suppose that $\lambda, \mu \in \mathbb{R}_{>0}$ without loss of generality. We can then write $a''_j = \frac{1}{\lambda}a'_j$ and $b''_j = \frac{1}{\mu}b'_j$ for $j \in \{1, \dots, n\}$ so that $\sum_{j=1}^n a''_j = \sum_{j=1}^n b''_j = 1$.

Then

$$\begin{aligned}\sum_{j=1}^n a_j b_j &= \sum_{j=1}^n (a'_j)^\alpha (b'_j)^\beta = \lambda^\alpha \mu^\beta \sum_{j=1}^n (a''_j)^\alpha (b''_j)^\beta \\ &\le \lambda^\alpha \mu^\beta \left(\sum_{j=1}^n a''_j\right)^\alpha \left(\sum_{j=1}^n b''_j\right)^\beta = \left(\sum_{j=1}^n a'_j\right)^\alpha \left(\sum_{j=1}^n b'_j\right)^\beta \\ &= \left(\sum_{j=1}^n a_j^p\right)^{1/p} \left(\sum_{j=1}^n b_j^{p'}\right)^{1/p'},\end{aligned}$$

giving the desired inequality. Moreover, from our previous computations, equality holds if and only if $a''_j = b''_j$, $j \in \{1, \dots, n\}$. This, in turn, holds if and only if $\mu a'_j = \lambda b'_j$, $j \in \{1, \dots, n\}$. In turn, this holds if and only if

$$\mu a_j^p = \lambda b_j^{p'}, \qquad j \in \{1, \dots, n\},$$

which is the result. ■

3.8.2 Notation (Conjugate index) For $p \in (1, \infty)$ the number $p' \in (1, \infty)$ such that $\frac{1}{p} + \frac{1}{p'} = 1$ is called the ***conjugate index*** for the index p. As we shall see, principally in Section 3.10.1, the conjugate index plays a surprisingly important rôle, although at this point it comes up as something of a conjurer's trick. Note that when $p = 2$ we have $p' = 2$. This in the important special case when the norm is derived from an inner product. We hope that the reader is tantalised at this moment. •

A variant of Hölder's inequality holds when $p = 1$, and we refer to Exercise 3.8.1 for this.

We next prove the useful Minkowski inequality.

3.8.3 Lemma (Minkowski's inequality) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *if* $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{F}$, *and if* $p \in [1, \infty)$ *then*

$$\left(\sum_{j=1}^n |a_j + b_j|^p\right)^{1/p} \le \left(\sum_{j=1}^n |a_j|^p\right)^{1/p} + \left(\sum_{j=1}^n |b_j|^p\right)^{1/p}.$$

Moreover, equality holds if and only if the following conditions hold:

(i) $p = 1$*: for each* $j \in \{1, \dots, n\}$ *there exists* $\alpha_j, \beta_j \in \mathbb{R}_{\ge 0}$*, not both zero, such that* $\alpha_j a_j = \beta_j b_j$*;*

(ii) $p \in (1, \infty)$*: there exists* $\alpha, \beta \in \mathbb{R}_{\ge 0}$*, not both zero, such that* $\alpha a_j = \beta b_j$ *for every* $j \in \{1, \dots, n\}$.

Proof The first part of the lemma has been proved for $p = 1$ in Example 3.1.3–3. Let us also prove the second part of the lemma for $p = 1$. First of all, it is easy to check that (i) is sufficient for equality in the Minkowski inequality. For the converse, note that, no matter whether $\mathbb{F} = \mathbb{R}$ or $\mathbb{F} = \mathbb{C}$, equality holds in the triangle inequality $|a+b| \leq |a|+|b|$, $a, b \in \mathbb{F}$, if and only if there exists $\alpha, \beta \in \mathbb{R}_{\geq 0}$, not both zero, such that $\alpha a = \beta b$. The reader not seeing this is encouraged to do the elementary geometry needed to verify this. From this observation,

$$\sum_{j=1}^{n} |a_j + b_j| = \sum_{j=1}^{n} |a_j| + \sum_{j=1}^{n} |b_j|$$

if and only if (i) holds.

Since the case of $p = 1$ has already been proved, we consider $p \in (1, \infty)$. We compute, using Lemma 3.8.1,

$$\begin{aligned}
\sum_{j=1}^{n} |a_j + b_j|^p &= \sum_{j=1}^{n} |a_j + b_j||a_j + b_j|^{p-1} \\
&\leq \sum_{j=1}^{n} |a_j||a_j + b_j|^{p-1} + \sum_{j=1}^{n} |b_j||a_j + b_j|^{p-1} \\
&\leq \left(\sum_{j=1}^{n} |a_j|^p\right)^{1/p} \left(\sum_{j=1}^{n} |a_j + b_j|^p\right)^{1/p'} + \left(\sum_{j=1}^{n} |b_j|^p\right)^{1/p} \left(\sum_{j=1}^{n} |a_j + b_j|^p\right)^{1/p'} \\
&= \left(\left(\sum_{j=1}^{n} |a_j|^p\right)^{1/p} + \left(\sum_{j=1}^{n} |b_j|^p\right)^{1/p}\right) \left(\sum_{j=1}^{n} |a_j + b_j|^p\right)^{1/p'}
\end{aligned}$$

from which we deduce, using the fact that $\frac{1}{p} = 1 - \frac{1}{p'}$,

$$\left(\sum_{j=1}^{n} |a_j + b_j|^p\right)^{1/p} \leq \left(\sum_{j=1}^{n} |a_j|^p\right)^{1/p} + \left(\sum_{j=1}^{n} |b_j|^p\right)^{1/p},$$

as desired. By considering where the possible inequality is introduced in the preceding computation, and in view of Lemma 3.8.1, equality in the statement of the sublemma holds if and only if

1. for each $j \in \{1, \ldots, n\}$ there exists $\alpha_j, \beta_j \in \mathbb{R}_{\geq 0}$, not both zero, such that $\alpha_j a_j = \beta_j b_j$ and
2. both $(|a_1|^p, \ldots, |a_n|^p)$ and $(|b_1|^p, \ldots, |b_n|^p)$ are collinear with $(|a_1 + b_1|^p, \ldots, |a_n + b_n|^p)$.

The second of these conditions is equivalent to the existence of $\alpha, \lambda \in \mathbb{R}_{\geq 0}$, not both zero, and $\beta, \mu \in \mathbb{R}_{\geq 0}$, not both zero, such that

$$\alpha |a_j|^p = \lambda |a_j + b_j|^p, \quad \beta |b_j|^p = \mu |a_j + b_j|^p.$$

We consider a few cases.

1. $a_j, b_j \neq 0$ for every j: In this case we must have $\alpha_j, \beta_j \in \mathbb{R}_{>0}$. Then $a_j = \delta_j b_j$ for $\delta_j = \frac{\beta_j}{\alpha_j} \in \mathbb{R}_{>0}$. We can then solve for δ_j to give $\delta_j = \frac{\lambda}{\alpha - \lambda}$. Note that $\alpha \neq \lambda$ since $b_j \neq 0$. This gives
$$(\alpha - \lambda) a_j = \lambda b_j$$
for every $j \in \{1, \ldots, n\}$, giving the result in this case.

2. $a_j, b_j \neq 0$ for some $j \in \{1, \dots, n\}$: In this case, whenever $a_j, b_j \neq 0$ the argument from the previous case gives
$$(\alpha - \lambda)a_j = \lambda b_j$$
Now we consider some subcases, taking into account that a_j and/or b_j might be zero for some j.
   (a) $a_j = 0$, $b_j \neq 0$: In this case we have $\lambda = \beta_j = 0$ and $\beta = \mu$. It, therefore, holds that
$$(\alpha - \lambda)a_j = \lambda b_j.$$
   (b) $a_j \neq 0$, $b_j = 0$: In this case $\mu = \alpha_j = 0$ and $\alpha = \lambda$. It, therefore, holds that
$$(\alpha - \lambda)a_j = \lambda b_j.$$
3. $a_j = 0$ for all $j \in \{1, \dots, n\}$: In this case we have, for any $\alpha \in \mathbb{R}_{>0}$ and with $\beta = 0$,
$$\alpha a_j = \beta b_j$$
for all $j \in \{1, \dots, n\}$.
4. $b_j = 0$ for all $j \in \{1, \dots, n\}$: In this case we have, for any $\beta \in \mathbb{R}_{>0}$ with $\alpha = 0$,
$$\alpha a_j = \beta b_j$$
for all $j \in \{1, \dots, n\}$.

The upshot of the preceding monotony is that condition (ii) holds when equality in the Minkowski inequality holds. ■

There is another version of the Minkowski inequality that is sometimes useful. We call this the "integral version" of the Minkowski inequality for reasons that are best made clear in . what?

3.8.4 Lemma (Integral version of Minkowski's inequality) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *if* $\mathrm{a_{jk}} \in \mathbb{F}$, $\mathrm{j} \in \{1, \dots, \mathrm{m}\}$, $\mathrm{k} \in \{1, \dots, \mathrm{n}\}$, *and if* $\mathrm{p} \in [1, \infty)$ *then* must prove when equality occurs
$$\left(\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p}\right)^{1/p} \leq \sum_{k=1}^{n}\left(\sum_{j=1}^{m}|a_{jk}|^{p}\right)^{1/p}.$$
Moreover, equality holds if and only if there exists $\mathrm{b_1, \dots, b_m, c_1, \dots, c_n} \in \mathbb{F}$ *such that* $\mathrm{a_{jk} = b_j c_k}$.

Proof For $p = 1$ we have
$$\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right| \leq \sum_{j=1}^{m}\left(\sum_{k=1}^{n}|a_{jk}|\right) = \sum_{k=1}^{n}\left(\sum_{j=1}^{m}|a_{jk}|\right),$$
giving the result in this case.

Now let $p \in (1, \infty)$. Here we compute

$$\begin{aligned}\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p} &= \sum_{j=1}^{m}\left(\left|\sum_{k=1}^{n} a_{jk}\right|^{p-1}\right)\left(\left|\sum_{l=1}^{n} a_{jl}\right|\right)\\ &\le \sum_{j=1}^{m}\left(\sum_{l=1}^{n}\left(|a_{jl}|\left|\sum_{k=1}^{n} a_{jk}\right|^{p-1}\right)\right)\\ &= \sum_{l=1}^{n}\left(\sum_{j=1}^{m}\left(|a_{jl}|\left|\sum_{k=1}^{n} a_{jk}\right|^{p-1}\right)\right),\end{aligned}$$

swapping the order of summation in the last step. Now let $p' = \frac{p}{p-1}$ be the conjugate index. Now, by Hölder's inequality,

$$\begin{aligned}\sum_{j=1}^{m}\left(|a_{jl}|\left|\sum_{k=1}^{n} a_{jk}\right|^{p-1}\right) &\le \left(\sum_{j=1}^{m}|a_{jl}|^{p}\right)^{1/p}\left(\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p'(p-1)}\right)^{1/p'}\\ &= \left(\sum_{j=1}^{m}|a_{jl}|^{p}\right)^{1/p}\left(\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p}\right)^{1/p'}.\end{aligned}$$

Substituting this last relation into the preceding equation yields

$$\begin{aligned}\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p} &\le \sum_{l=1}^{n}\left(\left(\sum_{j=1}^{m}|a_{jl}|^{p}\right)^{1/p}\left(\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p}\right)^{1/p'}\right)\\ &= \left(\sum_{l=1}^{n}\left(\sum_{j=1}^{m}|a_{jl}|^{p}\right)^{1/p}\right)\left(\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p}\right)^{1/p'}.\end{aligned}$$

Now we note that the lemma is obviously true when

$$\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p} = 0.$$

So we suppose that this quantity is nonzero and divide the above-derived inequality

$$\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p} \le \left(\sum_{l=1}^{n}\left(\sum_{j=1}^{m}|a_{jl}|^{p}\right)^{1/p}\right)\left(\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p}\right)^{1/p'}$$

by

$$\left(\sum_{j=1}^{m}\left|\sum_{k=1}^{n} a_{jk}\right|^{p}\right)^{1/p'},$$

which gives the desired inequality after noting that p' is conjugate to p. ■

From Minkowski's inequality we immediately have the following result.

3.8.5 Proposition ($(\mathbb{F}^n, \|\cdot\|_p)$ is a Banach space) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and if* $p \in [1, \infty]$ *then* $(\mathbb{F}^n, \|\cdot\|_p)$ *is an* $\mathbb{F}$*-Banach space.*

Moreover, we know from Theorem 3.1.15 that the norms $\|\cdot\|_p$ are equivalent. One can then wonder at why one would not just choose one of these norms and be done with it. There are at least two reasons why not.

1. Sometimes one norm is more convenient than another.
2. The finite-dimensional setting provides an opportunity to begin to understand the rôle p in how the norms "look." These sorts of p-norms will come up in increasingly more abstract settings, and the finite-dimensional example gives some useful intuition.

Along the lines of using the finite-dimensional setting to provide some intuition for more complicated ideas that will arise later, let us consider a variant of the p-norm for $p \in (0,1)$. The definition is the same. For $p \in (0,1)$ we define

$$\|(v_1, \dots, v_n)\|_p = \left(\sum_{j=1}^{n} |v_j|^p\right)^{1/p}.$$

The function $v \mapsto \|v\|_p$ clearly has the positivity and homogeneity properties needed for a norm. What we lose is the triangle inequality. Indeed, for $p \in (0,1)$ we have the following results which mirror Lemmata 3.8.1 and 3.8.3.

3.8.6 Lemma (Hölder's inequality for $p \in (0,1)$) *If* $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}_{\geq 0}$ *and if* $p \in (0,1)$ *then*

$$\sum_{j=1}^{n} a_j b_j \geq \left(\sum_{j=1}^{n} a_j^p\right)^{1/p} \left(\sum_{j=1}^{n} b_j^{p'}\right)^{1/p'},$$

where $\frac{1}{p} + \frac{1}{p'} = 1$*.*

Proof Let $q = p^{-1}$ so that $q \in (1, \infty)$ and define $c_j = b_j^{-1/q}$ and $d_j = a_j^{1/q} b_j^{1/q}$, $j \in \{1, \dots, n\}$. Let q' satisfy $\frac{1}{q} + \frac{1}{q'} = 1$. Then one shows directly that $c_j^{q'} = b_j^{p'}$ and $a_j^p = c_j d_j$, $j \in \{1, \dots, n\}$. Then we have, using Lemma 3.8.1,

$$\sum_{j=1}^{n} a_j^p = \sum_{j=1}^{n} c_j d_j \leq \left(\sum_{j=1}^{n} d_j^q\right)^{1/q} \left(\sum_{j=1}^{n} c_j^{q'}\right)^{1/q'} = \left(\sum_{j=1}^{n} a_j b_j\right)^{p} \left(\sum_{j=1}^{n} b_j^{p'}\right)^{1/q'},$$

from which we deduce that

$$\sum_{j=1}^{n} a_j b_j \geq \left(\sum_{j=1}^{n} a_j^p\right)^{1/p} \left(\sum_{j=1}^{n} b_j^{p'}\right)^{-1/(q'p)},$$

from which the result follows since $-\frac{1}{q'p} = \frac{1}{p'}$. ■

3.8.7 Lemma (Minkowski's inequality for p ∈ (0, 1)) *If* $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}_{\geq 0}$ *and if* $p \in (0, 1)$ *then*

$$\left(\sum_{j=1}^{n}(a_j + b_j)^p\right)^{1/p} \geq \left(\sum_{j=1}^{n} a_j^p\right)^{1/p} + \left(\sum_{j=1}^{n} b_j^p\right)^{1/p},$$

Proof This follows from Lemma 3.8.6 using the same sequence of computations used in proving that Lemma 3.8.3 follows from Lemma 3.8.1. ■

In Figure 3.4 we depict the boundaries of the balls $\mathsf{B}_p(1, \mathbf{0})$ in $\mathbb{R}^2$. The main

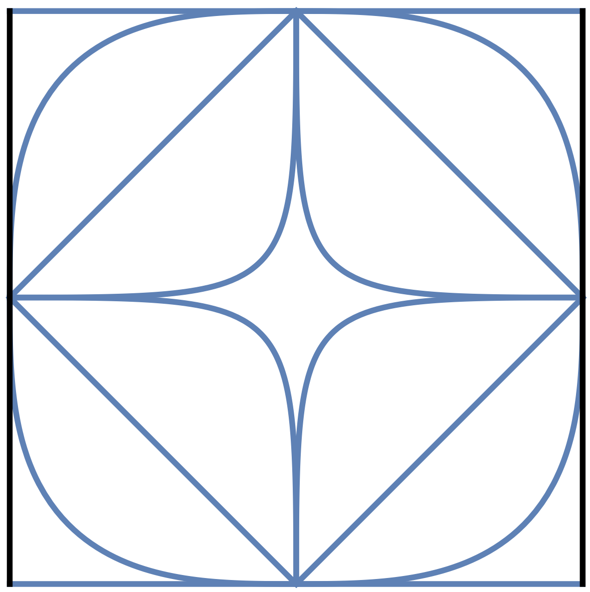


Figure 3.4 The unit spheres for the (if $p > 1$, at least) norms $\|\cdot\|_p$ on $\mathbb{R}^2$ (shown are, from inside to out, $p \in \{1/3, 1, 3, \infty\}$)

point is that the balls are convex of and only if $p \in [1, \infty)$. In the present finite-dimensional setting this has no consequences. One can define a topology on $\mathbb{F}^n$ as being generated by the open balls, even though they are not convex. This topology is equivalent to the standard topology (one can see this by applying), and so all the usual notions of convergence, continuity, etc., carry over to this case. However, when we generalise this to infinite-dimensions, it turns out that the lack of convexity causes problems. For example, in we shall see that the lack of convexity causes the topological dual to consist only of the zero functional.

what?

what?

3.8.2 Banach spaces of sequences

Among the more important classes of Banach spaces we will encounter are those that are sequences characterised by certain summability properties. As we shall expound on in detail in , such Banach spaces are models for discrete time- and frequency-domain representations of signals. Here we are merely interested in some basic definitions and properties. what?

The most fundamental Banach space of sequences are those that are bounded.

3.8.8 Definition ($\ell^\infty(\mathbb{F})$) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Define a subspace $\ell^\infty(\mathbb{F})$ of $\mathbb{F}^{\mathbb{Z}_{>0}}$ by

$$\ell^\infty(\mathbb{F}) = \{(a_j)_{j\in\infty} \mid \text{ there exists } M \in \mathbb{R}_{>0} \text{ such that } |a_j| \leq M,\ j \in \mathbb{Z}_{>0}\}$$

and define

$$\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_\infty = \sup\{|a_j| \mid j \in \mathbb{Z}_{>0}\}$$

for $(a_j)_{j\in\mathbb{Z}_{>0}} \in \ell^\infty(\mathbb{F})$. •

Thus $\ell^\infty(\mathbb{F})$ consists of the set of bounded sequences in $\mathbb{F}$ and $\|\cdot\|_\infty$ is the least upper bound for the terms in the sequence.

3.8.9 Theorem ($\ell^\infty(\mathbb{F})$ is a Banach space) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ then $(\ell^\infty(\mathbb{F}), \|\cdot\|_\infty)$ is an $\mathbb{F}$-Banach space.*

Proof The only not entirely trivial norm property to verify for $\|\cdot\|_\infty$ is the triangle inequality:

$$\begin{aligned}\|(a_j)_{j\in\mathbb{Z}_{>0}} + (b_j)_{j\in\mathbb{Z}_{>0}}\|_\infty &= \sup\{|a_j + b_j| \mid j \in \mathbb{Z}_{>0}\}\\ &\leq \sup\{|a_j| + |b_j| \mid j \in \mathbb{Z}_{>0}\}\\ &= \sup\{|a_j| \mid j \in \mathbb{Z}_{>0}\} + \sup\{|b_j| \mid j \in \mathbb{Z}_{>0}\}\\ &= \|(a_j)_{j\in\mathbb{Z}_{>0}}\|_\infty + \|(b_j)_{j\in\mathbb{Z}_{>0}}\|_\infty,\end{aligned}$$

where we have used Proposition I-2.2.27.

Now let us verify that $(\ell^\infty(\mathbb{F}), \|\cdot\|_\infty)$ is complete. We let $((a_j^{(l)})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\ell^\infty(\mathbb{F})$. We claim that, for each $j \in \mathbb{Z}_{>0}$, $(a_j^{(l)})_{l\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathbb{F}$. To see this, let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\left\|(a_j^{(l)})_{j\in\mathbb{Z}_{>0}} - (a_j^{(k)})_{j\in\mathbb{Z}_{>0}}\right\|_\infty < \epsilon$$

for $k, l \geq N$. Then, by definition of $\|\cdot\|_\infty$,

$$\left|a_j^{(l)} - a_j^{(k)}\right| < \epsilon$$

for $k, l \geq N$ and for $j \in \mathbb{Z}_{>0}$. Thus $(a_j^{(l)})_{l\in\mathbb{Z}_{>0}}$ is indeed a Cauchy sequence, and so converges to some $a_j \in \mathbb{F}$. We now claim that the sequence $((a_j^{(l)})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$ converges to $(a_j)_{j\in\mathbb{Z}_{>0}}$. To see this, let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\left\|(a_j^{(l)})_{j\in\mathbb{Z}_{>0}} - (a_j^{(k)})_{j\in\mathbb{Z}_{>0}}\right\|_\infty < \tfrac{\epsilon}{2}$$

for $k, l \geq N$. Thus
$$\left|a_j^{(l)} - a_j^{(l)}\right| < \tfrac{\epsilon}{2}, \qquad k, l \geq N.$$
Now, for fixed $j \in \mathbb{Z}_{>0}$, let $N' \in \mathbb{Z}_{>0}$ be sufficiently large that $\left|a_j^{(k)} - a_j\right| < \frac{\epsilon}{2}$ for $k \geq N'$. In this case, if $l \geq N$ and $k \geq \max\{N, N'\}$, we have
$$\left|a_j^{(l)} - a_j\right| \leq \left|a_j^{(l)} - a_j^{(k)}\right| + \left|a_j^{(k)} - a_j\right| < \epsilon.$$
Since this holds for each $j \in \mathbb{Z}_{>0}$ we have
$$\left\|(a_j^{(l)})_{j\in\mathbb{Z}_{>0}} - (a_j)_{j\in\mathbb{Z}_{>0}}\right\|_\infty \leq \epsilon,$$
as desired. ■

One property of $\ell^\infty(\mathbb{F})$ that makes it different than some of the other Banach spaces we consider is the following.

3.8.10 Proposition ($\ell^\infty(\mathbb{F})$ is not separable) *For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, the Banach space $(\ell^\infty(\mathbb{F}), \|\cdot\|_\infty)$ is not separable.*

Proof Let $\mathscr{U}$ be the collection of sequences $(a_j)_{j\in\mathbb{Z}_{>0}} \in \ell^\infty(\mathbb{F})$ such that $a_j \in \{-1, 1\}$, $j \in \mathbb{Z}_{>0}$. It follows from Exercises I-1.7.4, I-1.7.5, and I-2.1.4 that $\mathscr{U}$ is countable. Note that if $(a_j)_{j\in\mathbb{Z}_{>0}}, (b_j)_{j\in\mathbb{Z}_{>0}} \in \mathscr{U}$ are distinct then
$$\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_\infty = 1, \quad \|(a_j)_{j\in\mathbb{Z}_{>0}} - (b_j)_{j\in\mathbb{Z}_{>0}}\|_\infty = 2.$$
Let $(a_j)_{j\in\mathbb{Z}_{>0}} \in \mathscr{U}$ and let $(b_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{B}(1, (a_j)_{j\in\mathbb{Z}_{>0}})$. By Exercise 3.1.3 we have
$$\begin{aligned} & \left|\|(b_j)_{j\in\mathbb{Z}_{>0}}\| - \|(a_j)_{j\in\mathbb{Z}_{>0}}\|_\infty\right| \leq \left\|(b_j)_{j\in\mathbb{Z}_{>0}} - (a_j)_{j\in\mathbb{Z}_{>0}}\right\|_\infty \\ \implies\quad & \left|\|(b_j)_{j\in\mathbb{Z}_{>0}}\| - 1\right| \leq 1 \\ \implies\quad & \left\|(b_j)_{j\in\mathbb{Z}_{>0}}\right\|_\infty \leq 2. \end{aligned}$$
Thus $\mathsf{B}(1, (a_j)_{j\in\mathbb{Z}_{>0}}) \subseteq \mathsf{B}(2, 0_{\ell^\infty(\mathbb{F})})$ for each $(a_j)_{j\in\mathbb{Z}_{>0}} \in \mathscr{U}$. If
$$(a_j)_{j\in\mathbb{Z}_{>0}}, (b_j)_{j\in\mathbb{Z}_{>0}} \in \mathscr{U}$$
and
$$(c_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{B}(1, (a_j)_{j\in\mathbb{Z}_{>0}}), \quad (d_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{B}(1, (b_j)_{j\in\mathbb{Z}_{>0}})$$
then
$$\begin{aligned} &\|(c_j)_{j\in\mathbb{Z}_{>0}} - (b_j)_{j\in\mathbb{Z}_{>0}}\|_\infty \\ &\qquad \geq \left|\|(c_j)_{j\in\mathbb{Z}_{>0}} - (a_j)_{j\in\mathbb{Z}_{>0}}\|_\infty - \|(a_j)_{j\in\mathbb{Z}_{>0}} - (b_j)_{j\in\mathbb{Z}_{>0}}\|_\infty\right| \geq 2 \end{aligned}$$
using Proposition 1.1.3. Thus $(c_j)_{j\in\mathbb{Z}_{>0}} \notin \mathsf{B}(1, (b_j)_{j\in\mathbb{Z}_{>0}})$. One similarly shows that $(d_j)_{j\in\mathbb{Z}_{>0}} \notin \mathsf{B}(1, (a_j)_{j\in\mathbb{Z}_{>0}})$. This shows that $\mathsf{B}(2, 0_{\ell^\infty(\mathbb{F})})$ contains the collection
$$\{\mathsf{B}(1, (a_j)_{j\in\mathbb{Z}_{>0}}) \mid (a_j)_{j\in\mathbb{Z}_{>0}} \in \mathscr{U}\}$$

of disjoint open balls. In particular, if $((b_j^{(l)})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$ is any countable subset of $\ell^\infty(\mathbb{F})$ then there is a countable subset $((a_j^{(\alpha)})_{j\in\mathbb{Z}_{>0}})_{\alpha\in A}$ of $\mathscr{U}$ in which are contained all of the sequences $((b_j^{(l)})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$. Note that

$$\mathrm{cl}\Big(((b_j^{(l)})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}\Big) \subseteq \cup_{\alpha\in A}\overline{\mathsf{B}}(1,(a_j^{(\alpha)})_{j\in\mathbb{Z}_{>0}}).$$

Therefore, any of the set of balls

$$\{\mathsf{B}(1,(a_j)_{j\in\mathbb{Z}_{>0}}) \mid (a_j)_{j\in\mathbb{Z}_{>0}} \in \mathscr{U},\ (a_j)_{j\in\mathbb{Z}_{>0}} \neq (a_j^{(\alpha)})_{j\in\mathbb{Z}_{>0}},\ \alpha\in A\}$$

cannot lie in $\mathrm{cl}\Big(((b_j^{(l)})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}\Big)$ which prohibits $((b_j^{(l)})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$ from being dense. ■

Now we begin looking at subspaces of $\ell^\infty(\mathbb{F})$. We begin with subspaces of sequences that converge.

3.8.11 Definition ($\mathsf{c}(\mathbb{F})$ and $\mathsf{c}_0(\mathbb{F})$) Let $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$. Define subspaces $\mathsf{c}(\mathbb{F})$ and $\mathsf{c}_0(\mathbb{F})$ of $\mathbb{F}^{\mathbb{Z}_{>0}}$ by

$$\mathsf{c}(\mathbb{F}) = \left\{(a_j)_{j\in\mathbb{Z}_{>0}} \;\middle|\; \text{there exists } a\in\mathbb{F} \text{ such that } \lim_{j\to\infty} a_j = a\right\}$$

and

$$\mathsf{c}_0(\mathbb{F}) = \left\{(a_j)_{j\in\mathbb{Z}_{>0}} \;\middle|\; \lim_{j\to\infty} a_j = 0\right\},$$

respectively. •

Note that by Propositions I-2.3.23 and it follows that $\mathsf{c}(\mathbb{F})$ and $\mathsf{c}_0(\mathbb{F})$ are subspaces. Moreover, by Propositions I-2.3.3 and it follows that $\mathsf{c}(\mathbb{F})$ and $\mathsf{c}_0(\mathbb{F})$ are subspaces of $\ell^\infty(\mathbb{F})$. The appropriate norm to use on the spaces of sequences $\mathsf{c}(\mathbb{F})$ and $\mathsf{c}_0(\mathbb{F})$ is the restriction of norm $\|\cdot\|_\infty$ on $\ell^\infty(\mathbb{F})$. We denote this norm simply by $\|\cdot\|_\infty$. With this norm our spaces of convergent sequences are Banach spaces.

complex versions

complex version

3.8.12 Theorem (($\mathsf{c}(\mathbb{F}),\|\cdot\|_\infty$) and ($\mathsf{c}_0(\mathbb{F}),\|\cdot\|_\infty$) are Banach spaces) *If* $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ *then* $(\mathsf{c}(\mathbb{F}),\|\cdot\|_\infty)$ *and* $(\mathsf{c}_0(\mathbb{F}),\|\cdot\|_\infty)$ *are* $\mathbb{F}$*-Banach spaces.*

Proof Let $((a_j^{(l)})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathsf{c}(\mathbb{F})$. By Theorem 3.8.9 this means that the sequence converges to $(a_j)_{j\in\mathbb{Z}_{>0}}\in\ell^\infty(\mathbb{F})$. Since each sequence $(a_j^{(l)})_{j\in\mathbb{Z}_{>0}}$ is in $\mathsf{c}(\mathbb{F})$ there exists $a^{(l)}\in\mathbb{F}$ such that $\lim_{j\to\infty} a_j^{(l)} = a^{(l)}$. We claim that $(a^{(l)})_{l\in\mathbb{Z}_{>0}}$ is a Cauchy sequence. For $\epsilon\in\mathbb{R}_{>0}$ let $N\in\mathbb{Z}_{>0}$ be sufficiently large that

$$\left\|(a_j^{(k)})_{j\in\mathbb{Z}_{>0}} - (a_j^{(l)})_{j\in\mathbb{Z}_{>0}}\right\|_\infty < \tfrac{\epsilon}{3},$$

which implies that

$$\left|a_j^{(k)} - a_j^{(l)}\right| < \tfrac{\epsilon}{3}, \qquad k,l\geq N,\ j\in\mathbb{Z}_{>0}.$$

Now let $k, l \geq N$ and let $j \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\left|a_j^{(k)} - a^{(k)}\right| < \tfrac{\epsilon}{3}, \quad \left|a_j^{(l)} - a^{(k)}\right| < \tfrac{\epsilon}{3}.$$

Then

$$\left|a^{(k)} - a^{(l)}\right| \leq \left|a^{(k)} - a_j^{(k)}\right| + \left|a_j^{(k)} - a_j^{(l)}\right| + \left|a_j^{(l)} - a^{(l)}\right| < \epsilon.$$

As this holds for every $k, l \geq N$ it follows that $(a^{(l)})_{l\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathbb{F}$. We denote its limit by a.

Finally we show that $\lim_{j\to\infty} a_j = a$, which shows that $(a_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{c}(\mathbb{F})$. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N' \in \mathbb{Z}_{>0}$ be sufficiently large that $\left|a^{(k)} - a\right| < \frac{\epsilon}{3}$ for $k \geq N'$. Now fix $k \geq N'$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\left|a_j - a_j^{(k)}\right| < \tfrac{\epsilon}{3}, \quad \left|a_j^{(k)} - a^{(k)}\right| < \tfrac{\epsilon}{3}, \qquad j \geq N.$$

Then, for $j \geq N$,

$$|a_j - a| \leq \left|a_j - a_j^{(k)}\right| + \left|a_j^{(k)} - a^{(k)}\right| + \left|a^{(k)} - a\right| < \epsilon,$$

which completes the proof that $(\mathsf{c}(\mathbb{F}), \|\cdot\|_\infty)$ is a Banach space.

If $((a_j^{(l)})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathsf{c}_0(\mathbb{F}) \subseteq \mathsf{c}(\mathbb{F})$ the above argument is easily modified to show that the limit sequence, denoted $(a_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{c}(\mathbb{F})$ above is actually in $\mathsf{c}_0(\mathbb{F})$. The key point is that $a^{(l)} = 0$ for each $l \in \mathbb{Z}_{>0}$ and so $a = 0$ as well. Thus $(\mathsf{c}_0(\mathbb{F}), \|\cdot\|_\infty)$ is also a Banach space. ■

The Banach spaces $\mathsf{c}(\mathbb{F})$ an $\mathsf{c}_0(\mathbb{F})$ have the friendly property of being separable.

3.8.13 Proposition ($\mathsf{c}(\mathbb{F})$ and $\mathsf{c}_0(\mathbb{F})$ are separable) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ then the Banach spaces $(\mathsf{c}(\mathbb{F}), \|\cdot\|_\infty)$ and $(\mathsf{c}_0(\mathbb{F}), \|\cdot\|_\infty)$ are separable.*

Proof It suffices to prove the proposition for $\mathsf{c}(\mathbb{F})$. We first take the case when $\mathbb{F} = \mathbb{R}$. In this case, for $q \in \mathbb{Q}$, we let $\mathscr{D}_q(\mathbb{R})$ be the subset of $\mathsf{c}(\mathbb{R})$ consisting of sequences $(q_j)_{j\in\mathbb{Z}_{>0}}$ with $q_j \in \mathbb{Q}$, $j \in \mathbb{Z}_{>0}$, and such that $q_j = q$ for all j sufficiently large. We then take

$$\mathscr{D}(\mathbb{R}) = \cup_{q\in\mathbb{Q}} \mathscr{D}_q(\mathbb{R}).$$

We claim that $\mathscr{D}(\mathbb{R})$ is countable. We note that $\mathscr{D}_q(\mathbb{R})$ is a countable (indexed by $\mathbb{Z}_{>0}$) disjoint union of copies of $\mathbb{Q}$ and so is countable by Proposition I-1.7.16. Thus $\mathscr{D}(\mathbb{R})$ is a countable union of countable sets, and so is again countable by Proposition I-1.7.16. We should also show that $\mathscr{D}(\mathbb{R})$ is dense in $\mathsf{c}(\mathbb{R})$. Let $(a_j)_{j\in\mathbb{Z}_{>0}}$ and let $\epsilon \in \mathbb{R}_{>0}$. Suppose that $q \in \mathbb{Q}$ is such that

$$\left|\lim_{j\to\infty} a_j - q\right| < \epsilon$$

and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $|a_j - q| < \epsilon$ for $j \geq N$. Now choose $q_1, \dots, q_N \in \mathbb{Q}$ such that $|a_j - q_j| < \epsilon$ for $j \in \{1, \dots, N\}$. Now define $(q_j)_{j\in\mathbb{Z}_{>0}}$ by asking that $q_j = q$ for $j > N$. Then $(q_j)_{j\in\mathbb{Z}_{>0}} \in \mathscr{D}(\mathbb{R})$ and

$$\|(a_j)_{j\in\mathbb{Z}_{>0}} - (q_j)_{j\in\mathbb{Z}_{>0}}\|_\infty < \epsilon.$$

Thus $\mathscr{D}(\mathbb{R})$ is dense in $\mathsf{c}(\mathbb{R})$.

For $\mathbb{F} = \mathbb{C}$ the procedure above can be duplicated by letting $\mathscr{D}(\mathbb{C})$ be the set of sequences $(q_j + \mathrm{i}r_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{c}(\mathbb{C})$ with $(q_j)_{j\in\mathbb{Z}_{>0}}, (r_j)_{j\in\mathbb{Z}_{>0}} \in \mathscr{D}(\mathbb{R})$. ■

The result has the following interesting corollary.

3.8.14 Corollary ($\mathsf{c}_0(\mathbb{F})$ is the completion of $\mathbb{F}_0^\infty$) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *then* $(\mathsf{c}_0(\mathbb{F}), \|\cdot\|_\infty)$ *is the completion of* $(\mathbb{F}_0^\infty, \|\cdot\|_\infty)$.

Proof Borrowing the notation from the proof of Proposition 3.8.13 we have

$$\mathscr{D}_0(\mathbb{F}) \subseteq \mathbb{F}_0^\infty \subseteq \mathsf{c}_0(\mathbb{F})$$

from which we deduce that

$$\mathsf{c}_0(\mathbb{F}) = \mathrm{cl}(\mathscr{D}_0(\mathbb{F})) \subseteq \mathrm{cl}(\mathbb{F}_0^\infty) \subseteq \mathrm{cl}(\mathsf{c}_0(\mathbb{F})) = \mathsf{c}_0(\mathbb{F}).$$

Therefore, $\mathrm{cl}(\mathbb{F}_0^\infty) = \mathsf{c}_0(\mathbb{F})$, as desired. ■

Now we consider Banach spaces of sequences which naturally use a different norm that the ∞-norm.

3.8.15 Definition ($\ell^p(\mathbb{F})$) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $p \in [1, \infty)$. Define a subspace $\ell^p(\mathbb{F})$ of $\mathbb{F}^{\mathbb{Z}_{>0}}$ by

$$\ell^p(\mathbb{F}) = \left\{ (a_j)_{j\in\mathbb{Z}_{>0}} \;\middle|\; \sum_{j=1}^\infty |a_j|^p < \infty \right\}$$

and define

$$\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_p = \left(\sum_{j=1}^\infty |a_j|^p \right)^{1/p}$$

for $(a_j)_{j\in\mathbb{Z}_{>0}} \in \ell^p(\mathbb{F})$. •

At this point it is not necessarily clear that $\ell^p(\mathbb{F})$ is actually a subspace of $\mathbb{F}^{\mathbb{Z}_{>0}}$, but we shall show shortly that it is, and is in fact a Banach space when equipped with $\|\cdot\|_p$ as a norm.

Let us give some properties of the function $\|\cdot\|_p$ analogous to Lemmata 3.8.1 and 3.8.3.

3.8.16 Lemma (Hölder's inequality) *If* $p \in (1, \infty)$ *and if* $(a_j)_{j\in\mathbb{Z}_{>0}} \in \ell^p(\mathbb{F})$ *and* $(b_j)_{j\in\mathbb{Z}_{>0}} \in \ell^{p'}(\mathbb{F})$, *then*

$$\sum_{j=1}^\infty |a_j b_j| \le \left(\sum_{j=1}^\infty |a_j|^p \right)^{1/p} \left(\sum_{j=1}^\infty |b_j|^{p'} \right)^{1/p'},$$

where $\frac{1}{p} + \frac{1}{p'} = 1$. *Moreover, equality holds if and only if* $(|a_j|^p)_{j\in\mathbb{Z}_{>0}}$ *and* $(|b_j|^{p'})_{j\in\mathbb{Z}_{>0}}$ *are collinear.*

Proof For $N \in \mathbb{Z}_{>0}$, by Lemma 3.8.1 we have

$$\sum_{j=1}^{N} |a_j b_j| \le \left(\sum_{j=1}^{N} |a_j|^p\right)^{1/p} \left(\sum_{j=1}^{N} |b_j|^{p'}\right)^{1/p'} \le \left(\sum_{j=1}^{\infty} |a_j|^p\right)^{1/p} \left(\sum_{j=1}^{\infty} |b_j|^{p'}\right)^{1/p'}.$$

Thus

$$\sum_{j=1}^{\infty} |a_j b_j| = \lim_{N\to\infty} \sum_{j=1}^{N} |a_j b_j| \le \left(\sum_{j=1}^{\infty} |a_j|^p\right)^{1/p} \left(\sum_{j=1}^{\infty} |b_j|^{p'}\right)^{1/p'},$$

as desired.

For the final assertion of the lemma, first note that a direction computation shows that equality holds in the Hölder equality if $(|a_j|^p)_{j\in\mathbb{Z}_{>0}}$ and $(|b_j|^{p'})_{j\in\mathbb{Z}_{>0}}$ are collinear. For the converse, suppose that $(|a_j|^p)_{j\in\mathbb{Z}_{>0}}$ and $(|b_j|^{p'})_{j\in\mathbb{Z}_{>0}}$ are not collinear. Then there exists $N \in \mathbb{Z}_{>0}$ such that $(|a_1|^p, \dots, |a_N|^p)$ and $(|b_1|^{p'}, \dots, |b_N|^{p'})$ are not collinear. By Lemma 3.8.1 we then have

$$\sum_{j=1}^{N} |a_j b_j| < \left(\sum_{j=1}^{N} |a_j|^p\right)^{1/p} \left(\sum_{j=1}^{N} |b j|^{p'}\right)^{1/p'}.$$

Since

$$\sum_{j=N+1}^{\infty} |a_j b_j| < \left(\sum_{j=N+1}^{\infty} |a_j|^p\right)^{1/p} \left(\sum_{j=N+1}^{\infty} |b j|^{p'}\right)^{1/p'}$$

it follows that equality cannot hold in the Hölder inequality. ■

A version of Hölder's inequality holds for $p = 1$ and we refer to Exercise 3.8.2 for this.

The Minkowski inequality also holds in this case.

3.8.17 Lemma (Minkowski's inequality) *If* $p \in [1, \infty)$ *and if* $(a_j)_{j\in\mathbb{Z}_{>0}}, (b_j)_{j\in\mathbb{Z}_{>0}} \in \ell^p(\mathbb{F})$ *then*

$$\left(\sum_{j=1}^{\infty} |a_j + b_j|^p\right)^{1/p} \le \left(\sum_{j=1}^{\infty} |a_j|^p\right)^{1/p} + \left(\sum_{j=1}^{\infty} |b_j|^p\right)^{1/p}.$$

Moreover, equality holds if and only if the following conditions hold:

(i) $p = 1$*: for each* $j \in \mathbb{Z}_{>0}$ *there exists* $\alpha_j, \beta_j \in \mathbb{R}_{\ge 0}$*, not both zero, such that* $\alpha_j a_j = \beta_j b_j$*;*

(ii) $p \in (1, \infty)$*: there exists* $\alpha, \beta \in \mathbb{R}_{\ge 0}$*, not both zero, such that* $\alpha a_j = \beta b_j$ *for every* $j \in \mathbb{Z}_{>0}$*.*

Proof Let $p \in [1, \infty)$ and let $(a_j)_{j\in\mathbb{Z}_{>0}}, (b_j)_{j\in\mathbb{Z}_{>0}} \in \ell^p(\mathbb{F})$. For each $N \in \mathbb{Z}_{>0}$

$$\left(\sum_{j=1}^{N} |a_j + b_j|^p\right)^{1/p} \le \left(\sum_{j=1}^{N} |a_j|^p\right)^{1/p} + \left(\sum_{j=1}^{N} |b_j|^p\right)^{1/p} \le \|(a_j)_{j\in\mathbb{Z}_{>0}}\|_p + \|(b_j)_{j\in\mathbb{Z}_{>0}}\|_p$$

by Lemma 3.8.3. Therefore,

$$\|(a_j)_{j\in\mathbb{Z}_{>0}} + (b_j)_{j\in\mathbb{Z}_{>0}}\|_p = \lim_{N\to\infty}\left(\sum_{j=1}^{N}|a_j + b_j|^p\right)^{1/p} \le \|(a_j)_{j\in\mathbb{Z}_{>0}}\|_p + \|(b_j)_{j\in\mathbb{Z}_{>0}}\|_p.$$

This shows that $(a_j)_{j\in\mathbb{Z}_{>0}} + (b_j)_{j\in\mathbb{Z}_{>0}} \in \ell^p(\mathbb{F})$.

An argument similar to that used in the last part of Lemma 3.8.16 can be used to prove the last assertion of the lemma. ■

The integral version of Minkowski's inequality also holds in this case.

3.8.18 Lemma (Integral version of Minkowski's inequality) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, if* $\mathrm{p} \in [1,\infty)$*, if* $\mathrm{a}_{\mathrm{jk}} \in \mathbb{F}$*,* $\mathrm{j},\mathrm{k} \in \mathbb{Z}_{>0}$*, are such that* $(\mathrm{a}_{\mathrm{jk}})_{\mathrm{j}\in\mathbb{Z}_{>0}} \in \ell^{\mathrm{p}}(\mathbb{F})$ *for every* $\mathrm{k} \in \mathbb{Z}_{>0}$ *and* $(\mathrm{a}_{\mathrm{jk}})_{\mathrm{k}\in\mathbb{Z}_{>0}} \in \ell^{\mathrm{p}}(\mathbb{F})$ *for every* $\mathrm{j} \in \mathbb{Z}_{>0}$*, then*

Must prove when equality occurs

$$\left(\sum_{\mathrm{j}=1}^{\infty}\left|\sum_{\mathrm{k}=1}^{\infty}\mathrm{a}_{\mathrm{jk}}\right|^{\mathrm{p}}\right)^{1/\mathrm{p}} \le \sum_{\mathrm{k}=1}^{\infty}\left(\sum_{\mathrm{j}=1}^{\infty}|\mathrm{a}_{\mathrm{jk}}|^{\mathrm{p}}\right)^{1/\mathrm{p}}.$$

Moreover, equality holds if and only if there exists $\mathrm{b}_{\mathrm{j}},,\mathrm{c}_{\mathrm{k}} \in \mathbb{F}$*,* $\mathrm{j},\mathrm{k} \in$]integerp*, such that* $\mathrm{a}_{\mathrm{jk}} = \mathrm{b}_{\mathrm{j}}\mathrm{c}_{\mathrm{k}}$*.*

Proof For $p = 1$ we have

$$\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right| \le \sum_{j=1}^{\infty}\left(\sum_{k=1}^{\infty}|a_{jk}|\right) = \sum_{k=1}^{\infty}\left(\sum_{j=1}^{\infty}|a_{jk}|\right),$$

giving the result in this case.

Now let $p \in (1,\infty)$. Here we compute

$$\begin{aligned}\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right|^p &= \sum_{j=1}^{\infty}\left(\left|\sum_{k=1}^{\infty}a_{jk}\right|^{p-1}\right)\left(\left|\sum_{l=1}^{\infty}a_{jl}\right|\right)\\ &\le \sum_{j=1}^{\infty}\left(\sum_{l=1}^{\infty}\left(|a_{jl}|\left|\sum_{k=1}^{\infty}a_{jk}\right|^{p-1}\right)\right)\\ &= \sum_{l=1}^{\infty}\left(\sum_{j=1}^{\infty}\left(|a_{jl}|\left|\sum_{k=1}^{\infty}a_{jk}\right|^{p-1}\right)\right),\end{aligned}$$

swapping the order of summation in the last step. Now let $p' = \frac{p}{p-1}$ be the conjugate index. Now, by Hölder's inequality,

$$\begin{aligned}\sum_{j=1}^{\infty}\left(|a_{jl}|\left|\sum_{k=1}^{\infty}a_{jk}\right|^{p-1}\right) &\le \left(\sum_{j=1}^{\infty}|a_{jl}|^{\infty}\right)^{1/p}\left(\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right|^{p'(p-1)}\right)^{1/p'}\\ &= \left(\sum_{j=1}^{\infty}|a_{jl}|^{p}\right)^{1/p}\left(\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right|^{p}\right)^{1/p'}.\end{aligned}$$

Substituting this last relation into the preceding equation yields

$$\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right|^p \leq \sum_{l=1}^{\infty}\left(\left(\sum_{j=1}^{\infty}|a_{jl}|^p\right)^{1/p}\left(\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right|^p\right)^{1/p'}\right)$$
$$= \left(\sum_{l=1}^{\infty}\left(\sum_{j=1}^{\infty}|a_{jl}|^p\right)^{1/p}\right)\left(\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right|^p\right)^{1/p'}.$$

Now we note that the lemma is obviously true when

$$\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right|^p = 0.$$

So we suppose that this quantity is nonzero and divide the above-derived inequality

$$\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right|^p \leq \left(\sum_{l=1}^{\infty}\left(\sum_{j=1}^{\infty}|a_{jl}|^p\right)^{1/p}\right)\left(\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right|^p\right)^{1/p'}$$

by

$$\left(\sum_{j=1}^{\infty}\left|\sum_{k=1}^{\infty}a_{jk}\right|^p\right)^{1/p'},$$

which gives the desired inequality after noting that p' is conjugate to p. ■

Now we can prove that $\ell^p(\mathbb{F})$ is a Banach space.

3.8.19 Theorem ($(\ell^p(\mathbb{F}), \|\cdot\|_p)$ is a Banach space) *If $\mathbb{F} \in \mathbb{R}, \mathbb{C}\}$ and if $p \in [1, \infty)$ then $(\ell^p(\mathbb{F}), \|\cdot\|_p)$ is an $\mathbb{F}$-Banach space.*

Proof Let us first verify that $\ell^p(\mathbb{F})$ is a subspace. We first consider the case of $p = 1$. Let $(a_j)_{j\in\mathbb{Z}_{>0}}, (b_j)_{j\in\mathbb{Z}_{>0}} \in \ell^1(\mathbb{F})$. By Lemma 3.8.17 we have $(a_j + b_j)_{j\in\mathbb{Z}_{>0}} \in \ell^1(\mathbb{F})$. If $\alpha \in \mathbb{F}$ we have

$$\sum_{j=1}^{\infty}|\alpha a_j| = |\alpha|\sum_{j=1}^{\infty}|a_j|$$

by Proposition I-2.4.30. Thus $\alpha(a_j)_{j\in\mathbb{Z}_{>0}} \in \ell^1(\mathbb{F})$, which shows that $\ell^1(\mathbb{F})$ is a subspace of $\mathbb{F}^{\mathbb{Z}_{>0}}$.

By Lemma 3.8.17, if $(a_j)_{j\in\mathbb{Z}_{>0}}, (b_j)_{j\in\mathbb{Z}_{>0}} \in \ell^p(\mathbb{F})$ then $(a_j + b_j)_{j\in\mathbb{Z}_{>0}} \in \ell^p(\mathbb{F})$. It is easy to see, just as for the case of $p = 1$, that $\alpha(a_j)_{j\in\mathbb{Z}_{>0}} \in \ell^p(\mathbb{F})$ if $\alpha \in \mathbb{F}$ and if $(a_j)_{j\in\mathbb{Z}_{>0}} \in \ell^p(\mathbb{F})$, $p \in (1, \infty)$.

As we have shown the triangle inequality for $\|\cdot\|_p$ already in Lemma 3.8.17, and since the other norm properties for $\|\cdot\|_p$ hold trivially, it follows that $\ell^p(\mathbb{F})$ is a normed vector space. It remains to show that it is complete. Let $((a_j^{(l)})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\ell^p(\mathbb{F})$. We claim that the sequence $(a_j^{(l)})_{l\in\mathbb{Z}_{>0}}$ is a Cauchy sequence for each $j \in \mathbb{Z}_{>0}$. For every $\epsilon \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{Z}_{>0}$ such that

$$\left\|(a_j^{(k)})_{j\in\mathbb{Z}_{>0}} - (a_j^{(l)})_{j\in\mathbb{Z}_{>0}}\right\|_p = \left(\sum_{j=1}^{\infty}\left|a_j^{(k)} - a_j^{(l)}\right|^p\right)^{1/p} < \epsilon.$$

Now let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that

$$\left\|(a_j^{(k)})_{j\in\mathbb{Z}_{>0}} - (b_j^{(l)})_{j\in\mathbb{Z}_{>0}}\right\|_p < \epsilon.$$

Then

$$\left|a_j^{(k)} - a_j^{(l)}\right|^p \leq \sum_{j=1}^{\infty}\left|a_j^{(k)} - a_j^{(l)}\right|^p < \epsilon^p,$$

giving $(a_j^{(l)})_{l\in\mathbb{Z}_{>0}}$ as a Cauchy sequence. Denote its limit by $a_j \in \mathbb{F}$. We next claim that $(a_j^{(l)})_{l\in\mathbb{Z}_{>0}}$ converges to $(a_j)_{j\in\mathbb{Z}_{>0}}$ in $\ell^p(\mathbb{F})$. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that

$$\left\|(a_j^{(l)})_{j\in\mathbb{Z}_{>0}} - (a_j^{(k)})_{j\in\mathbb{Z}_{>0}}\right\|_p < \tfrac{\epsilon}{2}, \qquad l,k \geq N.$$

For $n \in \mathbb{Z}_{>0}$ the sequence $((a_j^{(l)})_{j=1}^n)_{l\in\mathbb{Z}}$ converges to $(a_j)_{j=1}^n$ in $\mathbb{F}^n$ with respect to the norm $\|\cdot\|_p$ by Theorem 3.3.3. Thus there exists $N' \in \mathbb{Z}_{>0}$ such that

$$\left(\sum_{j=1}^{n}\left|a_j^{(k)} - a_j\right|^p\right)^{1/p} < \frac{\epsilon}{2}, \qquad k \geq N'.$$

Then, for $k \geq \max\{N, N'\}$,

$$\begin{aligned}\left(\sum_{j=1}^{n}\left|a_j^{(l)} - a_j\right|^p\right)^{1/p} &\leq \left(\sum_{j=1}^{n}\left|a_j^{(l)} - a_j^{(k)}\right|^p\right)^{1/p} + \left(\sum_{j=1}^{n}\left|a_j^{(k)} - a_j\right|^p\right)^{1/p}\\ &\leq \left\|(a_j^{(l)})_{j\in\mathbb{Z}_{>0}} - (a_j^{(k)})_{j\in\mathbb{Z}_{>0}}\right\|_p + \left(\sum_{j=1}^{n}\left|a_j^{(k)} - a_j\right|^p\right)^{1/p} < \epsilon.\end{aligned}$$

Now we have

$$\left\|(a_j^{(l)})_{j\in\mathbb{Z}_{>0}} - (a_j)_{j\in\mathbb{Z}_{>0}}\right\|_p = \lim_{n\to\infty}\left(\sum_{j=1}^{n}\left|a_j^{(l)} - a_j\right|^p\right)^{1/p} \leq \epsilon.$$

This gives convergence of $(a_j^{(l)})_{l\in\mathbb{Z}_{>0}}$ to $(a_j)_{j\in\mathbb{Z}_{>0}}$ in $\ell^p(\mathbb{F})$, as desired. ■

Let us show that, unlike $\ell^\infty(\mathbb{F})$, the Banach spaces $\ell^p(\mathbb{F})$, $p \in [1, \infty)$, have the property of being separable.

3.8.20 Proposition ($\ell^p(\mathbb{F})$ is separable for $p \in [1,\infty)$) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if $p \in [1,\infty)$ then the Banach space $(\ell^p(\mathbb{F}), \|\cdot\|_p)$ is separable.*

Proof We recall the definition of $\mathscr{D}_q(\mathbb{F})$ from the proof of Proposition 3.8.13 for $q \in \mathbb{Q}$. There we showed that $\mathscr{D}(\mathbb{F})$ was countable. We will show that $\mathscr{D}_0(\mathbb{R})$ is dense in $\ell^p(\mathbb{F})$. It is clear that $\mathscr{D}_0(\mathbb{F}) \subseteq \ell^p(\mathbb{F})$ for $p \in [1,\infty)$. To show that it is dense in $\ell^p(\mathbb{F})$ let $\epsilon \in \mathbb{R}_{>0}$ and let $(a_j)_{j\in\mathbb{Z}_{>0}} \in \ell^p(\mathbb{F})$. Let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\left(\sum_{j=N+1}^{\infty} |a_j|^p\right)^{1/p} < \frac{\epsilon}{2}.$$

Now let $q_1, \dots, q_N \in \mathbb{Q}$ be such that

$$\left(\sum_{j=1}^{N} |a_j - q_j|^p\right)^{1/p} < \frac{\epsilon}{2}.$$

Then, taking $q_j = 0$ for $j > N$,

$$\left\|(a_j)_{j\in\mathbb{Z}_{>0}} - (q_j)_{j\in\mathbb{Z}_{>0}}\right\|_p = \left(\sum_{j=1}^{N} |a_j - q_j|^p\right)^{1/p} + \left(\sum_{j=N+1}^{\infty} |a_j|^p\right)^{1/p} < \epsilon.$$

Since $(q_j)_{j\in\mathbb{Z}_{>0}} \in \mathscr{D}_0(\mathbb{R})$ the result follows. ■

From this result we have the following useful corollary which finishes off Example 3.3.1–1.

3.8.21 Corollary ($\ell^p(\mathbb{F})$ is the completion of $\mathbb{F}_0^\infty$) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if $p \in [1,\infty)$ then $(\ell^p(\mathbb{F}), \|\cdot\|_p)$ is the completion of $(\mathbb{F}_0^\infty, \|\cdot\|_p)$.*

Proof Borrowing the notation from the proof of Proposition 3.8.20 we have

$$\mathscr{D}_0(\mathbb{F}) \subseteq \mathbb{F}_0^\infty \subseteq \ell^p(\mathbb{F})$$

from which we deduce, using the proof of Proposition 3.8.20, that

$$\ell^p(\mathbb{F}) = \operatorname{cl}(\mathscr{D}_0(\mathbb{F})) \subseteq \operatorname{cl}(\mathbb{F}_0^\infty) \subseteq \operatorname{cl}(\ell^p(\mathbb{F})) = \ell^p(\mathbb{F}).$$

Therefore, $\operatorname{cl}(\mathbb{F}_0^\infty) = \ell^p(\mathbb{F})$, as desired. ■

3.8.3 Banach spaces of direct sums

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$ be a family of nontrivial $\mathbb{F}$-Banach spaces. We shall generalise the situation of Proposition 3.3.4 as follows. For $p \in [1,\infty]$ we define a norm $\|\cdot\|_{I,p}$ on $\bigoplus_{i\in I} \mathsf{V}_i$ by

$$\|(v_i)_{i\in I}\|_{I,p} = \begin{cases} \left(\sum_{i\in I} \|v_i\|_i^p\right)^{1/p}, & p \in [1,\infty), \\ \sup\{\|v_i\|_i \mid i \in I\}, & p = \infty. \end{cases}$$

The argument in the proof of Proposition 3.3.4 used to show incompleteness of $(\bigoplus_{i\in I} \mathsf{V}_i, \|\cdot\|_{I,1})$ when I is infinite is easily adapted to the case when $p \in [1,\infty)$. Moreover, for $p = \infty$ one can also show that $(\bigoplus_{i\in I} \mathsf{V}_i, \|\cdot\|_{I,\infty})$ is incomplete; we leave this to the reader as Exercise 3.3.6.

Note that the situation we consider here is a generalisation of the spaces of sequences considered in detail in Section 3.8.2. Indeed, the situation in Section 3.8.2 occurs upon taking $I = \mathbb{Z}_{>0}$ and $\mathsf{V}_i = \mathbb{F}$ for each $i \in I$. For this reason, many of the particulars in this section go just as they do in Section 3.8.2, and we encourage the reader to understand this. It will be helpful in understanding the further generalisations we will make from families to functions.

That $\|\cdot\|_p$ is a norm for each $p \in [1,\infty)$ is not difficult to show, but we will show this as we go along in any event. In fact, we shall follow closely the course set out in Section 3.8.2. In keeping with this, we start off making the following definition.

3.8.22 Definition ($\ell^\infty(\bigoplus_{i\in I} \mathsf{V}_i)$) If $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ and if $((\mathsf{V}_i, \|\cdot\|_i)_{i\in I}$ is a family of normed $\mathbb{F}$-vector spaces then we define

$$\ell^\infty(\textstyle\bigoplus_{i\in I} \mathsf{V}_i) = \{(v_i)_{i\in I} \in \prod_{i\in I} \mathsf{V}_i \mid \sup\{\|v_i\|_i \mid i \in I\} < \infty\}$$

and define

$$\|(v_i)_{i\in I}\|_{I,\infty} = \sup\{\|v_i\|_i \mid i \in I\}$$

for $(v_i)_{i\in I} \in \ell^\infty(\bigoplus_{i\in I} \mathsf{V}_i$. •

It is evident (and see Exercise 3.8.5) that it is necessary that each of the normed vector spaces V_i be a Banach space if $\ell^\infty(\bigoplus_{i\in I} \mathsf{V}_i)$ is to be a Banach space. Moreover, this is sufficient.

3.8.23 Theorem ($(\ell^\infty(\bigoplus_{i\in I} \mathsf{V}_i), \|\cdot\|_{I,\infty})$ is a Banach space) *Let $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ and let $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$ be a family of $\mathbb{F}$-Banach spaces. Then $(\ell^\infty(\bigoplus_{i\in I} \mathsf{V}_i), \|\cdot\|_{I,\infty})$ is an $\mathbb{F}$-Banach space.*

Proof The only not entirely trivial norm property to verify for $\|\cdot\|_{I,\infty}$ is the triangle inequality:

$$\begin{aligned}
\|(u_i)_{i\in I} + (v_i)_{i\in I}\|_{I,\infty} &= \sup\{\|u_i + v_i\|_i \mid i \in I\} \\
&\le \sup\{\|u_i\|_i + \|v_i\|_i \mid i \in I\} \\
&= \sup\{\|u_i\|_i \mid i \in I\} + \sup\{\|v_i\|_i \mid i \in I\} \\
&= \|(u_i)_{i\in I}\|_\infty + \|(v_i)_{i\in I}\|_\infty,
\end{aligned}$$

where we have used Proposition I-2.2.27.

Now let us verify that $(\ell^\infty(\bigoplus_{i\in I} \mathsf{V}_i), \|\cdot\|_{I,\infty})$ is complete. We let $((v_i^{(l)})_{i\in I})_{l\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\ell^\infty(\bigoplus_{i\in I} \mathsf{V}_i)$. We claim that, for each $i \in I$, $(v_i^{(l)})_{l\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in V_i. To see this, let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\left\|(v_i^{(l)})_{i\in I} - (v_i^{(k)})_{i\in I}\right\|_{I,\infty} < \epsilon$$

for $k, l \geq N$. Then, by definition of $\|\cdot\|_{I,\infty}$,

$$\left\|v_i^{(l)} - v_i^{(k)}\right\|_i < \epsilon$$

for $k, l \geq N$ and for $i \in I$. Thus $(v_i^{(l)})_{l\in\mathbb{Z}_{>0}}$ is indeed a Cauchy sequence, and so converges to some $v_i \in \mathsf{V}_i$. We now claim that the sequence $((v_i^{(l)})_{i\in I})_{l\in\mathbb{Z}_{>0}}$ converges to $(v_i)_{i\in I}$. To see this, let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\left\|(v_i^{(l)})_{i\in I} - (v_i^{(k)})_{i\in I}\right\|_{I,\infty} < \tfrac{\epsilon}{2}$$

for $k, l \geq N$. Thus

$$\left\|v_i^{(l)} - v_i^{(l)}\right\|_i < \tfrac{\epsilon}{2}, \qquad k, l \geq N.$$

Now, for fixed $i \in I$, let $N' \in \mathbb{Z}_{>0}$ be sufficiently large that $\left\|v_i^{(k)} - v_i\right\|_i < \frac{\epsilon}{2}$ for $k \geq N'$. In this case, if $l \geq N$ and $k \geq \max\{N, N'\}$, we have

$$\left\|v_i^{(l)} - v_i\right\|_i \leq \left\|v_i^{(l)} - v_i^{(k)}\right\|_i + \left\|v_i^{(k)} - v_i\right\|_i < \epsilon.$$

Since this holds for each $i \in I$ we have

$$\left\|(v_i^{(l)})_{i\in I} - (v_i)_{i\in I}\right\|_{I,\infty} \leq \epsilon,$$

as desired. ■

Again sticking with the plan of Section 3.8.2, let us consider a subspace of $\ell^\infty(\bigoplus_{i\in I}\mathsf{V}_i)$ that is analogous to the subspace $\mathsf{c}_0(\mathbb{F})$ of $\ell^\infty(\mathbb{F})$.

3.8.24 Definition ($\mathsf{c}_0(\bigoplus_{i\in I}\mathsf{V}_i)$) If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if $((\mathsf{V}_i, \|\cdot\|_i)_{i\in I}$ is a family of normed $\mathbb{F}$-vector spaces then we define $\mathsf{c}_0(\bigoplus_{i\in I}\mathsf{V}_i)$ to be the elements $(v_i)_{i\in I} \in \ell^\infty(\bigoplus_{i\in I}\mathsf{V}_i)$ with the property that, for each $\epsilon \in \mathbb{R}_{>0}$ the set $\{i \in I \mid \|v_i\|_i \geq \epsilon\}$ is finite. •

As with the corresponding conclusion in Section 3.8.2, we have the following result.

3.8.25 Theorem ($\mathsf{c}_0(\bigoplus_{i\in I}\mathsf{V}_i)$ is a Banach space) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$ be a family of $\mathbb{F}$-Banach spaces. Then $(\mathsf{c}_0(\bigoplus_{i\in I}\mathsf{V}_i), \|\cdot\|_{I,\infty})$ is an $\mathbb{F}$-Banach space, and is moreover the completion of $\bigoplus_{i\in I}\mathsf{V}_i$ with respect to the norm $\|\cdot\|_{I,\infty}$.*

Proof Let $((v_i^{(l)})_{i\in I})_{l\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathsf{c}_0(\bigoplus_{i\in I}\mathsf{V}_i)$. By Theorem 3.8.23 this means that the sequence converges to $(v_i)_{i\in I} \in \ell^\infty(\bigoplus_{i\in I}\mathsf{V}_i)$. We next show that $(v_i)_{i\in I} \in \mathsf{c}_0(\bigoplus_{i\in I}\mathsf{V}_i)$. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\left\|v_i - v_i^{(k)}\right\|_i < \tfrac{\epsilon}{2}, \qquad k \geq N,\ i \in I.$$

For fixed $k \geq N$ let $J \subseteq I$ be a finite set such that $\left\|v_i^{(k)}\right\| < \frac{\epsilon}{2}$ for each $i \in I \setminus J$. Then, for $i \in I \setminus J$,

$$\|v_i\|_i \leq \left\|v_i - v_i^{(k)}\right\|_i + \left\|v_i^{(k)}\right\|_i < \epsilon,$$

which completes the proof that $\mathsf{c}_0(\bigoplus_{i\in I}\mathsf{V}_i)$ is a Banach space.

To see that $\mathsf{c}_0(\bigoplus_{i\in I}\mathsf{V}_i)$ is the completion of $\bigoplus_{i\in I}\mathsf{V}_i$, let $\epsilon\in\mathbb{R}_{>0}$ and let $(v_i)_{i\in I}\in\mathsf{c}_0(\bigoplus_{i\in I}\mathsf{V}_i)$. Let $J\subseteq I$ be a finite set such that $\|v_i\|_i<\epsilon$ for each $i\in I\setminus J$. Then define $(u_i)_{i\in I}\in\bigoplus_{i\in I}\mathsf{V}_i$ by

$$u_i=\begin{cases}v_i, & i\in J,\\ 0_{\mathsf{V}_i}, & i\in I\setminus J.\end{cases}$$

It then follows immediately that $\|(v_i)_{i\in I}-(u_i)_{i\in I}\|_{I,p}<\epsilon$, and so $\bigoplus_{i\in I}\mathsf{V}_i$ is dense in $\mathsf{c}_0(\bigoplus_{i\in I}\mathsf{V}_i)$. ■

Now let us turn to the case of $p\in[1,\infty)$.

3.8.26 Definition ($\ell^{\mathrm{p}}(\bigoplus_{\mathrm{i}\in\mathrm{I}}\mathsf{V}_{\mathrm{i}})$) Let $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and let $((\mathsf{V}_i,\|\cdot\|_i))_{i\in I}$ be a family of normed $\mathbb{F}$-vector spaces. For $p\in[1,\infty)$ we define

$$\ell^p(\textstyle\bigoplus_{i\in I}\mathsf{V}_i)=\left\{(v_i)_{i\in I}\in\prod_{i\in I}\mathsf{V}_i\;\middle|\;\displaystyle\sum_{i\in I}\|v_i\|_i^p<\infty\right\}$$

and

$$\|(v_i)_{i\in I}\|_{I,p}=\left(\sum_{i\in I}\|v_i\|_i^p\right)^{1/p},$$

for $(v_i)_{i\in I}\in\ell^p(\bigoplus_{i\in I}\mathsf{V}_i)$. •

Since the sum in the definition of $\|\cdot\|_{I,p}$ for $p\in[1,\infty)$ is over a general index set, it must be interpreted as in Section I-2.4.7 (see also Section 3.4.6).

We now have the expected result that $\ell^p(\bigoplus_{i\in I}\mathsf{V}_i)$ is a Banach space.

3.8.27 Theorem (($\ell^{\mathrm{p}}(\bigoplus_{\mathrm{i}\in\mathrm{I}}\mathsf{V}_{\mathrm{i}}),\|\cdot\|_{\mathrm{I,p}}$) is a Banach space) *Let $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and let $((\mathsf{V}_\mathrm{i},\|\cdot\|_\mathrm{i}))_{\mathrm{i}\in\mathrm{I}}$ be a family of $\mathbb{F}$-Banach spaces. Then $(\ell^\mathrm{p}(\bigoplus_{\mathrm{i}\in\mathrm{I}}\mathsf{V}_\mathrm{i}),\|\cdot\|_{\mathrm{I,p}})$ is an $\mathbb{F}$-Banach space, and is moreover the completion of $\bigoplus_{\mathrm{i}\in\mathrm{I}}\mathsf{V}_\mathrm{i}$ with respect to the norm $\|\cdot\|_{\mathrm{I,p}}$.*

Proof Let us first verify that $\ell^p(\bigoplus_{i\in I}\mathsf{V}_i)$ is a subspace. We first consider the case of $p=1$. Let $(u_i)_{i\in I},(v_i)_{i\in I}\in\ell^1(\bigoplus_{i\in I}\mathsf{V}_i)$ and note that for each finite subset $J\subseteq I$ we have

$$\sum_{j\in J}\|u_j+v_j\|_j\le\sum_{j\in J}\|u_j\|_j+\sum_{j\in J}\|v_j\|_j\le\|(u_i)_{i\in I}\|_{I,1}+\|(v_i)_{i\in I}\|_{I,1},$$

where we have used the triangle inequality for $\|\cdot\|_i$, $i\in I$. Therefore, by definition of sums over arbitrary index sets,

$$\|(u_i)_{i\in I}+(v_i)_{i\in I}\|_{I,1}=\sum_{i\in I}\|u_i+v_i\|_i\le\|(u_i)_{i\in I}\|_{I,1}+\|(v_i)_{i\in I}\|_{I,1}.\qquad(3.15)$$

This shows that $(u_i)_{i\in I}+(v_i)_{i\in I}\in\ell^1(\bigoplus_{i\in I}\mathsf{V}_i)$. If $\alpha\in\mathbb{F}$ we have

$$\sum_{i\in I}\|\alpha v_i\|_i=|\alpha|\sum_{i\in I}\|v_i\|_i$$

by Proposition I-2.4.30 (noting that the sum is over a countable subset of I). Thus $\alpha(v_i)_{i\in I} \in \ell^1(\bigoplus_{i\in I} \mathsf{V}_i)$, which shows that $\ell^1(\bigoplus_{i\in I} \mathsf{V}_i)$ is a subspace of $\prod_{i\in I} \mathsf{V}_i$.

Now let $p \in (1,\infty)$ and let $(u_i)_{i\in I}, (v_i)_{i\in I} \in \ell^p(\bigoplus_{i\in I} \mathsf{V}_i)$. For each finite subset $J \subseteq I$

$$\left(\sum_{j\in J} \|u_j + v_j\|_j^p\right)^{1/p} \leq \left(\sum_{j\in J} \|u_j\|_j^p\right)^{1/p} + \left(\sum_{j\in J} \|v_j\|)_j^p\right)^{1/p} \leq \|(u_i)_{i\in I}\|_{I,p} + \|(v_i)_{i\in I}\|_{I,p}$$

by Lemma 3.8.3. Therefore,

$$\|(u_i)_{i\in I} + (v_i)_{i\in I}\|_{I,p} = \left(\sum_{i\in I} \|u_i + v_i\|_i^p\right)^{1/p} \leq \|(u_i)_{i\in I}\|_{I,p} + \|(v_i)_{i\in I}\|_{I,p}. \tag{3.16}$$

This shows that $(u_i)_{i\in I} + (v_i)_{i\in I} \in \ell^p(\bigoplus_{i\in I} \mathsf{V}_i)$. It is easy to see, just as for the case of $p = 1$, that $\alpha(v_i)_{i\in I} \in \ell^p(\bigoplus_{i\in I} \mathsf{V}_i)$ if $\alpha \in \mathbb{F}$ and if $(v_i)_{i\in I} \in \ell^p(\bigoplus_{i\in I} \mathsf{V}_i)$, $p \in (1,\infty)$.

As we have shown the triangle inequality for $\|\cdot\|_{I,p}$ already in (3.15) and (3.16), and since the other norm properties for $\|\cdot\|_{I,p}$ hold trivially, it follows that $\ell^p(\bigoplus_{i\in I} \mathsf{V}_i)$ is a normed vector space. It remains to show that it is complete. Let $((v_i^{(l)})_{i\in I})_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\ell^p(\bigoplus_{i\in I} \mathsf{V}_i)$. We claim that the sequence $(v_i^{(l)})_{l\in\mathbb{Z}_{>0}}$ is a Cauchy sequence for each $i \in I$. For every $\epsilon \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{Z}_{>0}$ such that

$$\left\|(v_i^{(k)})_{i\in I} - (v_i^{(l)})_{i\in I}\right\|_{I,p} = \left(\sum_{i\in I} \left\|v_i^{(k)} - v_i^{(l)}\right\|_i^p\right)^{1/p} < \epsilon.$$

Now let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that

$$\left\|(v_i^{(k)})_{i\in I} - (v_i^{(l)})_{i\in I}\right\|_{I,p} < \epsilon.$$

Then

$$\left\|v_i^{(k)} - v_i^{(l)}\right\|_i^p \leq \sum_{i\in I} \left\|v_i^{(k)} - v_i^{(l)}\right\|_i^p < \epsilon^p,$$

giving $(v_i^{(l)})_{l\in\mathbb{Z}_{>0}}$ as a Cauchy sequence. Denote its limit by $v_i \in \mathsf{V}_i$. We next claim that $(v_i^{(l)})_{l\in\mathbb{Z}_{>0}}$ converges to $(v_i)_{i\in I}$ in $\ell^p(\bigoplus_{i\in I} \mathsf{V}_i)$. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that

$$\left\|(v_i^{(l)})_{i\in I} - (v_i^{(k)})_{i\in I}\right\|_p < \tfrac{\epsilon}{2}, \qquad l,k \geq N.$$

For any finite subset $J \subseteq I$ we claim that the sequence $((v_j^{(l)})_{j\in J})_{l\in\mathbb{Z}}$ converges to $(v_j)_{j\in J}$ in $\bigoplus_{j\in J} \mathsf{V}_j$ with respect to the norm $\|\cdot\|_{J,p}$ defined by

$$\|(v_j)_{j\in J}\|_{J,p} = \left(\sum_{j\in J} \|v_j\|_j^p\right)^{1/p}.$$

This claim is proved for $p = 1$ in Proposition 3.3.4. The proof for $p \in (1, \infty)$ is exactly the same, save for notation. Thus there exists $N' \in \mathbb{Z}_{>0}$ such that

$$\left(\sum_{j\in J} \left\|v_j^{(k)} - v_j\right\|_j^p\right)^{1/p} < \frac{\epsilon}{2}, \qquad k \geq N'.$$

Then, for $k \geq \max\{N, N'\}$,

$$\begin{aligned}\left(\sum_{j\in J} \left\|v_j^{(l)} - v_j\right\|_j^p\right)^{1/p} &\leq \left(\sum_{j\in J} \left\|v_j^{(l)} - v_j^{(k)}\right\|_j^p\right)^{1/p} + \left(\sum_{j\in J} \left\|v_j^{(k)} - v_j\right\|_j^p\right)^{1/p} \\ &\leq \left\|(v_i^{(l)})_{i\in I} - (v_i^{(k)})_{i\in I}\right\|_{I,p} + \left(\sum_{j\in J} \left\|v_j^{(k)} - v_j\right\|_j^p\right)^{1/p} < \epsilon.\end{aligned}$$

Since this can be done for any finite set $J \subseteq I$ we have

$$\left\|(v_i^{(l)})_{i\in I} - (a_i)_{i\in I}\right\|_p \leq \epsilon.$$

This gives convergence of $(v_i^{(l)})_{l\in\mathbb{Z}_{>0}}$ to $(a_i)_{i\in I}$ in $\ell^p(\bigoplus_{i\in I} \mathsf{V}_i)$, as desired. ∎

Of significant interest is the case when I is finite. In this case, all of the Banach spaces $\ell^p(\bigoplus_{i\in I} \mathsf{V}_i)$, $p \in [1, \infty]$, and $\mathsf{c}_0(\bigoplus_{i\in I} \mathsf{V}_i)$ are the same and equal to $\bigoplus_{i\in I} \mathsf{V}_i$. In particular, $\bigoplus_{i\in I} \mathsf{V}_i$ is a Banach space if I is finite and if all of the normed vector spaces V_i, $i \in I$, are complete.

3.8.4 Banach spaces of continuous functions on $\mathbb{R}$

One way to think of this section is as giving a generalisation of the construction of $\ell^\infty(\mathbb{F})$ and its subspaces in Section 3.8.2. The generalisation is to functions on the real line from sequences, which can be thought of as functions on $\mathbb{Z}_{>0}$. For functions on the real line one has the possible property of continuity that one is compelled to keep track of.

We begin by providing the classes of continuous functions we will talk about. We recall from Definition II-1.2.49 that if $I \subseteq \mathbb{R}$ is an interval and if $A \subseteq I$ then $\mathrm{cl}_I(A) = \mathrm{cl}(A) \cap I$.

3.8.28 Definition ($\mathsf{C}^0(I;\mathbb{F})$, $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$, $\mathsf{C}^0_{\mathrm{bdd}}(I;\mathbb{F})$, $\mathsf{C}^0_0(I;\mathbb{F})$) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $I \subseteq \mathbb{R}$ be an interval.

(i) $\mathsf{C}^0(I;\mathbb{F}) = \{f : I \to \mathbb{F} \mid f \text{ is continuous}\}$.

(ii) If $f \in \mathsf{C}^0(I;\mathbb{F})$ then the ***support*** of f is

$$\mathrm{supp}(f) = \mathrm{cl}_I(\{x \in I \mid f(x) \neq 0\}).$$

Is this the best place for this?

(iii) $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F}) = \{f \in \mathsf{C}^0(I;\mathbb{F}) \mid f \text{ has compact support}\}$.

(iv) $\mathsf{C}^0_0(I;\mathbb{F}) = \{f \in \mathsf{C}^0(I;\mathbb{F}) \mid$ for every $\epsilon \in \mathbb{R}_{>0}$ there exists a compact set $K \subseteq I$ such that $\{x \in I \mid |f(x)| \geq \epsilon\} \subseteq K\}$.

(v) $\mathsf{C}^0_{\mathrm{bdd}}(I;\mathbb{F}) = \{f \in \mathsf{C}^0(I;\mathbb{F}) \mid$ there exists $M \in \mathbb{R}_{>0}$ such that $|f(x)| \leq M$ for all $x \in I\}$. •

One should be a little careful about the meaning of compact support when I is not closed. For example, the function $f \in \mathsf{C}^0_{\mathrm{bdd}}((0,1];\mathbb{F})$ defined by $f(x) = 1$ does not have compact support since its support is $(0,1]$.

We first understand the case when $I = \mathbb{R}$. In this case, one can verify that

$$\mathsf{C}^0_0(\mathbb{R};\mathbb{F}) = \left\{ f \in \mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F}) \;\middle|\; \lim_{|x|\to\infty} |f(x)| = 0 \right\} \tag{3.17}$$

(this is Exercise 3.8.6). Thus $\mathsf{C}^0_0(\mathbb{R};\mathbb{F})$ consists of those functions which "die off" at infinity.

Clearly

$$\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F}) \subset \mathsf{C}^0_0(\mathbb{R};\mathbb{F}) \subset \mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F}) \subset \mathsf{C}^0(\mathbb{R};\mathbb{F}). \tag{3.18}$$

For $I = \mathbb{R}$ the vector space $\mathsf{C}^0(I;\mathbb{F})$ is too large to be of interest for the purposes of the discussion here. This is simply because continuous functions on $\mathbb{R}$ can be unbounded, and we wish to use a norm that is reliant on functions being bounded. Indeed, we define $\|\cdot\|_\infty$ by

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in \mathbb{R}\}$$

for $f \in \mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$. That this is a norm follows just as do the norm properties of Example 3.1.3–10.

Let us get the ball rolling by giving an important property of $\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F})$. This result should be thought of as being analogous to $(\mathbb{F}^\infty_0, \|\cdot\|_\infty)$ not being complete.

3.8.29 Proposition ($(\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F}), \|\cdot\|_\infty)$ is not complete) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ then $(\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F}), \|\cdot\|_\infty)$ is not complete.*

Proof Let us define a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F})$ by

$$f_j(x) = \begin{cases} \frac{1}{1+x^2}, & x \in [-j, j], \\ 0, & \text{otherwise}. \end{cases}$$

Let $\epsilon \in \mathbb{R}_{>0}$. Since $\lim_{x\to\infty} \frac{1}{1+x^2} = 0$ it follows that there exists $N \in \mathbb{Z}_{>0}$ such that $\left|\frac{1}{1+x_1^2} - \frac{1}{1+x_2^2}\right| < \epsilon$ for every $x_1, x_2 \geq N$. It then holds that $|f_j(x) - f_k(x)| < \epsilon$ for every $j, k \geq N$ and for every $x \in \mathbb{R}$. This shows that $(f_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence. We next claim that this sequence does not converge. The argument used in the lemma in Example 3.3.1–2 can be adapted to show that if $g \in \mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$ is a function to which the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges then $g(x) = \frac{1}{1+x^2}$ for every $x \in \mathbb{R}$. In particular, the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ does not converge in $\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F})$, and so $\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F})$ is not complete. ■

With this in our back pocket let us proceed in a manner entirely analogous to what we did in Section 3.8.2 in looking at $\ell^\infty(\mathbb{F})$ and its subspaces. Here the key observation is the following fairly obvious translation from the language of Section I-3.6.2 to the current language of convergence in normed vector spaces.

3.8.30 Proposition (Characterisation of convergence in $(\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F}), \|\cdot\|_\infty)$) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and if* $(f_j)_{j\in\mathbb{Z}_{>0}}$ *is a sequence in* $\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$ *then the following statements are equivalent:*

(i) *the sequence* $(f_j)_{j\in\mathbb{Z}_{>0}}$ *converges uniformly to* $f \in \mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$*;*

(ii) *the sequence* $(f_j)_{j\in\mathbb{Z}_{>0}}$ *converges to* $f \in \mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$ *with respect to the norm* $\|\cdot\|_\infty$*.*

Proof This just follows directly from the definitions of each sort of convergence. If the reader does not see this, they ought to convince themselves that this is the case. ■

The following theorem is now fairly easily proved, given what we already did in Section I-3.6.2.

3.8.31 Theorem ($(\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F}), \|\cdot\|_\infty)$ is a Banach space) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *then* $(\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F}), \|\cdot\|_\infty)$ *is an* $\mathbb{F}$*-Banach space.*

Proof Let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$. By Theorem I-3.6.8 it follows that this sequence converges to a function $f \in \mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$, and so the theorem follows. complex version?

■

As with $\ell^\infty(\mathbb{F})$, $\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$ is not separable.

3.8.32 Proposition ($\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$ is not separable) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *then* $(\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F}), \|\cdot\|_\infty)$ *is not separable.*

Proof Define a function $g_0 \colon \mathbb{R} \to \mathbb{F}$ by

$$g_0(x) = \begin{cases} 1 + x, & x \in [-\frac{1}{2}, 0], \\ 1 - x, & x \in (0, \frac{1}{2}], \\ 0, & \text{otherwise.} \end{cases}$$

Then let $\mathscr{U}$ be the collection of functions $f \in \mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$ of the form

$$f(x) = \sum_{j\in\mathbb{Z}_{>0}} (-1)^{k_j} g_0(x - j)$$

where $(k_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence in $\{0, 1\}$. The reader ought to sketch the graph of a typical function in $\mathscr{U}$ to understand what they are doing. Upon doing this it will be clear that, if $f \in \mathscr{U}$ then $\|f\|_\infty = 1$ and if $f_1, f_2 \in \mathscr{U}$ are distinct then $\|f_1 - f_2\|_\infty = 2$. The remainder of the proof follows the proof of Proposition 3.8.10, but we give it here for completeness.

Note that there are as many distinct functions in $\mathscr{U}$ as there are maps from $\mathbb{Z}_{>0}$ into $\{0, 1\}$. Thus $\mathrm{card}(\mathscr{U}) = 2^{\aleph_0}$. It then follows from Exercises I-1.7.4, I-1.7.5, and I-2.1.4

that $\mathscr{U}$ is uncountable. By Exercise 3.1.3 we have

$$\begin{aligned} & |\|g\|_\infty - \|f\|_\infty| \le \|g - f\|_\infty \\ \implies & |\|g\|_\infty - 1| \le 1 \\ \implies & \|g\|_\infty \le 2 \end{aligned}$$

for $f \in \mathscr{U}$. Thus $\mathsf{B}(1, f) \subseteq \mathsf{B}(2, 0_{\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})})$ for each $f \in \mathscr{U}$. If $f, g \in \mathscr{U}$ are distinct, and $\alpha \in \mathsf{B}(1, f)$ and $\beta \in \mathsf{B}(1, g)$ then

$$\|\alpha - g\|_\infty \ge |\|\alpha - f\|_\infty - \|f - g\|_\infty| \ge 2$$

using Proposition 1.1.3. Thus $\alpha \notin \mathsf{B}(1, g)$. One similarly shows that $\beta \notin \mathsf{B}(1, f)$. This shows that $\mathsf{B}(2, 0_{\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})})$ contains the collection

$$\{\mathsf{B}(1, f) \mid f \in \mathscr{U}\}$$

of disjoint open balls. In particular, if $(g_j)_{j\in\mathbb{Z}_{>0}}$ is any countable subset of $\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$ then there is a countable subset $(f_\alpha)_{\alpha\in A}$ of $\mathscr{U}$ in which are contained all of the functions $(g_j)_{j\in\mathbb{Z}_{>0}}$. Note that

$$\mathrm{cl}((g_j)_{j\in\mathbb{Z}_{>0}}) \subseteq \cup_{\alpha\in A}\overline{\mathsf{B}}(1, f_\alpha).$$

Therefore, any of the set of balls

$$\{\mathsf{B}(1, f) \mid f \in \mathscr{U},\ f \ne f_\alpha,\ \alpha \in A\}$$

cannot lie in $\mathrm{cl}((g_j)_{j\in\mathbb{Z}_{>0}})$ which prohibits $(g_j)_{j\in\mathbb{Z}_{>0}}$ from being dense. ■

Next let us characterise the completion of $\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F})$. The following result is entirely analogous to Corollary 3.8.14 which asserts that $\mathsf{c}_0(\mathbb{F})$ is the completion of $\mathbb{F}^\infty_0$.

3.8.33 Theorem ($(\mathsf{C}^0_0(\mathbb{R};\mathbb{F}), \|\cdot\|_\infty)$ is a Banach space) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ then $\mathsf{C}^0_0(\mathbb{R};\mathbb{F})$ is an $\mathbb{F}$-Banach space, and moreover is the completion of $(\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F}), \|\cdot\|_\infty)$.*

Proof We first make the observation that $\mathsf{C}^0_0(\mathbb{R};\mathbb{F})$ is a subspace of $\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$. This follows from Propositions I-2.3.23 and I-2.3.29. Now suppose that $(f_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathsf{C}^0_0(\mathbb{R};\mathbb{F})$. By Theorem 3.8.31 there exists a function $f \in \mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$ such that $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges to f. We need only show that $f \in \mathsf{C}^0_0(\mathbb{R};\mathbb{F})$. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $|f(x) - f_j(x)| < \frac{\epsilon}{2}$ for all $x \in \mathbb{R}$ provided that $j \ge N$. Let $K \subseteq \mathbb{R}$ be a compact set such that $|f_N(x)| < \frac{\epsilon}{2}$ for $x \in \mathbb{R} \setminus K$. Then, for $x \in \mathbb{R} \setminus K$ we have

$$|f(x)| \le |f(x) - f_N(x)| + |f_N(x)| < \epsilon,$$

giving $f \in \mathsf{C}^0_0(\mathbb{R};\mathbb{F})$, as desired.

To show that $\mathsf{C}^0_0(\mathbb{R};\mathbb{F})$ is the completion of $\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F})$, let $f \in \mathsf{C}^0_0(\mathbb{R};\mathbb{F})$ and define $(f_j)_{j\in\mathbb{Z}_{>0}}$ by

$$f_j(x) = \begin{cases} f(x), & x \in [-j, j], \\ f(-j)(j + 1 + x), & x \in [-j - 1, -j), \\ f(j)(j + 1 - x), & x \in (j, j + 1], \\ 0, & \text{otherwise}. \end{cases}$$

We claim that this sequence converges to f. For $\epsilon \in \mathbb{R}_{>0}$ let $N \in \mathbb{Z}_{>0}$ have the property that $|f(x)| < \epsilon$ if $|x| \geq N$. Then we immediately have $|f(x) - f_j(x)| < \epsilon$ for $j \geq N$, giving the desired convergence, and showing that $\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F})$ is dense in $\mathsf{C}^0_0(\mathbb{R};\mathbb{F})$. ■

Just as $\mathsf{c}_0(\mathbb{F})$ is separable, so too is $\mathsf{C}^0_0(\mathbb{R};\mathbb{F})$.

3.8.34 Proposition ($\mathsf{C}^0_0(\mathbb{R};\mathbb{F})$ is separable) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *then* $(\mathsf{C}^0_0(\mathbb{R};\mathbb{F}), \|\cdot\|_\infty)$ *is separable.*

Proof For $N \in \mathbb{Z}_{>0}$ let us denote by $P_N(\mathbb{F})$ the set of functions $f\colon \mathbb{R} \to \mathbb{F}$ having the form

$$f(x) = \begin{cases} z_k x^k + \cdots + z_1 x + z_0, & x \in [-N, N], \\ (z_k N^k + \cdots + z_1 N + z_0)(N + 1 - x), & x \in (N, N+1), \\ ((-1)^k z_k N^k + \cdots - z_1 N + z_0)(N + 1 + x), & x \in (-N-1, -N), \\ 0, & |x| \geq N + 1, \end{cases}$$

where $k \in \mathbb{Z}_{\geq 0}$ and $z_0, z_1, \dots, z_k \in \mathbb{F}$ are rational if $\mathbb{F} = \mathbb{R}$ and whose real and imaginary parts are rational of $\mathbb{F} = \mathbb{C}$. Note that functions in $P_N(\mathbb{F})$ are continuous. Moreover, for each $N \in \mathbb{Z}_{>0}$ the set $P_N(\mathbb{F})$ is countable by Proposition I-1.7.16. Thus $\cup_{N \in \mathbb{Z}_{>0}} P_N(\mathbb{F})$ is also countable, again by Proposition I-1.7.16.

We claim that $\cup_{N \in \mathbb{Z}_{>0}} P_N(\mathbb{F})$ is dense in $\mathsf{C}^0_0(\mathbb{R};\mathbb{F})$. Indeed, let $f \in \mathsf{C}^0_0(\mathbb{R};\mathbb{F})$ and let $\epsilon \in \mathbb{R}_{>0}$. Let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $|f(x)| < \epsilon$ for $|x| \geq N$. By the Weierstrass Approximation Theorem, Theorem I-3.6.21, let $g \in P_N(\mathbb{F})$ be such that $|f(x) - g(x)| < \epsilon$ for $x \in [-N, N]$. Our construction of functions in $P_N(\mathbb{F})$ then ensures that $|f(x) - g(x)| < \epsilon$ for all $x \in \mathbb{R}$. ■

In the preceding discussion we have pointed out various analogies with constructions concerning sequences in Section 3.8.2. In Table 3.1 we summarise the

Table 3.1 The relationships between the objects in the left column are analogous to the relationships between the objects in the right column

Sequence space	Function space
$\mathbb{F}^\infty_0$	$\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{F})$
$\ell^\infty(\mathbb{F})$	$\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F})$
$\mathsf{c}_0(\mathbb{F})$	$\mathsf{C}^0_0(\mathbb{R};\mathbb{F})$

correspondences. The correspondences for the sequence spaces $\ell^p(\mathbb{F})$ for $p \in [1, \infty)$ are more complicated, and we present these in Table 3.2.

Having now somewhat understood the structure of the spaces $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$, $\mathsf{C}^0_0(I;\mathbb{F})$, and $\mathsf{C}^0_{\mathrm{bdd}}(I;\mathbb{F})$ when $I = \mathbb{R}$, let us turn to the case of a general interval. It is fairly easy to carry out the programme directly in this case, adapting the arguments above. However, it is also the case that we shall do this in some generality in Section 3.8.5. Therefore, we abbreviate the discussion somewhat, mostly only giving outlines of proofs and referring to the more general results for complete arguments.

First let us observe that Proposition 3.8.30 holds for arbitrary intervals.

3.8.35 Proposition (Characterisation of convergence in $(\mathsf{C}^0_{\mathrm{bdd}}(\mathbb{R};\mathbb{F}),\|\cdot\|_\infty)$) *If $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$, if $\mathrm{I}\subseteq\mathbb{R}$, and if $(\mathrm{f}_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence in $\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{I};\mathbb{F})$ then the following statements are equivalent:*

(i) the sequence $(\mathrm{f}_j)_{j\in\mathbb{Z}_{>0}}$ converges uniformly to $\mathrm{f}\in\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{I};\mathbb{F})$;

(ii) the sequence $(\mathrm{f}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $\mathrm{f}\in\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{I};\mathbb{F})$ with respect to the norm $\|\cdot\|_\infty$.

Proof As with Proposition 3.8.30, this follows directly from the definitions. ■

Now let us indicate that things are significantly more trivial for compact intervals than for general intervals.

3.8.36 Proposition (Continuous function spaces for compact intervals) *If $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and if $\mathrm{I}\subseteq\mathbb{R}$ is a compact interval, then*

$$\mathsf{C}^0_{\mathrm{cpt}}(\mathrm{I};\mathbb{F})=\mathsf{C}^0_0(\mathrm{I};\mathbb{F})=\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{I};\mathbb{F})=\mathsf{C}^0(\mathrm{I};\mathbb{F}).$$

Proof This is a consequence of (3.18) along with the fact that $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})=\mathsf{C}^0(I;\mathbb{F})$ since every closed subset of I is compact according to Corollary I-2.5.28. ■

For compact intervals this gives the following characterisation of their continuous functions as forming a particularly nice Banach space.

3.8.37 Corollary (Properties of continuous function spaces for compact intervals) *If $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and if $\mathrm{I}\subseteq\mathbb{R}$ is a compact interval, then $\mathsf{C}^0_{\mathrm{cpt}}(\mathrm{I};\mathbb{F})$, $\mathsf{C}^0_0(\mathrm{I};\mathbb{F})$, $\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{I};\mathbb{F})$, and $\mathsf{C}^0(\mathrm{I};\mathbb{F})$ are separable $\mathbb{F}$-Banach spaces with the norm $\|\cdot\|_\infty$.*

complex version?

Proof That these are Banach spaces follows from Theorem I-3.6.8 since there we showed that in $\mathsf{C}^0_{\mathrm{bdd}}(I;\mathbb{F})$ all Cauchy sequences converge. Separability follows from the Weierstrass Approximation Theorem, just as does Proposition 3.8.34. ■

Since $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$ is the smallest of the spaces we consider, let us characterise precisely when it is a Banach space.

3.8.38 Proposition (Completeness of $(\mathsf{C}^0_{\mathrm{cpt}}(\mathrm{I};\mathbb{F}),\|\cdot\|_\infty)$) *If $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and if $\mathrm{I}\subseteq\mathbb{R}$ is an interval, then $(\mathsf{C}^0_{\mathrm{cpt}}(\mathrm{I};\mathbb{F}),\|\cdot\|_\infty)$ is complete if and only if I is compact.*

Proof For the noncompleteness of $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$ when I is not compact, we consider two cases of intervals: $I=(0,1]$ and $I=[0,\infty)$. The proof for an arbitrary noncompact interval follows by a trivial modification of these two cases.

First we show that $\mathsf{C}^0_{\mathrm{cpt}}((0,1];\mathbb{F})$ is not complete. We consider a sequence of functions $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{C}^0_{\mathrm{cpt}}((0,1];\mathbb{F})$ defined by

$$f_j(x)=\begin{cases}0, & x\in(0,\frac{1}{j}],\\ 2\frac{jx-1}{j-2}, & x\in[\frac{1}{j},\frac{1}{2}],\\ 1, & x\in[\frac{1}{2},1].\end{cases}$$

The reader is encouraged to plot the graphs of a few of the functions in this sequence to see what they are doing. Upon doing this it is easy to see that the sequence converges pointwise, in fact uniformly, to the function $f\colon (0,1] \to \mathbb{F}$ defined by

$$f(x) = \begin{cases} x, & x \in (0, \frac{1}{2}], \\ 1, & x \in (\frac{1}{2}, 1]. \end{cases}$$

We leave the elementary formal verification of this to the reader. Thus the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges in the normed vector space $(\mathsf{C}^0_{\mathrm{bdd}}((0,1];\mathbb{F}), \|\cdot\|_\infty)$. It is, therefore, a Cauchy sequence. However, since f does not have compact support, the sequence does not converge in $\mathsf{C}^0_{\mathrm{cpt}}((0,1];\mathbb{F})$, giving the incompleteness of $\mathsf{C}^0_{\mathrm{cpt}}((0,1];\mathbb{F})$.

Now we show that $\mathsf{C}^0_{\mathrm{cpt}}([0,\infty);\mathbb{F})$ is not complete. Let us define a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{C}^0_{\mathrm{cpt}}([0,\infty);\mathbb{F})$ by

$$f_j(x) = \begin{cases} \frac{1}{1+x^2}, & x \in [0, j], \\ 0, & \text{otherwise}. \end{cases}$$

It then follows, just as in the proof of Proposition 3.8.29, that this is a Cauchy sequence that does not converge.

That $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$ is complete when I is compact is Proposition 3.8.36. ■

The bounded continuous functions on I form a Banach space.

3.8.39 Theorem (($\mathsf{C}^0_{\mathrm{bdd}}$(I; $\mathbb{F}$), $\|\cdot\|_\infty$) is a Banach space) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if $\mathrm{I} \subseteq \mathbb{R}$ is an interval then $(\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{I};\mathbb{F}), \|\cdot\|_\infty)$ is an $\mathbb{F}$-Banach space. This Banach space is separable if and only if I is compact.*

Proof While the first assertion follows from Theorem I-3.6.8 just as does Theorem 3.8.31, we give a complete self-contained proof here, since this is an important result for us. complex version

Let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathsf{C}^0_{\mathrm{bdd}}(I;\mathbb{F})$ and for $x \in I$ define $f(x) = \lim_{j\to\infty} f_j(x)$. This pointwise limit exists since $(f_j(x))_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathbb{R}$ (why?).

First we claim that for any $\epsilon > 0$ there exists $N \in \mathbb{Z}_{>0}$ such that $|f(x) - f_j(x)| < \epsilon$ for all $x \in I$ whenever $j \geq N$. Let $\epsilon \in \mathbb{R}_{>0}$ and let $x \in I$. Since $(f_j)_{j\in\mathbb{Z}_{>0}}$ is Cauchy there exists $N \in \mathbb{Z}_{>0}$ such that $|f_j(x) - f_k(x)| < \frac{\epsilon}{2}$. We may also find $N(x) \in \mathbb{Z}_{>0}$ such that $|f(x) - f_j(x)| < \frac{\epsilon}{2}$ for $j \geq N(x)$. Let $k = \max\{N, N(x)\}$. For $j \geq N$ we then have

$$\begin{aligned} |f_j(x) - f(x)| &= |(f_j(x) - f_k(x)) + (f_k(x) - f(x))| \\ &\leq |f_j(x) - f_k(x)| + |f_k(x) - f(x)| \\ &< \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon, \end{aligned}$$

where we have used the triangle inequality. Note that this shows uniform convergence to f of the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$, and so convergence to f using the norm $\|\cdot\|_\infty$.

We next claim that f is bounded. To see this, for $\epsilon > 0$ let $N \in \mathbb{Z}_{>0}$ have the property that $\|f - f_N\|_\infty < \epsilon$. Then

$$|f(x)| \leq |f(x) - f_N(x)| + |f_N(x)| \leq \epsilon + \|f_N\|_\infty.$$

Since the expression on the right is independent of x, this gives the desired boundedness of f.

Finally we prove that the limit function f is continuous. As we showed above, for any $\epsilon > 0$ there exists $N \in \mathbb{Z}_{>0}$ such that $|f_N(x) - f(x)| < \frac{\epsilon}{3}$ for all $x \in I$. Now fix $x_0 \in I$, and consider the $N \in \mathbb{Z}_{>0}$ just defined. By continuity of f_N, there exists $\delta > 0$ such that if $x \in I$ satisfies $|x - x_0| < \delta$, then $|f_N(x) - f_N(x_0)| < \frac{\epsilon}{3}$. Then, for $x \in I$ satisfying $|x - x_0| < \delta$, we have

$$\begin{aligned}|f(x) - f(x_0)| &= |(f(x) - f_N(x)) + (f_N(x) - f_N(x_0)) + (f_N(x_0) - f(x_0))|\\ &\le |f(x) - f_N(x)| + |f_N(x) - f_N(x_0)| + |f_N(x_0) - f(x_0)|\\ &< \tfrac{\epsilon}{3} + \tfrac{\epsilon}{3} + \tfrac{\epsilon}{3} = \epsilon,\end{aligned}$$

where we have again used the triangle inequality. Since this argument is valid for any $x_0 \in I$, it follows that f is continuous.

Now let us turn to the separability of $\mathsf{C}^0_{\mathrm{bdd}}(I;\mathbb{F})$. The separability of $\mathsf{C}^0_{\mathrm{bdd}}(I;\mathbb{F})$ when I is compact is part of Corollary 3.8.37. If I is not compact, there are two cases to consider, when I is bounded and when I is not bounded. If I is not bounded a modification of the argument used in Proposition 3.8.32 can be used to show that $\mathsf{C}^0_{\mathrm{bdd}}(I;\mathbb{F})$ is not separable. Thus we need only consider the case when I is bounded but not compact.

We consider the case of $I = (0,1]$, the general case following, *mutatis mutandis*, from this. For $j \in \mathbb{Z}_{>0}$ define $g_j\colon (0,1] \to \mathbb{F}$ by

$$g_j(x) = \begin{cases} 2j(Herex(j+1) - 1), & x \in [\frac{1}{j+1}, \frac{1+2j}{2j(j+1)}],\\ 2(j+1)(1 - jx), & x \in (\frac{1+2j}{2j(j+1)}, \frac{1}{j}],\\ 0, & \text{otherwise.}\end{cases}$$

The reader would probably benefit from sketching the graph of this function to understand what the proof is achieving. We now let $\mathscr{U}$ be the collection of functions $f \in \mathsf{C}^0_{\mathrm{bdd}}((0,1];\mathbb{F})$ of the form

$$f(x) = \sum_{j\in\mathbb{Z}_{>0}} (-1)^{k_j} g_j(x).$$

One can now repeat the argument of Proposition 3.8.32 using this collection $\mathscr{U}$ of functions to show that $\mathsf{C}^0_{\mathrm{bdd}}((0,1];\mathbb{F})$ is not separable. ■

The generalisation of Theorem 3.8.33 also holds.

3.8.40 Theorem ($(\mathsf{C}^0_0(\mathrm{I};\mathbb{F}), \|\cdot\|_\infty)$ is a Banach space) *If $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ and if $\mathrm{I} \subseteq \mathbb{R}$ is an interval then $(\mathsf{C}^0_0(\mathrm{I};\mathbb{F}), \|\cdot\|_\infty)$ is a separable $\mathbb{F}$-Banach space, and moreover, is the completion of $(\mathsf{C}^0_{\mathrm{cpt}}(\mathrm{I};\mathbb{F}), \|\cdot\|_\infty)$.*

Proof A modification of the proof of Theorem 3.8.33 is easily made to give a direct proof; we leave the details to the reader. We also note that the present theorem also follows directly from the more general Theorem 3.8.43 below. ■

3.8.5 Banach spaces of continuous functions on metric spaces

We let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let (S, d) be a metric space and define

$$\mathsf{C}^0_{\mathrm{bdd}}(S; \mathbb{F}) = \{f : S \to \mathbb{F} \mid f \text{ is continuous and bounded}\}.$$

For $f \in \mathsf{C}^0_{\mathrm{bdd}}(S; \mathbb{F})$ we define

$$\|f\|_\infty = \sup\{|f(x)| \mid x \in S\}.$$

We claim that $(\mathsf{C}^0_{\mathrm{bdd}}(S, \mathbb{F}), \|\cdot\|_\infty)$ is a Banach space.

3.8.41 Theorem (($\mathsf{C}^0_{\mathrm{bdd}}$(S, $\mathbb{F}$), $\|\cdot\|_\infty$) is a Banach space) *$(\mathsf{C}^0_{\mathrm{bdd}}(S, \mathbb{F}), \|\cdot\|_\infty)$ is a Banach space.* Move this first part?

Proof First let us show that $\|\cdot\|_\infty$ is a norm. It is clear that $\|\lambda f\|_\infty = |\lambda| \|f\|_\infty$ for all $\lambda \in \mathbb{F}$ and $f \in \mathsf{C}^0_{\mathrm{bdd}}(S, \mathbb{F})$, and that $\|f\|_\infty \geq 0$ and $\|f\|_\infty = 0$ if and only if $f = 0$. We also compute, using Proposition I-2.2.27,

$$\begin{aligned} \|f + g\|_\infty &= \sup\{|f(x) + g(x)| \mid x \in S\} \\ &\leq \sup\{|f(x) + g(y)| \mid (x, y) \in S \times S\} \\ &\leq \sup\{|f(x)| \mid x \in S\} + \sup\{|g(y)| \mid y \in S\} \\ &= \|f\|_\infty + \|g\|_\infty \end{aligned}$$

for $f, g \in \mathsf{C}^0_{\mathrm{bdd}}(S, \mathbb{F})$. To show that $(\mathsf{C}^0_{\mathrm{bdd}}(S, \mathbb{F}), \|\cdot\|_\infty)$ is a complete normed vector space, we note that the norm topology is exactly the metric topology defined in general in Theorem 1.9.1. Since $(\mathbb{F}, |\cdot|)$ is complete, it then follows from Theorem 1.9.1 that $(\mathsf{C}^0_{\mathrm{bdd}}(S, \mathbb{F}), \|\cdot\|_\infty)$ is also complete. ■

Let us record some of the properties of the Banach space $\mathsf{C}^0_{\mathrm{bdd}}(S, \mathbb{F})$.

3.8.42 Proposition (Properties of $\mathsf{C}^0_{\mathrm{bdd}}$(S; $\mathbb{F}$)) finish

3.8.43 Theorem ((C^0_0(S; $\mathbb{F}$), $\|\cdot\|_\infty$) is a Banach space)

Proof ■

3.8.6 Banach spaces of continuous functions on locally compact topological spaces

3.8.7 Banach spaces of integrable functions on $\mathbb{R}$

In this section we look at an extremely important class of Banach spaces. In some sense, these are adaptations of the spaces of sequences considered in Section 3.8.2 to functions defined on intervals. These classes of functions play an essential rôle in Fourier analysis as we shall see in Chapters IV-5 and IV-6.

We begin, as we did with sequences, by considering functions that are, in the appropriate sense, bounded.

3.8.44 Definition ($\mathsf{L}^{(\infty)}(I;\mathbb{F})$) Let $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and let $I\subseteq\mathbb{R}$ be an interval. A measurable function $f\colon I\to\mathbb{F}$ is ***essentially bounded*** if there exists $M\in\mathbb{R}_{\geq 0}$ such that the set

$$\lambda(\{x\in I \mid |f(x)|>M\})=0.$$

The set of essentially bounded functions from I to $\mathbb{F}$ is denoted by $\mathsf{L}^{(\infty)}(I;\mathbb{F})$ and define

$$\|f\|_\infty=\inf\{M\in\mathbb{R}_{\geq 0} \mid \lambda(\{x\in I \mid |f(x)|>M\})=0\}$$

for $f\in\mathsf{L}^{(\infty)}(I;\mathbb{F})$. •

Let us give some initial properties of $\mathsf{L}^{(\infty)}(I;\mathbb{F})$.

3.8.45 Proposition (Properties of $(\mathsf{L}^{(\infty)}(I;\mathbb{F}),\|\cdot\|_\infty)$) *If $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and if $I\subseteq\mathbb{R}$ is an interval then $(\mathsf{L}^{(\infty)}(I;\mathbb{F}),\|\cdot\|_\infty)$ is a seminormed $\mathbb{F}$-vector space. Moreover, $\|f\|_\infty=0$ if and only if $f(x)=0$ for almost every $x\in I$.*

Proof The only seminorm property that is not completely trivial is the triangle inequality, so let us verify this. If $f\colon\phi\to\mathbb{R}$ is an arbitrary measurable function we denote

Move this defn

$$\operatorname{ess\,sup}\{\phi(x) \mid x\in I\}=\inf\{M\in\mathbb{R}_{\geq 0} \mid \lambda(\{x\in I \mid \phi(x)>M\}=0)\}.$$

If

$$Z_\phi=\{x\in I \mid \phi(x)>\operatorname{ess\,sup}\{\phi(x) \mid x\in I\}\}$$

and if Z is any set of measure zero containing Z_ϕ then

$$\operatorname{ess\,sup}\{\phi(x) \mid x\in I\}=\sup\{\phi(x) \mid x\in I\setminus Z\}.$$

Now let $f,g\in\mathsf{L}^{(\infty)}(I;\mathbb{F})$ and compute

$$\begin{aligned}
\|f+g\|_\infty &= \operatorname{ess\,sup}\{|f(x)+g(x)| \mid x\in I\}\\
&= \sup\{|f(x)+g(x)| \mid x\in I\setminus Z_{|f+g|}\}\\
&\leq \sup\{|f(x)|+|g(x)| \mid x\in I\setminus Z_{|f+g|}\}\\
&\leq \sup\{|f(x)| \mid x\in I\setminus Z_{|f+g|}\}+\sup\{|f(x)| \mid x\in I\setminus Z_{|f+g|}\}\\
&= \sup\{|f(x)| \mid x\in I\setminus (Z_{|f+g|}\cup Z_f)\}+\sup\{|f(x)| \mid x\in I\setminus (Z_{|f+g|}\cup Z_g)\}\\
&\leq \sup\{|f(x)| \mid x\in I\setminus Z_f\}+\sup\{|f(x)| \mid x\in I\setminus Z_g\}\\
&= \|f\|_\infty+\|g\|_\infty.
\end{aligned}$$

Thus $(\mathsf{L}^{(\infty)}(I;\mathbb{F}),\|\cdot\|_\infty)$ is a seminormed $\mathbb{F}$-vector space, as claimed.

The final assertion of the result is clear. ■

Now let

$$Z^\infty(I;\mathbb{F})=\{f\in\mathsf{L}^{(\infty)}(I;\mathbb{F}) \mid \|f\|_\infty=0\}.$$

By Theorem 3.1.8 we know that $\mathsf{L}^{(\infty)}(I;\mathbb{F})/Z^\infty(I;\mathbb{F})$,—i.e., the set of equivalence classes in $\mathsf{L}^{(\infty)}(I;\mathbb{F})$ where functions are equivalent if they agree almost everywhere—is a normed $\mathbb{F}$-vector space where the norm on the equivalence class $f+Z^\infty(I;\mathbb{F})$ is defined by

$$\|f+Z^\infty(I;\mathbb{F})\|_\infty=\|f\|_\infty;$$

it is convenient to use the same symbol for the norm.

3.8.46 Definition ($\mathsf{L}^\infty(I;\mathbb{F})$) For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and for an interval $I \subseteq \mathbb{R}$,

$$\mathsf{L}^\infty(I;\mathbb{F}) = \mathsf{L}^{(\infty)}(I;\mathbb{F})/Z^\infty(I;\mathbb{F}). \quad \bullet$$

Let us verify that $\mathsf{L}^\infty(I;\mathbb{F})$ is a Banach space.

3.8.47 Theorem (($\mathsf{L}^\infty(I;\mathbb{F}), \|\cdot\|_\infty$) is a Banach space) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if $\mathrm{I} \subseteq \mathbb{R}$ then $(\mathsf{L}^\infty(\mathrm{I};\mathbb{F}), \|\cdot\|_\infty)$ is an $\mathbb{F}$-Banach space.*

Proof For brevity, let us denote $[f] = f + Z^\infty(I;\mathbb{F})$ the equivalence class of $f \in \mathsf{L}^{(\infty)}(I;\mathbb{F})$ in $\mathsf{L}^\infty(I;\mathbb{F})$. We use the characterisation of completeness of Theorem 3.4.6. We let $\sum_{j=1}^\infty [f_j]$ be an absolutely convergent series. For $j \in \mathbb{Z}_{>0}$ define

$$Z_j = \{x \in I \mid |f_j(x)| > \|f_j\|_\infty\},$$

noting that $\lambda(Z_j) = 0$. For $x \notin \cup_{j=1}^\infty Z_j$ we have

$$\sum_{j=1}^\infty |f_j(x)| \le \sum_{j=1}^\infty \|f_j\|_\infty = \sum_{j=1}^\infty \|[f_j]\|_\infty < \infty$$

since $\sum_{j=1}^\infty [f_j]$ is absolutely convergent. This means that $\sum_{j=1}^\infty f_j(x)$ converges since absolute convergence in $\mathbb{F}$ implies convergence by Proposition I-2.4.3. Now define complex version

$$f(x) = \begin{cases} \sum_{j=1}^\infty f_j(x), & x \notin \cup_{j=1}^\infty Z_j \\ 0, & \text{otherwise.} \end{cases}$$

By Proposition 2.6.18 the function f is measurable. We then have

$$\begin{aligned} & f(x) - \sum_{j=1}^n f_j(x) = \sum_{j=n+1}^\infty f_j(x), \quad x \notin \cup_{j=1}^\infty Z_j \\ \implies\quad & \Big\| f - \sum_{j=1}^n f_j \Big\|_\infty \le \sum_{j=n+1}^\infty \|f_j\|_\infty \\ \implies\quad & \Big\| \Big[f - \sum_{j=1}^n f_j \Big] \Big\|_\infty \le \sum_{j=n+1}^\infty \|[f_j]\|_\infty \\ \implies\quad & \lim_{n\to\infty} \Big\| [f] - \sum_{j=1}^n [f_j] \Big\|_\infty \le \lim_{n\to\infty} \sum_{j=n+1}^\infty \|[f_j]\|_\infty = 0, \end{aligned}$$

thus giving convergence of $\sum_{j=1}^\infty [f_j]$ to $[f]$ in $\mathsf{L}^\infty(I;\mathbb{F})$. ■

3.8.48 Notation (Representing functions in $\mathsf{L}^\infty(I;\mathbb{F})$) While functions in $\mathsf{L}^\infty(I;\mathbb{F})$ are, by definition, equivalence classes of functions in $\mathsf{L}^{(\infty)}(I;\mathbb{F})$. The usual convention, however, is to in practice identify the equivalence class with one of its representatives. Most of the time the identification of an equivalence class with one of its representatives does not cause problems. However, there do arise instances where the distinction between these things becomes important, and so one must keep in mind what one is actually doing in writing "f" rather than "$f + Z^\infty(I;\mathbb{F})$." •

As with its brother $\ell^\infty(\mathbb{F})$, $\mathsf{L}^\infty(I;\mathbb{F})$ is not separable.

3.8.49 Proposition ($\mathsf{L}^\infty(I;\mathbb{F})$ is not separable) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if $I \subseteq \mathbb{R}$ is an interval with a nonempty interior then $\mathsf{L}^\infty(I;\mathbb{F})$ is not separable.*

Proof We shall only sketch the argument here as the details are already present in the proof of Proposition 3.8.32 and Theorem 3.8.39. If I is not bounded then an appropriate adaptation of the proof of the proof of Proposition 3.8.32 can be used to show that $\mathsf{L}^\infty(I;\mathbb{F})$ is not separable. If I is bounded then the idea in the proof of Theorem 3.8.39 can be used to give non-separability of $\mathsf{L}^\infty(I;\mathbb{F})$ in this case. Note that functions in $\mathsf{L}^\infty(I;\mathbb{F})$ are not required to be continuous and so the idea in the proof of Theorem 3.8.39 does indeed carry over to all bounded intervals, even those that are compact. ■

Let us relate convergence of sequences in $\mathsf{L}^\infty(I;\mathbb{F})$ with pointwise convergence. In fact, we can relate convergence in $\mathsf{L}^\infty(I;\mathbb{F})$ with a strong version of almost uniform convergence (see Definition 2.6.20(iv). In the statement of the result we denote $[f] = f + Z^\infty(I;\mathbb{F})$ for brevity.

3.8.50 Proposition (Uniform convergence and convergence in $\mathsf{L}^\infty(I;\mathbb{F})$) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $I \subseteq \mathbb{R}$ be an interval. Let $([f_j])_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{L}^\infty(I;\mathbb{F})$ converging to $[f] \in \mathsf{L}^\infty(I;\mathbb{F})$, and consider representatives $f_j \in [f_j]$, $j \in \mathbb{Z}_{>0}$, and $f \in [f]$. Then there exists a set $Z \subseteq I$ of measure zero such that the sequence $(f_j|(I \setminus Z))_{j\in\mathbb{Z}_{>0}}$ converges uniformly to $f|(I \setminus Z)$.*

Proof Let $([f_j])_{j\in\mathbb{Z}_{>0}}$ converge to $[f]$ in $\mathsf{L}^\infty(I;\mathbb{F})$ and let $f_j \in [f_j]$ and $f \in [f]$. Let

$$Z_j = \{x \in I \mid |f_j(x) - f(x)| > \|[f_j] - [f]\|_\infty\},$$

noting that Z_j, $j \in \mathbb{Z}_{>0}$, has measure zero. Let $Z = \cup_{j\in\mathbb{Z}_{>0}} Z_j$, noting that Z has measure zero. We claim that the assertion of the proposition holds for Z as defined. Indeed, let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that $\|[f_j] - [f]\|_\infty < \epsilon$ for $j \geq N$. Then, for $j \geq N$ and $x \in I \setminus Z$,

$$|f_j(x) - f(x)| \leq \|[f_j] - [f]\|_\infty < \epsilon,$$

giving the desired conclusion. ■

We have the following corollary, immediately from the definition of almost uniform convergence.

3.8.51 Corollary (Almost uniform convergence and convergence in $\mathsf{L}^\infty(I;\mathbb{F})$) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $I \subseteq \mathbb{R}$ be an interval. Let $([f_j])_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{L}^\infty(I;\mathbb{F})$ converging to $[f] \in \mathsf{L}^\infty(I;\mathbb{F})$, and consider representatives $f_j \in [f_j]$, $j \in \mathbb{Z}_{>0}$, and $f \in [f]$. Then $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges almost uniformly to f.*

The following example distinguishes between convergence in $\mathsf{L}^\infty(I;\mathbb{F})$ and almost uniform convergence.

3.8.52 Example (Almost uniform convergence does not imply convergence in $\mathsf{L}^\infty(\mathbf{I};\mathbb{F})$) Let $I = [-1,1]$ and define $f_j\colon [-1,1] \to \mathbb{R}$ by

$$f_j(x) = \begin{cases} 1, & j \in (-\frac{1}{j}, \frac{1}{j}), \\ 0, & \text{otherwise}. \end{cases}$$

Also define

$$f(x) = \begin{cases} 1, & x = 0, \\ 0, & \text{otherwise}. \end{cases}$$

We claim that $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges almost uniformly to f. Indeed, let $\epsilon, \delta \in \mathbb{R}_{>0}$. Take $N \in \mathbb{Z}_{>0}$ with $N > \frac{2}{\delta}$ and take $E_\delta = [-\frac{1}{N}, \frac{1}{N}]$. Then $|f(x) - f_j(x)| = 0$ for $j \geq N$ and $x \in I \setminus E_\delta$. This shows almost uniform convergence.

Next we claim that $([f_j])_{j\in\mathbb{Z}_{>0}}$ does not converge to $[f]$ in $\mathsf{L}^\infty(I;\mathbb{R})$. Let $\epsilon = \frac{1}{2}$. For $N > \frac{1}{\epsilon}$, $j \geq N$, and $x \in [-\frac{1}{N}, \frac{1}{N}]$, we have $|f(x) - f_j(x)| = 1 > \epsilon$. Thus $\|[f_j] - [f]\|_\infty > \epsilon$ for $j \geq N$, and this precludes convergence in $\mathsf{L}^\infty(I;\mathbb{R})$. •

Before we leave $\mathsf{L}^\infty(I;\mathbb{F})$ to talk about the spaces $\mathsf{L}^p(I;\mathbb{F})$ for $p \in [1,\infty)$ let us point out a possible source of confusion. We note that the Banach space $(\mathsf{L}^\infty(I;\mathbb{F}), \|(\|\cdot)_\infty)$ contains the Banach spaces $(\mathsf{C}^0_{\mathrm{bdd}}(I;\mathbb{F}); \|\cdot\|_\infty)$ and $\mathsf{C}^0_0(I;\mathbb{F})$ as a closed proper subspaces (they is a closed by Proposition 3.6.16 since it is complete). Thus $\mathsf{L}^\infty(I;\mathbb{F})$ is *not* the completion of these spaces. This is to be contrasted with the conclusion of Theorem 3.8.59 where we show that $\mathsf{L}^p(I;\mathbb{F})$ *is* the completion of a space of continuous functions when $p \in [1,\infty)$. This explains why the reader does not see $\mathsf{L}^\infty(I;\mathbb{F})$ in Tables 3.1 and 3.2.

Next we consider functions defined by their integrals. This is analogous to the sequence spaces $\ell^p(\mathbb{F})$, $p \in [1,\infty)$, being defined by their infinite sums.

3.8.53 Definition ($\mathsf{L}^{(p)}(\mathbf{I};\mathbb{F})$) Let $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$, let $p \in [1,\infty)$, and let $I \subseteq \mathbb{R}$ be an interval. Define a subspace $\mathsf{L}^{(p)}(I;\mathbb{F})$ of the measurable functions from I to $\mathbb{F}$ by

$$\mathsf{L}^{(p)}(I;\mathbb{F}) = \left\{ f\colon I \to \mathbb{F} \;\middle|\; f \text{ measurable}, \int_I |f|^p \,\mathrm{d}\lambda < \infty \right\}$$

and define

$$\|f\|_p = \left(\int_I |f|^p \,\mathrm{d}\lambda \right)^{1/p}$$

for $f \in \mathsf{L}^{(p)}(I;\mathbb{F})$. •

In the preceding definition it turns out to be crucial that the integral used is the Lebesgue integral. Indeed, many of the results we prove in this section simply do not hold if we instead attempt to use the Riemann integral. We shall, nonetheless, generally adopt the policy of writing the Lebesgue integral as $\int \mathrm{d}x$ rather than $\int \mathrm{d}\lambda$ for simplicity.

Let us give the analogues of Lemmata 3.8.16 and 3.8.17 in this setup.

3.8.54 Lemma (Hölder's inequality) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $\mathrm{I} \subseteq \mathbb{R}$ *be an interval, and let* $\mathrm{p} \in (1, \infty)$ *with* p' *defined by* $\frac{1}{\mathrm{p}} + \frac{1}{\mathrm{p}'} = 1$. *Then, for* $\mathrm{f} \in \mathsf{L}^{(\mathrm{p})}(\mathrm{I}; \mathbb{F})$ *and* $\mathrm{g} \in \mathsf{L}^{(\mathrm{p}')}(\mathrm{I}; \mathbb{F})$, $\mathrm{fg} \in \mathsf{L}^{(1)}(\mathrm{I}; \mathbb{F})$ *and*

$$\|\mathrm{fg}\|_1 \leq \|\mathrm{f}\|_{\mathrm{p}} \|\mathrm{g}\|_{\mathrm{p}'}.$$

Moreover, equality holds if and only if there exists $\alpha, \beta \in \mathbb{R}_{\geq 0}$, *not both zero, such that*

$$\alpha |\mathrm{f}(\mathrm{x})|^{\mathrm{p}} = \beta |\mathrm{g}(\mathrm{x})|^{\mathrm{p}'}, \qquad \textit{a.e. } \mathrm{x} \in \mathrm{I}.$$

Proof For $p, p' \in (1, \infty)$ satisfying $\frac{1}{p} + \frac{1}{p'} = 1$ we claim that for $x, y \in \mathbb{R}_{\geq 0}$ we have

$$xy \leq \frac{x^p}{p} + \frac{y^{p'}}{p'}.$$

This is trivial if either x or y are zero. So suppose that $x, y \in \mathbb{R}_{>0}$. Taking $\xi = \frac{x^p}{y^{p'}}$ we easily check that

$$xy \leq \frac{x^p}{p} + \frac{y^{p'}}{p'} \iff \xi^{1/p} \leq \frac{\xi}{p} + \frac{1}{p'}.$$

One can check using Theorem I-3.2.16 that the function

$$\xi \mapsto \frac{\xi}{p} + \frac{1}{p'} - \xi^{1/p}$$

has a minimum value of 0 attained at $\xi = 1$. Thus

$$\frac{\xi}{p} + \frac{1}{p'} - \xi^{1/p} \geq 0 \implies \quad \implies xy \leq \frac{x^p}{p} + \frac{y^{p'}}{p'},$$

as desired.

Now let us proceed with the proof. The result is clearly true if $\|f\|_p = 0$ or $\|g\|_{p'} = 0$. So we assume neither of these are true. For all $x \in I$ we have

$$|f(x)g(x)| \leq \frac{|f(x)|^p}{p} + \frac{|g(x)|^{p'}}{p'}.$$

Therefore, if $\|f\|_p = \|g\|_{p'} = 1$, we immediately have

$$\|fg\|_1 \leq \frac{1}{p} + \frac{1}{p'} = \|f\|_p \|g\|_{p'}.$$

In general we have

$$\|fg\|_1 = \|f\|_p \|g\|_{p'} \left\| \frac{f}{\|f\|_p} \frac{g}{\|g\|_{p'}} \right\|_1 \leq 1,$$

and the first part of the result follows.

If one chases through the argument above one sees that equality is achieved only when

$$|f(x)g(x)| \frac{|f(x)|^p}{p} + \frac{|g(x)|^{p'}}{p'}$$

for almost every $x \in I$. A tedious argument like that for the last part of Lemma 3.8.1, but replacing sums with integrals, shows that the above equality implies the final conclusion of the lemma. ∎

There is a version of Hölder's inequality for the case when $p = 1$, and we refer to Exercise 3.8.8 for this.

Let us prove the Minkowski inequality in this case.

3.8.55 Lemma (Minkowski's inequality) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $I \subseteq \mathbb{R}$ *be an interval, and let* $p \in [1, \infty)$*. Then, for* $f, g \in \mathsf{L}^{(p)}(I; \mathbb{F})$*, we have* $f + g \in \mathsf{L}^{(p)}(I; \mathbb{F})$ *and*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

Moreover, equality holds if and only if the following conditions hold:

(i) $p = 1$*: there exists nonnegative measurable functions* $\alpha, \beta\colon I \to \mathbb{R}_{\geq 0}$ *such that* $\alpha(x)f(x) = \beta(x)g(x)$ *and* $\alpha(x)$ *and* $\beta(x)$ *are not both zero for almost every* $\in I$*;*

(ii) $p \in (1, \infty)$*: there exists* $\alpha, \beta \in \mathbb{R}_{\geq 0}$*, not both zero, such that* $\alpha f(x) = \beta g(x)$ *for almost every* $x \in I$*.*

Proof For $p = 1$ we have

$$\|f + g\|_1 = \int_I |f(x) + g(x)|\, dx \leq \int_I |f(x)|\, dx + \int_I |g(x)|\, dx = \|f\|_1 + \|g\|_1.$$

The second assertion of the lemma for $p = 1$ follows from the fact, pointed out in the proof of Lemma 3.8.1, that $|a+b| = |a|+|b|$ for $a, b \in \mathbb{F}$ if and only if $\alpha a = \beta b$ for $\alpha, \beta \in \mathbb{R}_{\geq 0}$ not both zero. Note that the sets

$$A_f = \{x \in I \mid f(x) = 0\}, \quad A_g = \{x \in I \mid g(x) = 0\}, \quad A_{f,g} = \{x \in I \mid f(x)g(x) = 0\}$$

are measurable and so, therefore, are their complements. We then define $\alpha, \beta\colon I \to \mathbb{R}_{\geq 0}$ by

$$\alpha(x) = \begin{cases} g(x), & x \in I \setminus A_{f,g}, \\ g(x), & x \in A_{f,g} - A_f, \\ 0, & x \in A_g \end{cases}$$

and

$$\beta(x) = \begin{cases} f(x), & x \in I \setminus A_{f,g}, \\ 0, & x \in A_g, \\ f(x), & x \in A_{f,g} - A_g. \end{cases}$$

For $p \in (1, \infty)$ we let $\frac{1}{p} + \frac{1}{p'} = 1$. We then have

$$\left(|f(x) + g(x)|^{p-1}\right)^{p'} = |f(x) + g(x)|^p$$

from which we deduce that $|f + g|^{p-1} \in \mathsf{L}^{(p')}(I; \mathbb{F})$. Therefore, using Lemma 3.8.54,

$$\begin{aligned} \int_I |f(x) + g(x)|^p\, dx &\leq \int_I |f(x)||f(x) + g(x)|^{p-1}\, dx + \int_I |g(x)||f(x) + g(x)|^{p-1}\, dx \\ &\leq \|f\|_p \||f + g|^{p-1}\|_{p'} + \|g\|_p \||f + g|^{p-1}\|_{p'} \\ &= (\|f\|_p + \|g\|_p)\left(\int_I |f(x) + g(x)|^p\, dx\right)^{1/p'}, \end{aligned}$$

which implies that

$$\|f+g\|_p^{p-p/p'} \leq \|f\|_p + \|g\|_p,$$

provided that $\|f+g\|_p \neq 0$ (if it is zero, the result is trivial). The first part of the result follows since $p - p/p' = 1$. The second part of the result for $p \in (1,\infty)$ follows as does the second part of the proof of Lemma 3.8.1, replacing "for every $j \in \{1,\dots,n\}$" with "for almost every $x \in I$" and replacing "for some $j \in \{1,\dots,n\}$" with "for $x \in A$ with $A \subseteq I$ of positive measure." We leave the tedious details to the reader. ■

The following version of the Minkowski inequality is also useful.

Must prove when equality occurs

3.8.56 Lemma (Integral version of Minkowski inequality) *Let $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$, let $I, J \subseteq \mathbb{R}$ be intervals, and let $p \in [1,\infty)$. Let $f\colon I \times J \to \mathbb{F}$ have the property that $x \mapsto f(x,y)$ is in $\mathsf{L}^{(p)}(I;\mathbb{F})$ for almost every $y \in J$ and that $y \mapsto f(x,y)$ is in $\mathsf{L}^{(p)}(J;\mathbb{F})$ for almost every $x \in I$. Then, we have*

$$\left(\int_I \left|\int_J f(x,y)\,\mathrm{d}y\right|^p \mathrm{d}x\right)^{1/p} \leq \int_J \left(\int_I |f(x,y)|^p \mathrm{d}x\right)^{1/p} \mathrm{d}y.$$

Proof For $p = 1$ we have

$$\int_I \left|\int_J f(x,y)\,\mathrm{d}y\right| \mathrm{d}x \leq \int_I \left(\int_J |f(x,y)|\,\mathrm{d}y\right) \mathrm{d}x = \int_J \left(\int_I |f(x,y)|\,\mathrm{d}x\right) \mathrm{d}y,$$

giving the result in this case by Fubini's Theorem.

Now let $p \in (1,\infty)$. Here we compute

$$\begin{aligned}\int_I \left|\int_J f(x,y)\,\mathrm{d}y\right|^p \mathrm{d}x &= \int_I \left(\left(\left|\int_J f(x,y)\,\mathrm{d}y\right|^{p-1}\right)\left(\left|\int_J f(x,z)\,\mathrm{d}z\right|\right)\right) \mathrm{d}x \\ &\leq \int_I \left(\int_J \left(|f(x,z)| \left|\int_J f(x,y)\,\mathrm{d}y\right|^{p-1}\right) \mathrm{d}z\right) \mathrm{d}x \\ &= \int_J \left(\int_I \left(|f(x,z)| \left|\int_J f(x,y)\,\mathrm{d}y\right|^{p-1}\right) \mathrm{d}x\right) \mathrm{d}z\end{aligned}$$

using Fubini's Theorem in the last step. Now let $p' = \frac{p}{p-1}$ be the conjugate index. Now, by Hölder's inequality,

$$\begin{aligned}\int_I \left(|f(x,z)| \left|\int_J f(x,y)\,\mathrm{d}y\right|^{p-1}\right) \mathrm{d}x &\leq \left(\int_I |f(x,z)|^p\,\mathrm{d}x\right)^{1/p} \left(\int_I \left|\int_J f(x,y)\,\mathrm{d}y\right|^{p'(p-1)} \mathrm{d}x\right)^{1/p'} \\ &= \left(\int_I |f(x,z)|^p\,\mathrm{d}x\right)^{1/p} \left(\int_I \left|\int_J f(x,y)\,\mathrm{d}y\right|^{p} \mathrm{d}x\right)^{1/p'}.\end{aligned}$$

Substituting this last relation into the preceding equation yields

$$\begin{aligned}\int_I\left|\int_J f(x,y)\,\mathrm{d}y\right|^p\,\mathrm{d}x &\le \int_J\left(\left(\int_I|f(x,z)|^p\,\mathrm{d}x\right)^{1/p}\left(\int_I\left|\int_J f(x,y)\,\mathrm{d}y\right|^p\,\mathrm{d}x\right)^{1/p'}\right)\mathrm{d}z\\ &= \left(\int_J\left(\int_I|f(x,z)|^p\,\mathrm{d}x\right)^{1/p}\,\mathrm{d}z\right)\left(\int_I\left|\int_J f(x,y)\,\mathrm{d}y\right|^p\,\mathrm{d}x\right)^{1/p'}\end{aligned}$$

Now we note that the lemma is obviously true when

$$\int_I\left|\int_J f(x,y)\,\mathrm{d}y\right|^p\,\mathrm{d}x = 0.$$

So we suppose that this quantity is nonzero and divide the above-derived inequality

$$\int_I\left|\int_J f(x,y)\,\mathrm{d}y\right|^p\,\mathrm{d}x \le \left(\int_J\left(\int_I|f(x,z)|^p\,\mathrm{d}x\right)^{1/p}\,\mathrm{d}z\right)\left(\int_I\left|\int_J f(x,y)\,\mathrm{d}y\right|^p\,\mathrm{d}x\right)^{1/p'}$$

by

$$\left(\int_I\left|\int_J f(x,y)\,\mathrm{d}y\right|^p\,\mathrm{d}x\right)^{1/p'}$$

which gives the desired inequality after noting that p' is conjugate to p. ■

Now we can prove the basic fact about the spaces $\mathsf{L}^{(p)}(I;\mathbb{F})$.

3.8.57 Proposition (Properties of $(\mathsf{L}^{(p)}(I;\mathbb{F}),\|\cdot\|_\infty)$) *If* $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$, *if* $p\in[1,\infty)$, *and if* $I\subseteq\mathbb{R}$ *is an interval then* $(\mathsf{L}^{(p)}(I;\mathbb{F}),\|\cdot\|_\infty)$ *is a seminormed* $\mathbb{F}$*-vector space. Moreover,* $\|f\|_p=0$ *if and only if* $f(x)=0$ *for almost every* $x\in I$.

Proof That $\mathsf{L}^{(p)}(I;\mathbb{F})$ is a seminormed vector space follows from Lemma 3.8.55 which gives the triangle inequality; the other seminorm properties are clear. The final assertion is clear. ■

Now we proceed much as we did for $\mathsf{L}^{(\infty)}(I;\mathbb{F})$. That is, we define

$$Z^p(I;\mathbb{F}) = \{f\in\mathsf{L}^{(p)}(I;\mathbb{F}) \mid \|f\|_p=0\}$$

and note that, by Theorem 3.1.8, $\mathsf{L}^{(p)}(I;\mathbb{F})/Z^p(I;\mathbb{F})$ is a normed $\mathbb{F}$-vector space if we define the norm by

$$\|f+Z^p(I;\mathbb{F})\|_p = \|f\|_p.$$

This leads to the following definition.

3.8.58 Definition ($\mathsf{L}^p(I;\mathbb{F})$) For $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$, for $p\in[1,\infty)$, and for an interval $I\subseteq\mathbb{R}$,

$$\mathsf{L}^p(I;\mathbb{F}) = \mathsf{L}^{(p)}(I;\mathbb{F})/Z^p(I;\mathbb{F}).$$ •

We can prove that $\mathsf{L}^p(I;\mathbb{F})$ is a Banach space.

3.8.59 Theorem ($(\mathsf{L}^p(I;\mathbb{F}),\|\cdot\|_p)$ is a Banach space) *If $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$, if $p\in[1,\infty)$, and if $I\subseteq\mathbb{R}$ is an interval, then $(\mathsf{L}^p(I;\mathbb{F}),\|\cdot\|_p)$ is an $\mathbb{F}$-Banach space. Moreover, $\mathsf{L}^p(I;\mathbb{F})$ is isomorphic, as a normed vector space, to the completion of $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$.*

Proof For brevity let us denote $[f]=f+Z^p(I;\mathbb{F})$ for $f\in\mathsf{L}^{(p)}(I;\mathbb{F})$. We use the characterisation of completeness of Theorem 3.4.6. Let $\sum_{j=1}^\infty[f_j]$ be absolutely convergent. Define $g\colon I\to\mathbb{F}\cup\{\infty\}$ by

$$g(x)=\left(\sum_{j=1}^\infty|f_j(x)|\right)^p,$$

and note that Minkowski's inequality gives

$$\|g\|_1\le\sum_{j=1}^\infty\|f_j\|_p=\sum_{j=1}^\infty\|[f_j]\|_p<\infty$$

since $\sum_{j=1}^\infty[f_j]$ is absolutely convergent. Therefore, $g\in\mathsf{L}^{(1)}(I;\mathbb{F})$, and it, therefore, follows that g is finite for almost every $x\in I$. This implies that for almost every $x\in I$ the series $\sum_{j=1}^\infty f_j(x)$ is absolutely convergent and so convergent. Now define

$$f(x)=\begin{cases}\sum_{j=1}^\infty f_j(x), & g(x)<\infty\\ 0, & \text{otherwise.}\end{cases}$$

Since f is almost everywhere equal to the measurable function g, it is itself measurable, and further $\|f\|_p\le\|g\|_1<\infty$ so that $f\in\mathsf{L}^{(p)}(I;\mathbb{F})$. Furthermore, the Dominated Convergence Theorem gives

$$\begin{aligned}
&f(x)-\sum_{j=1}^n f_j(x)=\sum_{j=n+1}^\infty f_j(x),\quad \text{a.e. } x\in I\\
\implies\quad &\left|f(x)-\sum_{j=1}^n f_j(x)\right|\le\sum_{j=n+1}^\infty|f_j(x)|,\quad \text{a.e. } x\in I\\
\implies\quad &\lim_{n\to\infty}\left\|f-\sum_{j=1}^n f_j\right\|_p\le\lim_{n\to\infty}\sum_{j=n+1}^\infty\|f_j\|_p=0\\
\implies\quad &\lim_{n\to\infty}\left\|\left[f-\sum_{j=1}^n f_j\right]\right\|_p\le\lim_{n\to\infty}\sum_{j=n+1}^\infty\|[f_j]\|_p=0,
\end{aligned}$$

so giving convergence of $\sum_{j=1}^\infty[f_j]$.

Now let us prove that $\mathsf{L}^p(I;\mathbb{F})$ is isomorphic, as a normed vector space, to the completion of $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$. We first note that $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$ is a subspace of $\mathsf{L}^{(p)}(I;\mathbb{F})$. Moreover, by Exercise 2.9.8 it follows that if $\|f\|_p=0$ for $f\in\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$ then $f(x)=0$ for every $x\in I$. That is to say, the map

$$\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})\ni f\mapsto[f]\in\mathsf{L}^p(I;\mathbb{F})$$

is injective and so $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$ is a subspace of $\mathsf{L}^p(I;\mathbb{F})$. Thus to prove the theorem we need only show that $\mathsf{L}^p(I;\mathbb{F})$ is the closure of $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$. Thus we will show that if $f \in \mathsf{L}^{(p)}(I;\mathbb{F})$ then, for every $\epsilon \in \mathbb{R}_{>0}$ there exists $g \in \mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$ such that $\|f - g\|_p < \epsilon$. By Exercise 2.7.4, we can without loss of generality restrict to the case where f takes values in $\mathbb{R}_{\geq 0}$. We shall make this restriction in the arguments below.

Let us first consider the case when $I = [a, b]$ is compact and f is bounded. Let $M \in \mathbb{R}_{>0}$ be such that $f(x) \leq M$ for all $x \in I$. Let $\epsilon \in \mathbb{R}_{>0}$. By Theorem 2.9.3 there exists a continuous function $g\colon I \to \mathbb{R}_{\geq 0}$ such that

$$\lambda\left(\left\{x \in I \;\middle|\; |f(x) - g(x)| < \tfrac{\epsilon}{(2(b-a))^{1/p}}\right\}\right) < \frac{\epsilon^p}{2M^p}.$$

Then

$$\int_a^b |f(x) - g(x)|\,\mathrm{d}x < \frac{\epsilon^p}{2(b-a)}(b-a) + \frac{\epsilon^p}{2M^p}M^p < \epsilon^p.$$

Thus $\|f - g\|_p < \epsilon$, giving the result in this case.

Next we consider the case when $I = [a, b]$ is compact and f is possibly unbounded. Let $\epsilon \in \mathbb{R}_{>0}$. For $M \in \mathbb{R}_{>0}$ define

$$f_M(x) = \begin{cases} f(x), & f(x) \leq M, \\ M, & f(x) > M. \end{cases}$$

Since $f \in \mathsf{L}^{(p)}(I;\mathbb{F})$ there exists M sufficiently large that

$$\int_a^b |f(x) - f_M(x)|^p\,\mathrm{d}x < \frac{\epsilon^p}{2^p}.$$

By the argument in the previous paragraph there exists a continuous function $g\colon I \to \mathbb{R}_{\geq 0}$ such that $\|f_M - g\|_p < \frac{\epsilon}{2}$. Then, using the triangle inequality,

$$\|f - g\|_p \leq \|f - f_M\|_p + \|f_M - g\|_p < \epsilon,$$

giving the result in this case.

Finally, we consider the case when I is not compact. Let $\epsilon \in \mathbb{R}_{>0}$. We let $(I_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of compact intervals such that $I_j \subseteq I_{j+1}$ for each $j \in \mathbb{Z}_{>0}$ and such that $\cup_{j\in\mathbb{Z}_{>0}} I_j = I$. Define a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{L}^p(I;\mathbb{R})$ by

$$f_j(x) = \begin{cases} f(x), & x \in I_j, \\ 0, & \text{otherwise}. \end{cases}$$

By the Monotone Convergence Theorem we have

$$\lim_{j\to\infty} \int_I |f(x) - f_j(x)|^p\,\mathrm{d}x = \int_I \lim_{j\to\infty} |f(x) - f_j(x)|^p\,\mathrm{d}x = 0.$$

Thus $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges to f in $\mathsf{L}^{(p)}(I;\mathbb{F})$. Now, for each $j \in \mathbb{Z}_{>0}$, our arguments above ensure the existence of a continuous function $h_j\colon I_j \to \mathbb{R}_{\geq 0}$ such that $\|f_j|I_j - h_j\|_p^p < \frac{\epsilon^p}{2^{p+1}}$.

Note that if we extend h_j to I by asking that it be zero on $I \setminus I_j$ then this extension may not be continuous. However, we can linearly taper h_j to zero on $I \setminus I_j$ to arrive at a continuous function $g_j\colon I \to \mathbb{R}_{\geq 0}$ with compact support satisfying

$$\int_{I\setminus I_j} |g_j(x)|^p \,\mathrm{d}x < \frac{\epsilon^p}{2^{p+1}}.$$

Then

$$\int_I |f_j(x) - g_j(x)|^p \,\mathrm{d}x = \int_{I_j} |f_j(x) - h_j(x)|^p \,\mathrm{d}x + \int_{I\setminus I_j} |g_j(x)|^p \,\mathrm{d}x < \frac{\epsilon^p}{2^{p+1}} + \frac{\epsilon^p}{2^{p+1}} < \frac{\epsilon^p}{2^p}.$$

Now choose $N \in \mathbb{Z}_{>0}$ sufficiently large that $\|f - f_j\|_p < \frac{\epsilon}{2}$. Then, by the triangle inequality,

$$\|f - g_j\|_p \leq \|f - f_j\|_p + \|f_j - g_j\|_p < \epsilon,$$

as desired. ■

3.8.60 Notation (Representing functions in $\mathsf{L}^p(I;\mathbb{F})$) Just as we indicated for $\mathsf{L}^\infty(I;\mathbb{F})$ in Notation 3.8.48, we shall make use of the widespread and convenient convention of identifying an equivalence class in $\mathsf{L}^{(p)}(I;\mathbb{F})$, $p \in [1,\infty)$, with one of its representatives. This is mostly innocuous; however, there are times when this distinction must be made in order for things to make sense. While we do adopt the convention of writing elements of $\mathsf{L}^p(I;\mathbb{F})$ as f rather than $f + Z^p(I;\mathbb{F})$, we shall try to be careful to point out places where it really is the equivalence class that is being used. •

The second part of the Theorem 3.8.59 bears attention. As we commented after the proof of Theorem 3.3.6, although it is not difficult to demonstrate the existence of a completion of a normed vector space, it is not necessarily easy to understand what the meaning of points in the completion are relative to the original normed vector space. The second part of Theorem 3.8.59 says that although elements in the completion of $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$ are not functions, they are at least related to functions in that they are equivalence classes of functions. It might also be helpful to view the relationship between $\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$ and $\mathsf{L}^p(I;\mathbb{F})$ as being analogous to the relationship between $\mathbb{F}^\infty_0$ and $\ell^p(\mathbb{F})$, as born out in Table 3.2. What is interesting is that, to make

Table 3.2 The relationships between the objects in the left column are analogous to the relationships between the objects in the right column

Sequence space	Function space
$\mathbb{F}^\infty_0$	$\mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$
$\ell^p(\mathbb{F})$	$\mathsf{L}^p(I;\mathbb{F})$

this seemingly innocent analogy, one must go through the trials of defining the Lebesgue integral.

Let us prove the separability of $\mathsf{L}^p(I;\mathbb{F})$.

3.8.61 Proposition ($\mathsf{L}^p(I;\mathbb{F})$ is separable) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, if $I \subseteq \mathbb{R}$ is an interval, and if $p \in [1, \infty)$, then $\mathsf{L}^p(I;\mathbb{F})$ is separable.*

Proof From Theorem 3.8.40 we know that $\mathsf{C}^0_0(I;\mathbb{F})$ is separable and so $\mathsf{C}_{\mathrm{cpt}}(I;\mathbb{F})$ is also separable, being a subspace of $\mathsf{C}^0_0(I;\mathbb{F})$. Thus a countable dense subset $D \subseteq \mathsf{C}^0_{\mathrm{cpt}}(I;\mathbb{F})$ is also dense in $\mathsf{L}^p(I;\mathbb{F})$ by Exercise 3.6.2. ■

It is useful to be able to relate convergence in $\mathsf{L}^p(I;\mathbb{F})$ to pointwise convergence. The precise statement of this is as follows. Here we are careful to express the result in terms of equivalence classes of functions, since this is important to the meaning of the result. In the statement of the result we denote $[f] = f + Z^p(I;\mathbb{F})$ for brevity.

3.8.62 Proposition (Pointwise convergence and convergence in $\mathsf{L}^p(I;\mathbb{F})$) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $p \in [1, \infty]$, and let $I \subseteq \mathbb{R}$ be an interval. If $([f_j])_{j\in\mathbb{Z}_{>0}}$ is a sequence in $\mathsf{L}^p(I;\mathbb{F})$ converging to $[f] \in \mathsf{L}^p(I;\mathbb{F})$, then there exists a subsequence $([f_{j_k}])_{k\in\mathbb{Z}_{>0}}$ with the property that, for any representatives $f_{j_k} \in [f_{j_k}]$, $k \in \mathbb{Z}_{>0}$, and any representative $f \in [f]$, we have $\lim_{k\to\infty} f_{j_k}(x) = f(x)$ for almost every $x \in I$.*

Proof Throughout the proof we work with arbitrary representatives f_{j_k}, $k \in \mathbb{Z}_{>0}$, as stated in the proof. Since $\lim_{j\to\infty}\|f - f_j\|_p = 0$ there exists a subsequence $(f_{j_k})_{k\in\mathbb{Z}_{>0}}$ satisfying $\|f_{j_{k+1}} - f_{j_k}\|_p \le 2^{-k}$. We then define

$$g_k(x) = \sum_{\ell=1}^{k} |f_{j_{k+1}}(x) - f_{j_k}(x)|$$

and $g(x) = \lim_{k\to\infty} g_k(x)$ whenever these quantities are finite, taking them to be zero otherwise. Using Minkowski's inequality, $\|g_k\|_p \le 1$. Fatou's Lemma then gives $\|g\|_p \le 1$. This means that $g(x)$ is finite for almost every $x \in I$. Now define

$$f(x) = f_{j_1}(x) + \sum_{j=1}^{\infty} (f_{j_{k+1}}(x) - f_{j_k}(x)) \tag{3.19}$$

when this limit exists, taking it to be zero otherwise. Since the sum converges absolutely for almost every $x \in I$ this implies that the limit in (3.19) exists for almost every $x \in I$. The matter of showing that $f \in \mathsf{L}^p(I;\mathbb{F})$ goes like the last steps in the proof of the completeness of in Theorems 3.8.47 and 3.8.59. This gives the result for a particular representative of the limit class in $\mathsf{L}^p(I;\mathbb{F})$. That the result holds for any representative follows since any two representatives differ on a set of zero measure. ■

3.8.8 Banach spaces of integrable functions on measure spaces

3.8.9 Banach spaces of measures

In this section we let $(X, \mathscr{A})$ be a measurable space, and we recall from Section 2.3.10 the $\mathbb{R}$-vector spaces $\mathsf{M}((X,\mathscr{A});\mathbb{R})$ and $\mathsf{M}((X,\mathscr{A});\mathbb{R}^n)$ of finite signed and $\mathbb{R}^n$-valued vector measures on $\mathscr{A}$, and the $\mathbb{C}$-vector space $\mathsf{M}((X,\mathscr{A});\mathbb{C})$ of complex measures on $\mathscr{A}$. For μ in either $\mathsf{M}((X,\mathscr{A});\mathbb{R})$ or $\mathsf{M}((X,\mathscr{A});\mathbb{C})$ the total variation of μ

is defined to be

$$\|\mu\| = \sup\left\{\sum_{j=1}^{k}|\mu(A_j)| \;\middle|\; (A_1,\dots,A_k) \text{ is a partition of } X\right\}$$

(for signed measures this follows from Proposition 2.3.48). If $\boldsymbol{\mu} \in \mathsf{M}((X,\mathscr{A});\mathbb{R}^n)$ then the total variation of $\boldsymbol{\mu}$ is defined by

$$|||\boldsymbol{\mu}|||_{\mathbb{R}^n} = \sup\left\{\sum_{j=1}^{k}\|\boldsymbol{\mu}(A_j)\|_{\mathbb{R}^n} \;\middle|\; (A_1,\dots,A_k) \text{ is a partition of } X\right\}.$$

We can now state the main result of this section.

3.8.63 Theorem (Banach spaces of measures) *The pairs* $(\mathsf{M}((\mathsf{X},\mathscr{A});\mathbb{R}),\|\cdot\|)$ *and* $(\mathsf{M}((\mathsf{X},\mathscr{A});\mathbb{R}^{\mathrm{n}}),|||\cdot|||_{\mathbb{R}^{\mathrm{n}}})$ *are* $\mathbb{R}$*-Banach spaces and the pair* $(\mathsf{M}((\mathsf{X},\mathscr{A});\mathbb{C}),\|\cdot\|)$ *is a* $\mathbb{C}$*-Banach space.*

Proof We first must verify that $\|\cdot\|$ and $|||\cdot|||_{\mathbb{R}^n}$ are norms. For $\|\cdot\|$, we clearly have $\|\mu\| \in \mathbb{R}_{>0}$ for $\mu \in \mathsf{M}((X,\mathscr{A});\mathbb{R})$. Also, if $\alpha \in \mathbb{F}$ for $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$,

$$\begin{aligned}\|\alpha\mu\| &= \sup\left\{\sum_{j=1}^{k}|\alpha\mu(A_j)| \;\middle|\; (A_1,\dots,A_k) \text{ is a partition of } X\right\}\\ &= |\alpha|\sup\left\{\sum_{j=1}^{k}|\mu(A_j)| \;\middle|\; (A_1,\dots,A_k) \text{ is a partition of } X\right\} = |\alpha|\|\mu\|.\end{aligned}$$

If $\mu_1,\mu_2 \in \mathsf{M}((X,\mathscr{A});\mathbb{F})$ then we have

$$\begin{aligned}\|\mu_1+\mu_2\| &= \sup\left\{\sum_{j=1}^{k}|\mu_1(A_j)+\mu_2(A_j)| \;\middle|\; (A_1,\dots,A_k) \text{ is a partition of } X\right\}\\ &\leq \sup\left\{\sum_{j=1}^{k}|\mu_1(A_j)|+|\mu_2(A_j)| \;\middle|\; (A_1,\dots,A_k) \text{ is a partition of } X\right\}\\ &= \|\mu_1\|+\|\mu_2\|\end{aligned}$$

using Proposition I-2.2.27. This gives the triangle inequality for $\|\cdot\|$. Finally, we suppose that $\|\mu\| = 0$. For $A \in \mathscr{A}$ we have

$$|\mu(A)| \leq |\mu(A)| + |\mu(X\setminus A)| \leq \|\mu\|$$

since $(A, X\setminus A)$ is a partition of X. Thus it follows that $\mu(A) = 0$ for every $A \in \mathscr{A}$. Thus μ is the zero measure. This verifies positive-definiteness of $\|\cdot\|$ and so verifies that it is a norm. An entirely similar analysis yields the same conclusion for $|||\cdot|||_{\mathbb{R}^n}$.

It now remains to verify the completeness of the normed vector spaces. We consider the case of a signed or complex measure, letting $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$. We consider a

Cauchy sequence $(\mu_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{M}((X,\mathscr{A});\mathbb{F})$. Let $\epsilon\in\mathbb{R}_{>0}$ and let $N\in\mathbb{Z}_{>0}$ be such that $\|\mu_j-\mu_k\|\le\epsilon$ for $j,k\ge N$. Then, for $A\in\mathscr{A}$, we have, since $(A,X\setminus A)$ is a partition of X,

$$|\mu_j(A)-\mu_k(A)|\le|(\mu_j-\mu_k)(A)|+|(\mu_j-\mu_k)(X\setminus A)|\le\|\mu_j-\mu_k\|\le\epsilon$$

for $j,k\ge N$. Thus, for every $A\in\mathscr{A}$, $(\mu_j(A))_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathbb{F}$. We then denote the limit of this Cauchy sequence by $\mu(A)$. We must show that the map $A\mapsto\mu(A)$ is a signed or complex measure.

The following lemma will be useful, saying that the limit $\lim_{j\to\infty}\mu_j(A)=\mu(A)$ in uniform in A.

1 **Lemma** *For $\epsilon\in\mathbb{R}_{>0}$ there exists $\mathrm{N}\in\mathbb{Z}_{>0}$ such that $|\mu(\mathrm{A})-\mu_{\mathrm{j}}(\mathrm{A})|<\epsilon$ for each $\mathrm{j}\ge\mathrm{N}$ and $\mathrm{A}\in\mathscr{A}$.*

Proof Let $\epsilon\in\mathbb{R}_{>0}$ and choose $N\in\mathbb{Z}_{>0}$ such that $\|\mu_j-\mu_k\|<\frac{\epsilon}{2}$ for $j,k\ge N$. Thus, as we saw above, $|\mu_j(A)-\mu_k(A)|<\frac{\epsilon}{2}$ for $j,k\ge N$. Now let N_1 be sufficiently large that $|\mu(A)-\mu_k(A)|<\frac{\epsilon}{2}$ for $k\ge N_1$. Now, if $A\in\mathscr{A}$ and $j\ge N$ we have

$$|\mu(A)-\mu_j(A)|\le|\mu(A)-\mu_k(A)|+|\mu_k(A)-\mu_j(A)|<\epsilon,$$

where $k\ge\max\{N,N_1\}$. ▼

Since $\mu_j(\varnothing)=0$ for every $j\in\mathbb{Z}_{>0}$ we obviously have

$$\mu(\varnothing)=\lim_{j\to\infty}\mu_j(\varnothing)=0.$$

Let $A_1,\dots,A_m$ be a finite family of pairwise disjoint $\mathscr{A}$-measurable sets. Since μ_j, $j\in\mathbb{Z}_{>0}$, is countably-additive, it is finitely-additive, and so

$$\mu_j(\cup_{l=1}^m A_l)=\sum_{l=1}^m\mu_j(A_l),\qquad j\in\mathbb{Z}_{>0}.$$

Therefore,

$$\mu(\cup_{l=1}^m A_l)=\lim_{j\to\infty}\mu_j(\cup_{l=1}^m A_l)=\lim_{j\to\infty}\sum_{l=1}^m\mu_j(A_l)=\sum_{l=1}^m\mu(A_l),$$

swapping the finite sum with the limit. This gives finite-additivity of μ. It also holds that μ is consistent since, by construction, it takes values in $\mathbb{R}$.

Now let $(A_l)_{l\in\mathbb{Z}_{>0}}$ be a family of $\mathscr{A}$-measurable sets such that $A_{l+1}\subseteq A_l$, $l\in\mathbb{Z}_{>0}$, and such that $\cap_{l\in\mathbb{Z}_{>0}}A_l=\varnothing$. Since μ_j, $j\in\mathbb{Z}_{>0}$, is countably-additive and consistent, by Proposition 2.3.3 we have

$$\lim_{l\to\infty}\mu_j(A_l)=0,\qquad j\in\mathbb{Z}_{>0}.$$

Let $\epsilon\in\mathbb{R}_{>0}$ and, by Lemma 1, let $N_1\in\mathbb{Z}_{>0}$ be such that $|\mu(A)-\mu_j(A)|<\frac{\epsilon}{2}$ for $j\ge N_1$ and $A\in\mathscr{A}$. Let $N\in\mathbb{Z}_{>0}$ be such that $|\mu_{N_1}(A_l)|<\frac{\epsilon}{2}$ for $l\ge N$. Then, for $l\ge N$ we have

$$|\mu(A_l)|\le|\mu(A_l)-\mu_{N_1}(A_l)|+|\mu_{N_1}(A_l)|<\epsilon.$$

Thus $\lim_{l\to\infty}\mu(A_l)=0$ and so μ is countable additive by Proposition 2.3.3.

Finally, we must show that $(\mu_j)_{j\in\mathbb{Z}_{>0}}$ converges to μ. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that $\|\mu_j - \mu_k\| < \epsilon$ for $j, k \geq N$. Let $(A_1, \dots, A_m)$ be a partition of X and note that, by definition of $\|\cdot\|$,

$$\sum_{l=1}^{m} |\mu_j(A_l) - \mu_k(A_l)| = \sum_{l=1}^{m} |(\mu_j - \mu_k)(A_l)| \leq \|\mu_j - \mu_k\| < \epsilon$$

for $j, k \geq N$. Therefore,

$$\sum_{l=1}^{m} |\mu(A_l) - \mu_k(A_l)| = \lim_{j\to\infty} \sum_{l=1}^{m} |\mu_j(A_l) - \mu_k(A_l)| \leq \epsilon$$

for $k \geq N$. Since this holds for every partition $(A_1, \dots, A_m)$ of X, taking the supremum over all such partitions gives $\|\mu - \mu_k\| \leq \epsilon$ for $k \geq N$, so giving convergence of $(\mu_j)_{j\in\mathbb{Z}_{>0}}$ to μ. ■

3.8.10 Notes

Exercises

3.8.1 For $a_1, \dots, a_n, b_1, \dots, b_n \in \mathbb{R}_{>0}$ show that

$$\sum_{j=1}^{n} a_j b_j \leq \max\{b_1, \dots, b_n\} \sum_{j=1}^{n} a_j.$$

3.8.2 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. For $(a_j)_{j\in\mathbb{Z}_{>0}} \in \ell^1(\mathbb{F})$ and $(b_j)_{j\in\mathbb{Z}_{>0}} \in \ell^\infty(\mathbb{F})$, show that $(a_j b_j)_{j\in\mathbb{Z}_{>0}} \in \ell^1(\mathbb{F})$ and that

$$\|(a_j b_j)_{j\in\mathbb{Z}_{>0}}\|_1 \leq \|(a_j)_{j\in\mathbb{Z}_{>0}}\|_1 \|(b_j)_{j\in\mathbb{Z}_{>0}}\|_\infty.$$

3.8.3 Show that $\mathbb{F}_0^\infty$ is not dense in $\ell^\infty(\mathbb{F})$.

3.8.4 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

(a) Show that $\ell^p(\mathbb{F}) \subseteq \mathsf{c}_0(\mathbb{F})$ for $p \in [1, \infty)$.

(b) Is $\ell^\infty(\mathbb{F}) \subseteq \mathsf{c}_0(\mathbb{F})$?

3.8.5 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $((\mathsf{V}_i, \|\cdot\|_i))_{i\in I}$ be a family of normed F-vector spaces. Show that if $\ell^p(\bigoplus_{i\in I} \mathsf{V}_i)$ is a Banach space for any $p \in [1, \infty)$ then V_i is a Banach space for every $i \in I$.

3.8.6 Show that $\mathsf{C}_0^0(\mathbb{R}; \mathbb{F})$ can be defined alternatively by (3.17).

3.8.7 Show that $\mathsf{C}_{\mathrm{cpt}}^0((0,1); \mathbb{F})$ is not dense in $\mathsf{L}^\infty((0,1); \mathbb{R})$.
Hint: *Consider* $\mathrm{f}(\mathrm{x}) = 1$, $\mathrm{x} \in (0,1)$.

3.8.8 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $I \subseteq \mathbb{R}$ be an interval, and let $f \in \mathsf{L}^{(1)}(I; \mathbb{F})$ and $g \in \mathsf{L}^{(\infty)}(I; \mathbb{F})$. Show that $fg \in \mathsf{L}^{(1)}(I; \mathbb{F})$ and $\|fg\|_1 \leq \|f\|_1 \|g\|_\infty$.

Section 3.9

Topological duals of Banach spaces

Examples: integral, delta-function

3.9.1 Theorem (Characterisation of topological dual) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}; \|\cdot\|)$ be a normed $\mathbb{F}$-vector space. For $\alpha \in \mathsf{V}'$ the following conditions are equivalent:*

(i) α is continuous;

(ii) α is continuous at 0_{V};

(iii) α is uniformly continuous;

(iv) α is bounded;

(v) $\ker(\alpha)$ is a closed subspace of V.

Proof The equivalence of the first four conditions is simply a specialisation of Theorem 3.5.8. That the first four conditions imply the fifth is also a specialisation of Theorem 3.5.8.

(v) $\implies$ (i): Suppose that α is discontinuous; we claim that there exists a sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{B}_{\mathsf{V}}(1, 0_{\mathsf{U}})$ such that the sequence $(\|\alpha(v_j)\|)_{j\in\mathbb{Z}_{>0}}$ diverges. If not, then there exists $R \in \mathbb{R}_{>0}$ such that $\alpha(\mathsf{B}_{\mathsf{V}}(1, 0_{\mathsf{V}}) \subseteq \mathsf{B}_{\mathbb{F}}(R, 0)$. Linearity of α then implies that, for $\alpha(\mathsf{B}_{\mathsf{V}}(\epsilon, 0_{\mathsf{V}})) \subseteq \mathsf{B}_{\mathbb{F}}(R\epsilon, 0)$, which in turn implies continuity of α at 0_{V}. But this implies continuity of α. Thus a sequence as claimed exists. Let us further assume that the sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ is nonzero; this can be done without loss of generality. Now let $v \notin \ker(\alpha)$ and define $\tilde{v}_j = v - \frac{\alpha(v)}{\alpha(v_j)} v_j$. We then have $\alpha(\tilde{v}_j) = 0$. Moreover,

$$\lim_{j\to\infty} \|\tilde{v}_j - v\|_{\mathsf{V}} = \lim_{j\to\infty} \left\| \frac{\alpha(v)}{\alpha(v_j)} v_j \right\|_{\mathsf{V}} = 0.$$

Thus the sequence $(\tilde{v}_j)_{j\in\mathbb{Z}_{>0}}$ in $\ker(\alpha)$ converges to $v \notin \ker(\alpha)$, implying that $\ker(\alpha)$ is not closed by Proposition 3.6.8 below. ■

Our next technical result is one that is in actuality extremely useful on many occasions.

3.9.2 Theorem (Hahn–Banach Theorem) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, if $(\mathsf{V}, \|\cdot\|)$ is a normed $\mathbb{F}$-vector space, if $\mathsf{U} \subseteq \mathsf{V}$ is a subspace, and if $\alpha \in \mathsf{U}^*$, then there exists $\bar{\alpha} \in \mathsf{V}^*$ such that $\bar{\alpha}|\mathsf{U} = \alpha$ and $\|\bar{\alpha}\|_{\mathsf{V},\mathbb{F}} = \|\alpha\|_{\mathsf{U},\mathbb{F}}$.*

Proof First we assume that $\mathbb{F} = \mathbb{R}$. If $\alpha = 0$ the result is trivial, so suppose that $\|\alpha\|_{\mathsf{U},\mathscr{F}} > 0$. Define a new norm $|||\cdot|||$ on V by $|||v||| = \|\alpha\|_{\mathsf{U};\mathbb{F}}\|v\|$ and note that for $u \in \mathsf{U}$ we have

$$|\alpha(u)| \leq \|\alpha\|_{\mathsf{U};\mathbb{R}}\|u\| = |||u|||.$$

Thus we suppose that α is a continuous linear map for which $|\alpha(u)| \leq |||u|||$ for all $u \in \mathsf{U}$. The result is also trivial if $\mathsf{U} = \mathsf{V}$, so we assume this is not the case. Then we take

$v_0 \notin \mathsf{U}$. We then have, for $u_1, u_2 \in \mathsf{U}$,

$$\begin{aligned} & \alpha(u_1) + \alpha(u_2) = \alpha(u_1 + u_2) \le |||u_1 + u_2||| \le |||u_1 + v_0||| + |||u_2 - v_0||| \\ \implies \quad & \alpha(u_2) - |||u_2 - v_0||| \le |||u_1 + v_0||| - \alpha(u_1) \\ \implies \quad & \sup_{u_2 \in \mathsf{U}}\{\alpha(u_2) - |||u_2 - v_0|||\} \le \inf_{u_1 \in \mathsf{U}}\{|||u_1 + v_0||| - \alpha(u_1)\}. \end{aligned} \tag{3.20}$$

Choose a_0 to lie between the "sup" and the "inf" in the preceding equation and for $u + cv_0 \in \mathsf{U} \oplus \operatorname{span}_{\mathbb{R}}(v_0)$ define $\bar{\alpha}(u + cv_0) = \alpha(u) + ca_0$. Clearly $\bar{\alpha}$ is linear and $\bar{\alpha}|\mathsf{U} = \alpha$. By the rightmost of the inequalities of (3.20) we also have

$$|\bar{\alpha}(u + cv_0)| = |\alpha(u) + ca_0| \le |\alpha(u) + c|||u + v_0||| - c\alpha(u)| \le |||u + cv_0|||,$$

provided that $c \ge 0$. If $c < 0$ then we similarly have

$$|\bar{\alpha}(u + cv_0)| \le |||u + cv_0|||,$$

using the leftmost of the inequalities (3.20). Let S denote the collection of ordered pairs (W, β) where

1. W is a subspace of V satisfying $\mathsf{U} \subseteq \mathsf{W} \subseteq \mathsf{V}$,
2. $\beta\colon \mathsf{W} \to \mathbb{R}$ is a continuous linear map on W,
3. $|\beta(w)| \le |||w|||$ for all $w \in \mathsf{W}$.

On S place a partial order $\preceq$ by $(\mathsf{W}_1, \beta_1) \preceq (\mathsf{W}_2, \beta_2)$ if $\mathsf{W}_1 \subseteq \mathsf{W}_2$ and $\beta_2|\mathsf{W}_1 = \beta_1$. Since any $(\mathsf{W}, \beta) \in S$ must satisfy $\mathsf{W} \subseteq \mathsf{V}$ every well ordered subset must have an upper bound. Therefore, by Zorn's Lemma, there exists a maximal element (W_0, β_0) of S. If $\mathsf{W}_0 \neq \mathsf{V}$ then W_0 can always be extended by one-dimension as above. Therefore we must have $\mathsf{W}_0 = \mathsf{V}$, and we have shown that when $\mathbb{F} = \mathbb{R}$ we may find $\bar{\alpha}\colon \mathsf{V} \to \mathbb{R}$ so that $\bar{\alpha}|\mathsf{U} = \alpha$ and $|\bar{\alpha}(v)| \le |||v|||$ for all $v \in \mathsf{V}$.

If $\mathbb{F} = \mathbb{C}$ let $\alpha = \alpha_R + i\alpha_I$ and note that by complex linearity of α we have $\alpha_I(u) = -\alpha_R(iu)$ for all $u \in \mathsf{U}$. By the real part of the theorem let $\bar{\alpha}_R\colon \mathsf{V} \to \mathbb{R}$ have the property that $\bar{\alpha}_R|\mathsf{U} = \alpha_R$ and $|\bar{\alpha}_R(v)| \le |||v|||$ for all $v \in \mathsf{V}$. Then define $\bar{\alpha}\colon \mathsf{V} \to \mathbb{C}$ by $\bar{\alpha}(v) = \bar{\alpha}_R(v) - i\bar{\alpha}_R(iv)$. One readily checks that $\bar{\alpha}$ is $\mathbb{C}$-linear and that $\bar{\alpha}|\mathsf{U} = \alpha$. We also compute, using the polar form $\bar{\alpha}(v) = |\bar{\alpha}(v)|e^{i\theta}$,

$$|\bar{\alpha}(v)| = e^{-i\theta}\bar{\alpha}(v) = \bar{\alpha}_R(ve^{-i\theta}) \le |||ve^{-i\theta}||| = |||v|||.$$

We have now shown that there exists $\bar{\alpha}\colon \mathsf{U} \to \mathbb{F}$ so that $\bar{\alpha}|\mathsf{U} = \alpha$ and so that $|\bar{\alpha}(v)| \le |||v|||$ for all $v \in \mathsf{V}$. But this immediately gives $\|\bar{\alpha}\|_{\mathsf{V},\mathbb{F}} \le \|\alpha\|_{\mathsf{V},\mathbb{F}}$ by the definition of $|||\cdot|||$. However, since $\bar{\alpha}$ extends α we also have $\|\bar{\alpha}\|_{\mathsf{V},\mathbb{F}} \ge \|\alpha\|_{\mathsf{V},\mathbb{F}}$. This gives $\|\bar{\alpha}\|_{\mathsf{V},\mathbb{F}} = \|\alpha\|_{\mathsf{V},\mathbb{F}}$, as desired. ■

3.9.1 Reflexivity in Banach spaces

Section 3.10

Examples of duals of Banach spaces

3.10.1 The dual of $\mathsf{L}^p(\mathsf{I};\mathbb{F})$

We next consider two technical results that we present for completeness, as they are used in the proofs of Theorems IV-3.2.45 and IV-3.3.23. These next results can be skipped if one does not feel the need to understand the proofs of Theorems IV-3.2.45 and IV-3.3.23. The first result is an extension of the Riesz Representation Theorem. Note that the Riesz Representation Theorem says that the dual of $\mathsf{L}^2(\mathbb{T};\mathbb{F})$ is naturally isomorphic to $\mathsf{L}^2(\mathbb{T};\mathbb{F})$ for any continuous time-domain $\mathbb{T}$. This raises the question of the character of the dual of $\mathsf{L}^p(\mathbb{T};\mathbb{F})$ for $p \neq 2$. The following result gives this characterisation for $p \in [1,\infty)$.

3.10.1 Theorem (The dual of $\mathsf{L}^p(\mathbb{T};\mathbb{F})$) *Let $\mathbb{T}$ be a continuous time-domain and let $p \in [1,\infty)$. If $\alpha\colon \mathsf{L}^p(\mathbb{T};\mathbb{F}) \to \mathbb{F}$ is continuous then there exists a unique $g_\alpha \in \mathsf{L}^q(\mathbb{T};\mathbb{F})$ so that*

$$\alpha(f) = \int_{\mathbb{T}} f(t)g_\alpha(t)\,dt, \qquad f \in \mathsf{L}^p(\mathbb{T};\mathbb{F}),$$

where $\frac{1}{p} + \frac{1}{q} = 1$. Furthermore, $\|\alpha\|_{\mathsf{L}^p,\mathbb{F}} = \|g_\alpha\|_q$.

Proof We first a technical lemma.

1 Lemma *If $g \in \mathsf{L}^1([a,b];\mathbb{F})$ and if there exists $M > 0$ so that*

$$\Big|\int_a^b f(t)g(t)\,dt\Big| \le M\|f\|_p$$

for each $f \in \mathsf{L}^\infty([a,b];\mathbb{F})$ then $g \in \mathsf{L}^q([a,b];\mathbb{F})$ and $\|g\|_q \le M$ for $\frac{1}{p} + \frac{1}{q} = 1$.

Proof Let us first consider the case $p = 1$. Denote

$$A_\epsilon = \{t \mid |g(t)| \ge M + \epsilon\}$$

and define $f = \operatorname{sign}(g)\chi_{A_\epsilon}$. Then

$$M\lambda(A_\epsilon) = M\|f\|_1 \ge \Big|\int_a^b f(t)g(t)\,dt\Big| \ge (M+\epsilon)\lambda(A_\epsilon).$$

From this we deduce that $\lambda(A_\epsilon) = 0$ and so $\|g\|_\infty \le M$.

For $p \in (1,\infty)$ we define

$$g_n = \begin{cases} 0, & |g(t)| > n \\ g(t), & |g(t)| \le n. \end{cases}$$

Then take $f_n(t) = |g_n(t)|\,\mathrm{sign}(g_n(t))$. Note that $\|f_n\|_p = \|g_n\|_q^{q/p}$ and that $|g_n(t)|^q = f_n(t)g_n(t) = f_n(t)g(t)$. Therefore

$$\|g_n\|_q^q = \int_a^b f_n(t)g(t)\,\mathrm{d}t \le M\|f_n\|_p = M\|g_n\|_q^{q/p}.$$

Thus

$$\|g_n\|_q^q \le M\|g_n\|_q^{q/p} \quad \Longrightarrow \quad \|g_n\|_q \le M \quad \Longrightarrow \int_a^b |g_n(t)|^q\,\mathrm{d}t \le M^q,$$

using the fact that $q - \frac{q}{p} = 1$. Since $\lim_{n\to\infty} g_n(t) = g(t)$ for a.e. t we have, by the Monotone Convergence Theorem,

$$\|g\|_q^q = \int_a^b \lim_{n\to\infty} |g_n(t)|^q\,\mathrm{d}t = \lim_{n\to\infty} \int_a^b |g_n(t)|^q\,\mathrm{d}t \le M,$$

as desired. ▼

We now prove the theorem when $\mathbb{T}$ is bounded, say $\mathbb{T} \in \{[a,b],(a,b],[a,b),(a,b)\}$. Define $\mu_\alpha\colon [a,b] \to \mathbb{F}$ by $\mu_\alpha(t) = \alpha(\chi_{[a,t]})$. We claim that μ_α is absolutely continuous. For $\epsilon > 0$ let $\delta = \left(\frac{\epsilon}{\|\alpha\|_{\mathsf{L}^p,\mathbb{F}}}\right)^p$. For a collection $((t_{1,j},t_{2,j}))_{j\in\{1,\dots,n\}}$ of disjoint intervals of total measure less than δ, and define

$$f = \sum_{j=1}^k (\chi_{[a,t_{2,j}]} - \chi_{[a,t_{1,j}]})\,\mathrm{sign}(\mu_\alpha(t_{2,j}) - \mu_\alpha(t_{1,j})).$$

Note that $\|f\|_p^p < \delta$. We then have

$$\sum_{j=1}^k |\mu_\alpha(t_{2,j}) - \mu_\alpha(t_{1,j})| = \alpha(f) \le \|\alpha\|_{\mathsf{L}^p,\mathbb{F}}\|f\|_p < \|\alpha\|_{\mathsf{L}^p,\mathbb{F}}\delta^{1/p} < \epsilon.$$

Now, since μ_α is absolutely continuous there exists $g_\alpha \in \mathsf{L}^1(\mathbb{T};\mathbb{F})$ so that

$$\mu_\alpha(t) = \int_a^t g_\alpha(\tau)\,\mathrm{d}\tau$$

by Theorem 2.9.33. Since

$$\alpha(\chi_{[0,s]}) = \int_a^b g_\alpha(t)\chi_{[0,s]}(t)\,\mathrm{d}t$$

and since every step function is a.e. a finite linear combination of step functions of the sort $\chi_{[0,s]}$ for suitable choices of s, it follows that for any step function f on $\mathbb{T}$ we have

$$\alpha(f) = \int_a^b g_\alpha(t)f(t)\,\mathrm{d}t.$$

Now let f be an arbitrary bounded measurable signal on $\mathbb{T}$ and let $(f_n)_{n\in\mathbb{Z}_{>0}}$ be a sequence of step functions converging a.e. to f, this being possible by virtue of Theorem 2.9.2. Since f is bounded the sequence $(f - f_n)_{n\in\mathbb{Z}_{>0}}$ in bounded uniformly in n, and so the Dominated Convergence Theorem gives

$$\lim_{n\to\infty} \|f - f_n\|_p^p = \int_a^b \lim_{n\to\infty} |f(t) - f_n(t)|^p \, \mathrm{d}t = 0.$$

Therefore, by continuity of α,

$$\lim_{n\to\infty} |\alpha(f) - \alpha(f_n)| \le \lim_{n\to\infty} \|\alpha\|_{\mathsf{L}^p,\mathbb{F}} \|f - f_n\|_p = 0.$$

By the Dominated Convergence Theorem we then have

$$\alpha(f) = \lim_{n\to\infty} \alpha(f_n) = \lim_{n\to\infty} \int_a^b f_n(t) g_\alpha(t) \, \mathrm{d}t = \int_a^b f(t) g_\alpha(t) \, \mathrm{d}t,$$

by virtue of the fact that $|f_n(t)g_\alpha(t)| \le M|g_\alpha(t)|$ where M uniformly bounds $|f_n(t)|$ both in n and t. Since $\|\alpha(f)\| \le \|\alpha\|_{\mathsf{L}^p,\mathbb{F}} \|f\|_p$ this implies that $\|g_\alpha\|_q < \|\alpha\|_{\mathsf{L}^p,\mathbb{F}}$ by virtue of Lemma 1. It remains to show that the result holds if f is not bounded, but $f \in \mathsf{L}^p(\mathbb{T};\mathbb{F})$. By Theorem 3.8.59, for any $\epsilon > 0$ we can find a step function h for which $\|f - h\|_p < \epsilon$, and since h is bounded we have

$$\alpha(h) = \int_a^b h(t) g_\alpha(t) \, \mathrm{d}t.$$

Now we have

$$\begin{aligned}
\left|\alpha(f) - \int_a^b f(t) g_\alpha(t) \, \mathrm{d}t\right| &= \left|\alpha(f) - \alpha(h) + \int_a^b (h(t) - f(t)) g_\alpha(t) \, \mathrm{d}t\right| \\
&\le |\alpha(f - h)| + \int_a^b |(f(t) - h(t)) g_\alpha(t)| \, \mathrm{d}t \\
&\le \|\alpha\|_{\mathsf{L}^p;\mathbb{F}} \|f - h\|_p + \|g_\alpha\|_q \|f - h\|_p < (\|\alpha\|_{\mathsf{L}^p;\mathbb{F}} + \|g_\alpha\|_q)\epsilon,
\end{aligned}$$

by Lemma 3.8.54. To see that $\|\alpha\|_{\mathsf{L}^p,\mathbb{F}} = \|g_\alpha\|_q$ we take, for $p > 1$, $f(t) = |g_\alpha(t)|^{q/p} \operatorname{sign}(g_\alpha(t))$. We then have $|f(t)|^p = |g_\alpha(t)|^q = f(t) g_\alpha(t)$. Therefore $f \in \mathsf{L}^p(\mathbb{T};\mathbb{F})$ and

$$\alpha(f) = \int_a^b f(t) g_\alpha(t) \, \mathrm{d}t = \|g_\alpha\|_q^q = \|f\|_p \|g\|_q.$$

Thus $\|\alpha\|_{\mathsf{L}^p,\mathbb{F}} \ge \|g_\alpha\|_q$, provided that $p > 1$. For $p = 1$ and $q = \infty$ we argue this as follows. Let $\epsilon > 0$ and define

$$A_\epsilon = \{t \mid g_\alpha(t) \ge \|g_\alpha\|_\infty - \epsilon\}$$

and let $f_\epsilon = \chi_{A_\epsilon}$. Then

$$\begin{aligned}
\alpha(f_\epsilon) = \int_a^b f_\epsilon(t) g_\alpha(t) \, \mathrm{d}t &= \int_{A_\epsilon} g_\alpha(t) \, \mathrm{d}t \\
&\ge (\|g_\alpha\|_\infty - \epsilon)\lambda(A_\epsilon) = (\|g_\alpha\|_\infty - \epsilon)\|f_\epsilon\|_1.
\end{aligned}$$

The above all proves the theorem for bounded time-domains. If $\mathbb{T}$ is an unbounded time-domain then we let $(\mathbb{T}_j)_{j\in\mathbb{Z}_{>0}}$ be a collection of disjoint bounded time-domains for which $\mathbb{T} = \cup_{j\in\mathbb{Z}_{>0}}\mathbb{T}_j$. On each of the time-domains $\mathbb{T}_j$ we define $\alpha_j\colon \mathsf{L}^p(\mathbb{T}_j;\mathbb{F}) \to \mathbb{F}$ by $\alpha_j(f) = \alpha(\bar{f}\chi_{\mathbb{T}_j})$ where $\bar{f}$ is any signal on $\mathbb{T}$ agreeing with f on $\mathbb{T}_j$. We may then define $g_{\alpha_j} \in \mathsf{L}^q(\mathbb{T}_j;\mathbb{F})$ as above. We then define f_α on $\mathbb{T}$ by asking that it agree with f_{α_j} on $\mathbb{T}_j$. We must show that $f_\alpha \in \mathsf{L}^q(\mathbb{T};\mathbb{F})$ and that $\alpha(f) = \int_{\mathbb{T}} f(t)f_\alpha(t)\,\mathrm{d}t$ for all $f \in \mathsf{L}^p(\mathbb{T};\mathbb{F})$. For the former we proceed as follows. Define $A_n = \cup_{j=1}^n\mathbb{T}_j$ so that $(A_n)_{n\in\mathbb{Z}_{>0}}$ is an increasing sequence of subsets whose union is $\mathbb{T}$. Then define $g_n = g_\alpha\chi_{A_n}$ and $\alpha_n\colon \mathsf{L}^p(\mathbb{T};\mathbb{F}) \to \mathbb{F}$ by $\alpha_n(f) = \alpha(f\chi_{A_n})$. Note that we clearly have $\|\alpha_n\|_{\mathsf{L}^p,\mathbb{F}} \le \|\alpha\|_{\mathsf{L}^p;\mathbb{F}}$. From this it follows that that the sequence $(\|g_n\|_q)_{n\in\mathbb{Z}_{>0}}$ is bounded. It then follows that $\|g_\alpha\|_q < \infty$. If $f \in \mathsf{L}^p(\mathbb{T};\mathbb{F})$ then we define $f_n = f\chi_{A_n}$ and note that

$$\int_{\mathbb{T}} f(t)g_\alpha(t)\,\mathrm{d}t = \int_{\mathbb{T}} \lim_{n\to\infty} f_n(t)g_\alpha(t)\,\mathrm{d}t = \lim_{n\to\infty}\int_{\mathbb{T}} f_n(t)g_\alpha(t)\,\mathrm{d}t = \int_{\mathbb{T}} f(t)g_\alpha(t)\,\mathrm{d}t,$$

by the Dominated Convergence Theorem. ■

Chapter 4

Hilbert spaces

The notion of a Hilbert space is one of the most important in mathematics and applications of mathematics. It will arise in a crucial way in Fourier analysis in Chapters IV-5, IV-6, and IV-7. Hilbert space theory also plays an important rôle in optimisation theory, system theory, and partial differential equations, to name just as a few applications. As we shall see, Hilbert spaces are examples of Banach spaces, so all of our discussions of Chapter 3 apply to Hilbert spaces. However, the norm in a Hilbert space arises in a particular way, from an inner product. The inner product structure gives rise to important concepts such as orthogonality and self-duality, and it is concepts such as these that account for the importance of Hilbert spaces as examples of Banach spaces.

In this chapter we give a systematic overview of the notion of a Hilbert space, developing the theory starting in the simple but insightful finite-dimensional case. We endeavour to indicate how all of the concepts in general Banach space theory as developed in Chapter 3 specialise to Hilbert spaces.

Do I need to read this chapter? This chapter is an important one and most of the material in it is essential to the applied material that follows in later volumes. Certain specialised topics can be omitted on an initial reading. In particular, the details of uncountable orthonormal sets in Section 4.4.1 can be initially sidestepped, instead referring explicitly to the enumerable case considered in Section 4.4.3. •

Contents

4.1 Definitions and properties of inner product spaces 509
4.1.1 Inner products and semi-inner products 509
4.1.2 Inner product spaces as normed vector spaces 512
4.1.3 Orthogonality 517
4.1.4 Hilbert spaces and their subspaces 520
4.1.5 Minimising distance to a set 525
4.1.6 Norms 528
Exercises 528
4.2 Continuous maps between inner product spaces 531
4.2.1 The dual of an inner product space 531

4.2.2 Particular aspects of continuity for inner product spaces 534
4.2.3 The adjoint of a continuous linear map 535
4.2.4 Spectral properties for operators on Hilbert spaces 536
4.2.5 Notes 539
Exercises 539
4.3 Examples of Hilbert spaces 542
Exercises 542
4.4 Orthonormal bases in Hilbert spaces 543
4.4.1 General definitions and results 543
4.4.2 Finite orthonormal sets and finite Hilbert bases 554
4.4.3 Enumerable orthonormal sets and enumerable Hilbert bases 558
4.4.4 Generalised Fourier series 565
4.4.5 Classification of Hilbert spaces 567
4.4.6 Notes 572
Exercises 572

Section 4.1

Definitions and properties of inner product spaces

We have already encountered an important example of inner product, the standard inner product on $\mathbb{R}^n$ in Section II-1.2. The axioms defining a general inner product are exactly those for the standard inner product on $\mathbb{R}^n$, with the slight added generality that we allow for vector spaces over $\mathbb{C}$ as well as over $\mathbb{R}$.

Do I need to read this section? If you are reading this chapter then you should read this section. •

4.1.1 Inner products and semi-inner products

Just as we did in Chapter 3, we will simultaneously deal with the fields $\mathbb{R}$ and $\mathbb{C}$ by letting $\mathbb{F}$ denote wither $\mathbb{R}$ or $\mathbb{C}$, by letting $|a|$, $a \in \mathbb{F}$, denote the absolute value or modulus, and by letting $\bar{a}$, $a \in \mathbb{F}$, denote either a or the complex conjugate of a. We refer to Notation 3.1.1.

Just as in parts of Chapter 3 we considered seminorms, we will also consider semi-inner products in parts of this chapter. There is an additional caveat to make in this respect. The notion of a seminorm has an important independent life separate from its defining a norm as in Theorem 3.1.8. Indeed, in Chapter 6 we will devote significant time and effort to how seminorms arise in linear analysis. However, this is much less the case with the notion of a semi-inner product. Indeed, most authors do not mention the concept. We do so for two reasons: (1) there are examples of semi-inner products that arise *en route* to the construction of certain inner products; (2) we wish to maintain some consistency with the presentation in Chapter 3. Nonetheless, the reader is well-advised to not place much stock in the concept of a semi-inner product and to focus instead on the special case of an inner product.

With all that said, we can give the definitions.

4.1.1 Definition (Semi-inner product, inner product) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be an $\mathbb{F}$-vector space. A ***semi-inner product*** on V is a map $\mathsf{V} \times \mathsf{V} \ni (v_1, v_2) \mapsto \langle v_1, v_2 \rangle \in \mathbb{F}$ with the following properties:

(i) $\langle v_1, v_2 \rangle = \overline{\langle v_2, v_1 \rangle}$ for $v_1, v_2 \in \mathsf{V}$ (***symmetry***);

(ii) $\langle a_1 v_1 + a_2 v_2, v \rangle = a_1 \langle v_1, v \rangle + a_2 \langle v_2, v \rangle$ for $a_1, a_2 \in \mathbb{F}$ and $v_1, v_2 \in \mathsf{V}$ (***linearity***);

(iii) $\langle v, v \rangle \geq 0$ for $v \in \mathsf{V}$, (***positivity***).

An ***inner product*** on V is a semi-inner product $(v_1, v_2) \mapsto \langle v_1, v_2 \rangle$ with the additional property that

(iv) $\langle v, v \rangle = 0$ only if $v = 0_{\mathsf{V}}$ (***definiteness***).

We shall often denote a semi-inner product by $\langle \cdot, \cdot \rangle$. •

Note that the condition for positivity makes sense even when $\mathbb{F} = \mathbb{C}$ since $\langle v, v\rangle$ is always real. Indeed, using symmetry of the semi-inner product,

$$\overline{\langle v, v\rangle} = \overline{\overline{\langle v, v\rangle}} = \langle v, v\rangle,$$

and since the subset $\mathbb{R} \subseteq \mathbb{C}$ is exactly characterised by its being the subset fixed by complex conjugation, it follows that $\langle v, v\rangle \in \mathbb{R}$.

Let us record a trivial consequence of the properties of a semi-inner product.

4.1.2 Proposition (Bilinearity or sesquilinearity of a semi-inner product) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let V be an $\mathbb{F}$-vector space, and let $\langle\cdot,\cdot\rangle$ be a semi-inner product on V. Then, for $a_1, a_2, b_1, b_2 \in \mathbb{F}$ and $u_1, u_2, v_1, v_2 \in \mathsf{V}$ we have*

$$\langle a_1u_1 + a_2u_2, b_1v_1 + b_2v_2\rangle = a_1\bar{b}_1\langle u_1, v_1\rangle + a_1\bar{b}_2\langle u_1, v_2\rangle + a_2\bar{b}_1\langle u_2, v_1\rangle + a_2\bar{b}_2\langle u_2, v_2\rangle.$$

Proof We leave this as Exercise 4.1.1. ■

In the case when $\mathbb{F} = \mathbb{R}$ this property is called ***bilinearity*** and when $\mathbb{F} = \mathbb{C}$ this property is called ***sesquilinearity***.

Let us give some examples of inner products and semi-inner products.

4.1.3 Examples (Semi-inner product, inner product)

1. Any $\mathbb{F}$-vector space V has the useless semi-inner product defined by $\langle v_1, v_2\rangle = 0$ for all $v_1, v_2 \in \mathsf{V}$. This is only an inner product in the uninteresting case when $\mathsf{V} = \{0_{\mathsf{V}}\}$.
2. On $\mathbb{F}^n$ define
$$\langle \boldsymbol{u}, \boldsymbol{v}\rangle_2 = \sum_{j=1}^{n} u_j\bar{v}_j.$$
This is readily seen to be an inner product on $\mathbb{F}^n$. In the case when $\mathbb{F} = \mathbb{R}$ this specialises to the standard inner product on $\mathbb{R}^n$ discussed in Section II-1.2. Note that we use different notation for this object than was used in Chapter II-1, but we will still refer to it as the standard inner product.
3. Recall from Example I-4.5.2–4 that $\mathbb{F}_0^\infty$ denotes the sequences $(a_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{F}$ for which the set $\{j \in \mathbb{Z}_{>0} \mid v_j \neq 0\}$ is finite. Thus sequences in $\mathbb{F}_0^\infty$ are eventually zero. We define
$$\langle (a_j)_{j\in\mathbb{Z}_{>0}}, (b_j)_{j\in\mathbb{Z}_{>0}}\rangle_2 = \sum_{j=1}^{\infty} a_j\bar{b}_j,$$
noting that the sum makes sense since it is finite. It is a straightforward exercise to show that $\langle\cdot,\cdot\rangle_2$ is an inner product.

4. Finally, we consider the $\mathbb{F}$-vector space $\mathsf{C}^0([a,b];\mathbb{F})$ of continuous $\mathbb{F}$-valued functions on the compact interval $[a,b]$. Here we define an inner product on $\mathsf{C}^0([a,b];\mathbb{F})$ by

$$\langle f,g\rangle = \int_a^b f(x)\bar{g}(x)\,\mathrm{d}x.$$

One readily verifies all properties of the inner product, possibly resorting to Exercise I-3.4.1 for the positive-definiteness. •

Just as all vector spaces were shown to possess a norm in Proposition 3.1.4, we can use a similar strategy to show that all vector spaces possess an inner product.

4.1.4 Proposition (Vector spaces always have at least one inner product) *If* $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ *and if* V *is an* $\mathbb{F}$*-vector space then there is an inner product on* V*.*

Proof By Theorem I-4.5.45 we know the vector space V possesses a basis which establishes an isomorphism ι of V with $\mathbb{F}_0^J$ for some set J. Let us first define an inner product on $\mathbb{F}_0^J$. Writing a typical element of $\mathbb{F}_0^J$ as $(v_j)_{j\in J}$ we define

$$\langle (u_j)_{j\in J},(v_j)_{j\in J}\rangle_J = \sum_{j\in J}\bar{u}_j v_j,$$

the sum being well-defined since it is finite. To show that $\langle\cdot,\cdot\rangle_J$ is an inner product is a mere matter of checking the definitions. Now define

$$\langle u,v\rangle_{\mathsf{V}} = \langle\iota(u),\iota(v)\rangle_J,\qquad u,v\in\mathsf{V}.$$

To verify that $\langle\cdot,\cdot\rangle_{\mathsf{V}}$ is an inner product is straightforward. Symmetry is obvious. For linearity we compute

$$\langle a_1v_1+a_2v_2,v\rangle_{\mathsf{V}} = \langle\iota(a_1v_1+a_2v_2),v\rangle_J = \iota a_1\iota(v_1)+a_2\iota(v_2)v_J = a_1\langle v_1,v\rangle + a_2\langle v_2,v\rangle_{\mathsf{V}},$$

using linearity of $\langle\cdot,\cdot\rangle_J$ and ι. Positivity follows immediately from positivity of $\langle\cdot,\cdot\rangle_J$. Definiteness is shown as follows. Suppose that $\langle v,v\rangle_{\mathsf{V}}=0$. Then $\langle\iota(v),\iota(v)\rangle_J=0$ and so $\iota(v)=0_{\mathbb{F}_0^J}$ by definiteness of $\langle\cdot,\cdot\rangle_J$. Thus $v=0_{\mathsf{V}}$ since ι is an isomorphism. ■

As with the corresponding Proposition 3.1.4, one must take care to understand that the preceding result asserts neither the existence of a unique or even natural inner product. Moreover, there is no assurance that the inner product defined in the preceding result is useful. We refer to Corollary 3.6.27 to see why some vector spaces are incapable of supporting interesting norms; the same idea applies to inner products since, as we shall shortly see, inner products give rise to norms.

Analogous to normed vector spaces we have the following terminology.

4.1.5 Definition (Semi-inner product space, inner product space) Let $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$.

(i) An $\mathbb{F}$***-semi-inner product space*** is a pair $(\mathsf{V},\langle\cdot,\cdot\rangle)$ where V is a $\mathbb{F}$-vector space and $\langle\cdot,\cdot\rangle$ is a semi-inner product on V.

(ii) An $\mathbb{F}$***-inner product space*** is a pair $(\mathsf{V},\langle\cdot,\cdot\rangle)$ where V is a $\mathbb{F}$-vector space and $\langle\cdot,\cdot\rangle$ is an inner product on V. •

4.1.6 Notation ((Semi-)inner product spaces) As was the case when we were working with seminormed and normed vector spaces, it will be convenient to be able to state results for both semi-inner product spaces and inner product spaces at the same time. In order to facilitate this we will write "(semi-)inner product space" when we wish to mean that either sorts of objects may be used in the statement. •

4.1.2 Inner product spaces as normed vector spaces

In this section we show that a (semi-)inner product space gives rise in a natural way to an associated (semi)normed vector space. In order to do so, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a $\mathbb{F}$-(semi-)inner product space and define a map $\mathsf{V} \ni v \mapsto \|v\| \in \mathbb{R}_{\geq 0}$ by $\|v\| = \sqrt{\langle v, v\rangle}$. While we use the notation $\|\cdot\|$ as if this function is a (semi)norm, we do not in fact know that this is a (semi)norm at this point. It is, however, easy to see that $\|\cdot\|$ satisfies all (semi)norm properties except the triangle inequality. In order to verify this we first prove the following result that is of independent interest.

4.1.7 Theorem (Cauchy–Bunyakovsky–Schwarz inequality) *For an $\mathbb{F}$-(semi-)inner product space $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ we have*

$$|\langle v_1, v_2\rangle| \leq \|v_1\|\,\|v_2\|, \qquad v_1, v_2 \in \mathsf{V}.$$

Moreover, if $\langle\cdot,\cdot\rangle$ is an inner product then equality holds in the above expression if and only if v_1 and v_2 are collinear, i.e., if and only if

$$\operatorname{span}_{\mathbb{F}}(v_1) \subseteq \operatorname{span}_{\mathbb{F}}(v_2) \quad \textit{or} \quad \operatorname{span}_{\mathbb{F}}(v_2) \subseteq \operatorname{span}_{\mathbb{F}}(v_1).$$

Proof The result is obviously true for $v_2 = 0$, so we shall suppose that $v_2 \neq 0$. We first prove the result for $\|v_2\| = 1$. In this case we have

$$\begin{aligned}
0 &\leq \|v_1 - \langle v_1, v_2\rangle v_2\|^2 \\
&= \langle v_1 - \langle v_1, v_2\rangle v_2, v_1 - \langle v_1, v_2\rangle v_2\rangle \\
&= \langle v_1, v_1\rangle - \langle v_1, v_2\rangle\langle v_2, v_1\rangle - \overline{\langle v_1, v_2\rangle}\langle v_1, v_2\rangle + \langle v_1, v_2\rangle\overline{\langle v_1, v_2\rangle}\langle v_2, v_2\rangle \\
&= \|v_1\|^2 - |\langle v_1, v_2\rangle|^2,
\end{aligned}$$

where we have used Proposition 4.1.2. Thus we have shown that, provided $\|v_2\| = 1$,

$$|\langle v_1, v_2\rangle|^2 \leq \|v_1\|^2.$$

Taking square roots yields the result in this case. For $\|v_2\| \neq 1$ we define $v_3 = \frac{v_2}{\|v_2\|}$ so that $\|v_3\| = 1$. In this case

$$|\langle v_1, v_3\rangle| \leq \|v_1\| \quad \Longrightarrow \quad \frac{|\langle v_1, v_2\rangle|}{\|v_2\|} \leq \|v_1\|,$$

and so the inequality in the theorem holds.

Note that $\operatorname{span}_{\mathbb{F}}(v_1) \subset \operatorname{span}_{\mathbb{F}}(v_2)$ if and only if $v_1 = 0_{\mathsf{V}}$. In this case it is obvious that equality holds in the stated inequality. Similarly, equality holds if $\operatorname{span}_{\mathbb{F}}(v_2) \subset$

$\text{span}_{\mathbb{F}}(v_1)$. If $\text{span}_{\mathbb{F}}(v_1) = \text{span}_{\mathbb{F}}(v_2)$ then $v_1 = av_2$ for some $a \in \mathbb{F}$. In this case it is a direct computation, using properties of the inner product, to show that the stated inequality is in fact achieved with equality.

Conversely, suppose that the inequality in the theorem is achieved with equality. If equality is achieved with zero on each side then either or both of $\|v_1\|$ and $\|v_2\| = 0$ hold, i.e., either or both of v_1 and v_2 are zero. In this case we have either

$$\text{span}_{\mathbb{F}}(v_1) \subseteq \text{span}_{\mathbb{F}}(v_2) \quad \text{or} \quad \text{span}_{\mathbb{F}}(v_2) \subseteq \text{span}_{\mathbb{F}}(v_1),$$

as desired. Thus the final assertion to prove is that one of the preceding inclusions holds when equality is obtained with both sides of the equality being strictly positive. In this case both of v_1 and v_2 are nonzero. Let us first suppose that $\|v_2\| = 1$. If equality holds in the theorem statement then, going backwards through the argument in the first part of the proof, we must have

$$\|v_1 - \langle v_1, v_2\rangle v_2\|^2 = 0 \quad \Longrightarrow \quad v_1 = \langle v_1, v_2\rangle v_2,$$

giving the result in this case. If $\|v_2\| \neq 0$ then define $v_3 = \frac{v_2}{\|v_3\|}$ so that $\|v_3\| = 1$. Moreover,

$$|\langle v_1, v_3\rangle| = \frac{|\langle v_1, v_2\rangle|}{\|v_2\|} = \|v_1\| = \|v_1\|\|v_3\|,$$

and so equality holds for v_1 and v_3 in the inequality in the theorem. By the preceding argument we then have

$$v_1 = \langle v_1, v_3\rangle v_3 \quad \Longrightarrow \quad v_1 = \frac{\langle v_1, v_2\rangle}{\|v_2\|^2} v_2,$$

giving the final assertion for $\|v_2\| \neq 0$. ■

Using the Cauchy–Bunyakovsky–Schwarz inequality it is possible to show that the quantity $\|\cdot\|$ associated with an inner product is indeed a norm.

4.1.8 Theorem ((Semi-)inner product spaces are (semi)normed vector spaces) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-(semi-)inner product space, and define $\mathsf{V} \ni \mathrm{v} \mapsto \|\mathrm{v}\| \in \mathbb{R}_{\geq 0}$ be defined by $\|\mathrm{v}\| = \sqrt{\langle \mathrm{v}, \mathrm{v}\rangle}$. Then $(\mathsf{V}, \|\cdot\|)$ is a (semi)normed vector space.*

Proof All (semi)norm properties except the triangle inequality are easily verified. To verify the triangle inequality, for $v_1, v_2 \in \mathsf{V}$, we compute

$$\begin{aligned}\|v_1 + v_2\|^2 &= \langle v_1 + v_2, v_1 + v_2\rangle = \|v_1\|^2 + \langle v_1, v_2\rangle + \langle v_2, v_1\rangle + \|v_2\|^2 \\ &= \|v_1\|^2 + \langle v_1, v_2\rangle + \overline{\langle v_1, v_2\rangle} + \|v_2\|^2 = \|v_1\|^2 + 2\,\text{Re}(\langle v_1, v_2\rangle) + \|v_2\|^2 \\ &\leq \|v_1\|^2 + 2|\text{Re}(\langle v_1, v_2\rangle)| + \|v_2\|^2 \leq \|v_1\|^2 + 2|\langle v_1, v_2\rangle| + \|v_2\|^2 \\ &\leq \|v_1\|^2 + 2\|v_1\|\|v_2\| + \|v_2\|^2 = (\|v_1\| + \|v_2\|)^2,\end{aligned}$$

using the Cauchy–Bunyakovsky–Schwartz inequality. Taking square roots gives the result. ■

Needless to say, when we talk about the (semi)norm on a (semi-)inner product space, it is the norm of the preceding theorem to which we will refer.

A natural question that arises is then, "Given a norm on a vector space, can one tell when it comes from an inner product?" This question admits an easily stated, but not so easily proved, answer.

4.1.9 Theorem (When does a norm come from an inner product?) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and if* $(\mathsf{V}, \|\cdot\|)$ *is a (semi)normed* $\mathbb{F}$*-vector space, then the following statements are equivalent:*

(i) there exists an (semi-)inner product $\langle\cdot,\cdot\rangle$ *on* V *such that* $\|\mathrm{v}\| = \sqrt{\langle \mathrm{v}, \mathrm{v}\rangle}$ *for all* $\mathrm{v} \in \mathsf{V}$*;*

(ii) $\|\mathrm{v}_1 + \mathrm{v}_2\|^2 + \|\mathrm{v}_1 - \mathrm{v}_2\|^2 = 2\Big(\|\mathrm{v}_1\|^2 + \|\mathrm{v}_2\|^2\Big)$ *for every* $\mathrm{v}_1, \mathrm{v}_2 \in \mathsf{V}$ *(**parallelogram law**).*

Proof We leave to the reader as Exercise 4.1.4 the fairly easy task of showing that a (semi)norm derived from a (semi-)inner product satisfies the parallelogram law. Here we show the converse.

The proof for $\mathbb{F} = \mathbb{R}$ and $\mathbb{F} = \mathbb{C}$ are carried out separately. Let us consider the case of $\mathbb{F} = \mathbb{R}$ first. We claim that if a (semi)norm satisfies the parallelogram law then

$$\langle u, v\rangle \triangleq \tfrac{1}{4}\Big(\|u+v\|^2 - \|u-v\|^2\Big)$$

is a (semi-)inner product on V. It is clear that $\langle u, v\rangle = \langle v, u\rangle$ and that $\langle v, v\rangle \geq 0$ for all $v \in \mathsf{V}$ and that (in the case when $\|\cdot\|$ is a norm) $\langle v, v\rangle = 0$ if and only if $v = 0_{\mathsf{V}}$.

Let $u, v_1, v_2 \in \mathsf{V}$. Then

$$\begin{aligned}\langle u, v_1\rangle + \langle u, v_2\rangle &= \tfrac{1}{4}\Big(\|u+v_1\|^2 - \|u-v_1\|^2 + \|u+v_2\|^2 - \|u-v_2\|^2\Big)\\ &= \tfrac{1}{4}\Big(\Big\|u + \tfrac{1}{2}(v_1+v_2) + \tfrac{1}{2}(v_1-v_2)\Big\|^2 - \Big\|u - \tfrac{1}{2}(v_1+v_2) - \tfrac{1}{2}(v_1-v_2)\Big\|^2 +\\ &\quad \Big\|u + \tfrac{1}{2}(v_1+v_2) - \tfrac{1}{2}(v_1-v_2)\Big\|^2 - \Big\|u - \tfrac{1}{2}(v_1+v_2) + \tfrac{1}{2}(v_1-v_2)\Big\|^2\Big).\end{aligned} \tag{4.1}$$

By the parallelogram law we have

$$\Big\|u + \tfrac{1}{2}(v_1+v_2) + \tfrac{1}{2}(v_1-v_2)\Big\|^2 + \Big\|u + \tfrac{1}{2}(v_1+v_2) - \tfrac{1}{2}(v_1-v_2)\Big\|^2 = 2\Big\|u + \tfrac{1}{2}(v_1+v_2)\Big\|^2 + 2\Big\|\tfrac{1}{2}(v_1-v_2)\Big\|^2 \tag{4.2}$$

and

$$\Big\|u - \tfrac{1}{2}(v_1+v_2) + \tfrac{1}{2}(v_1-v_2)\Big\|^2 + \Big\|u - \tfrac{1}{2}(v_1+v_2) - \tfrac{1}{2}(v_1-v_2)\Big\|^2 = 2\Big\|u - \tfrac{1}{2}(v_1+v_2)\Big\|^2 + 2\Big\|\tfrac{1}{2}(v_1-v_2)\Big\|^2. \tag{4.3}$$

If we substitute (4.2) and (4.3) into (4.1) we get

$$\langle u, v_1\rangle + \langle u, v_2\rangle = \tfrac{1}{4}\Big(\Big\|u + \tfrac{1}{2}(v_1+v_2)\Big\|^2 - \Big\|u - \tfrac{1}{2}(v_1+v_2)\Big\|^2\Big) = 2\Big\langle u, \tfrac{1}{2}(v_1+v_2)\Big\rangle. \tag{4.4}$$

With this we prove a lemma.

1 Lemma *If* $\mathrm{k} \in \mathbb{Z}_{\geq 0}$ *then* $\Big\langle \frac{1}{2^{\mathrm{k}}}\mathrm{u}, \mathrm{v}\Big\rangle = \frac{1}{2^{\mathrm{k}}}\langle \mathrm{u}, \mathrm{v}\rangle$ *for all* $\mathrm{u}, \mathrm{v} \in \mathsf{V}$.

Proof The result is vacuously true for $k = 0$. If we let $v_2 = 0$ in (4.4) we have $\Big\langle \frac{1}{2}u, v\Big\rangle = \frac{1}{2}\langle u, v\rangle$, giving the lemma for $k = 1$. Now we proceed by induction. Suppose that the lemma holds for $k = m \geq 2$. Then

$$\Big\langle \tfrac{1}{2^{m+1}}u, v\Big\rangle = \Big\langle \tfrac{1}{2^m}\tfrac{1}{2}u, v\Big\rangle = \tfrac{1}{2^m}\Big\langle \tfrac{1}{2}u, v\Big\rangle = \tfrac{1}{2^{m+1}}\langle u, v\rangle,$$

using the induction hypotheses. ▼

Note that we now have

$$\langle u, v_1 + v_2\rangle = \langle v_1 + v_2, u\rangle = 2\Big\langle \tfrac{1}{2}(v_1 + v_1), u\Big\rangle = 2\Big\langle u, \tfrac{1}{2}(v_1 + v_2)\Big\rangle = \langle u, v_1\rangle + \langle u, v_2\rangle$$

where we have used (4.4).

Now we give another lemma.

2 Lemma *We have* $\left\langle \frac{m}{2^k}u, v\right\rangle = \frac{m}{2^k}\langle u, v\rangle$ *for all* $u, v \in V$, $m \in \mathbb{Z}$, *and* $k \in \mathbb{Z}_{\geq 0}$.

Proof We shall prove the result for $m \in \mathbb{Z}_{>0}$. The result for $m = 0$ is trivial, and the proof for $m \in \mathbb{Z}_{<0}$ follows along the same lines as the proof for $m \in \mathbb{Z}_{>0}$.

The result is clearly true for $m = 1$. Now suppose it is true for $m = l \geq 2$. Then we have

$$\begin{aligned}\left\langle \tfrac{l+1}{2^k}u, v\right\rangle = \left\langle \tfrac{l+1}{2^k}u, v\right\rangle &= \left\langle \tfrac{l}{2^k}u + \tfrac{1}{2^k}u, v\right\rangle \\ &= \left\langle \tfrac{l}{2^k}u, v\right\rangle + \left\langle \tfrac{1}{2^k}u, v\right\rangle = \left\langle \tfrac{l}{2^k}u, v\right\rangle + \left\langle \tfrac{1}{2^k}u, v\right\rangle \\ &= \tfrac{l}{2^k}\langle u, v\rangle + \tfrac{1}{2^k}\langle u, v\rangle = \tfrac{l+1}{2^k}\langle u, v\rangle,\end{aligned}$$

using the induction hypotheses. ▼

Now need a pair of technical lemmata.

3 Lemma *Let* $a, b \in \mathbb{R}$ *be such that* $a < b$. *Then there exist* $m \in \mathbb{Z}$ *and* $k \in \mathbb{Z}_{\geq 0}$ *such that* $a < \frac{m}{2^k} < b$.

Proof This is Exercise I-2.1.5. ▼

4 Lemma *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(V, \|\cdot\|)$ *be a (semi)normed* $\mathbb{F}$*-vector space. Fix* $u, v \in V$ *and define* $\phi\colon \mathbb{F} \to \mathbb{R}$ *by* $\phi(a) = \|au + v\|$. *Then* ϕ *is continuous.*

Proof This follows from Proposition 3.5.4 along with the fact that the composition of continuous maps is continuous. ▼

We may now prove the final property needed to show that $\langle\cdot,\cdot\rangle$ is a (semi-)inner product. That is, we show that $\langle au, v\rangle = a\langle u, v\rangle$ for all $a \in \mathbb{F}$ and $u, v \in V$. We will show that $|\langle au, v\rangle - a\langle u, v\rangle| < \epsilon$ for any $\epsilon \in \mathbb{R}_{>0}$. Let $\delta_{m,k} = a - m/2^k$ for $m \in \mathbb{Z}$ and $k \in \mathbb{Z}_{\geq 0}$. Note that we can make $|\delta_{m,k}|$ as small as we like by appropriately choosing m and k. We thus have

$$\begin{aligned}|\langle au, v\rangle - a\langle u, v\rangle| &= |\langle (m2^n + \delta_{m,k})u, v\rangle - (m/2^n + \delta_{m,k})\langle u, v\rangle| \\ &= |\langle \delta_{m,k}u, v\rangle - \delta_{m,k}\langle u, v\rangle| \leq |\langle \delta_{m,k}u, v\rangle| + |\delta_{m,k}\langle u, v\rangle|.\end{aligned}$$

For $\epsilon > 0$ let $\delta_1 = \left|\frac{\epsilon}{2\langle u, v\rangle}\right|$ and let δ_2 be such that $|\langle \delta_2 u, v\rangle| \leq \epsilon/2$. This is possible since $a \mapsto \langle au, v\rangle$ is continuous by Proposition 4.2.1. Now choose $m \in \mathbb{Z}$ and $k \in \mathbb{Z}_{>0}$ so that $\delta_{m,k} < \min(\delta_1, \delta_2)$. Then

$$|\langle au, v\rangle - a\langle u, v\rangle| \leq |\langle \delta_{m,k}u, v\rangle| + |\delta_{m,k}\langle u, v\rangle| < \epsilon/2 + \epsilon/2 = \epsilon,$$

as desired. This shows that $\langle\cdot,\cdot\rangle$ is a (semi-)inner product. Now we show that $\|\cdot\|$ is derived from this (semi-)inner product. This is easy since

$$\langle v, v\rangle = \tfrac{1}{4}\|v + v\|^2 = \|v\|^2.$$

This completes the proof for the case when $\mathbb{F} = \mathbb{R}$.

When $\mathbb{F} = \mathbb{C}$ we claim that

$$\langle u, v\rangle \triangleq \tfrac{1}{4}\Big(\|u+v\|^2 - \|u-v\|^2\Big) + \tfrac{\mathrm{i}}{4}\Big(\|u+\mathrm{i}v\|^2 - \|u-\mathrm{i}v\|^2\Big)$$

is a (semi-)inner product on V. First note that

$$\begin{aligned}\overline{\langle v, u\rangle} &= \tfrac{1}{4}\Big(\|v+u\|^2 - \|v-u\|^2\Big) - \tfrac{\mathrm{i}}{4}\Big(\|v+\mathrm{i}u\|^2 - \|v-\mathrm{i}u\|^2\Big)\\ &= \tfrac{1}{4}\Big(\|u+v\|^2 - \|u-v\|^2\Big) - \tfrac{\mathrm{i}}{4}\Big(\|-\mathrm{i}\mathrm{i}v+\mathrm{i}u\|^2 - \|-\mathrm{i}\mathrm{i}v-\mathrm{i}u\|^2\Big)\\ &= \tfrac{1}{4}\Big(\|u+v\|^2 - \|u-v\|^2\Big) + \tfrac{\mathrm{i}}{4}\Big(\|\mathrm{i}\mathrm{i}v+\mathrm{i}u\|^2 - \|\mathrm{i}\mathrm{i}v-\mathrm{i}u\|^2\Big)\\ &= \tfrac{1}{4}\Big(\|u+v\|^2 - \|u-v\|^2\Big) + \tfrac{\mathrm{i}}{4}\Big(\|u+\mathrm{i}v\|^2 - \|u-\mathrm{i}v\|^2\Big)\\ &= \langle u, v\rangle.\end{aligned}$$

We also compute

$$\langle u, v_1 + v_2\rangle = \langle u, v_1\rangle + \langle u, v_2\rangle, \quad \langle au, v\rangle = a\langle u, v\rangle$$

for $u, v, v_1, v_2 \in \mathsf{V}$ and for $a \in \mathbb{R}$. We also compute

$$\begin{aligned}\langle \mathrm{i}u, v\rangle &= \tfrac{1}{4}\Big(\|\mathrm{i}u+v\|^2 - \|\mathrm{i}u-v\|^2\Big) + \tfrac{\mathrm{i}}{4}\Big(\|\mathrm{i}u+\mathrm{i}v\|^2 - \|\mathrm{i}u-\mathrm{i}v\|^2\Big)\\ &= \tfrac{\mathrm{i}}{4}\Big(\|u+v\|^2 - \|u-v\|^2\Big) + \tfrac{1}{4}\Big(\|\mathrm{i}u-\mathrm{i}\mathrm{i}v\| - \|\mathrm{i}u+\mathrm{i}\mathrm{i}v\|\Big)\\ &= \mathrm{i}\tfrac{1}{4}\Big(\|u+v\|^2 - \|u-v\|^2\Big) + \mathrm{i}\tfrac{\mathrm{i}}{4}\Big(\|u+\mathrm{i}v\|^2 - \|u-\mathrm{i}v\|^2\Big)\\ &= \mathrm{i}\langle u, v\rangle\end{aligned}$$

We can then readily check that $\langle au, v\rangle = a\langle u, v\rangle$ for every $u, v \in \mathsf{V}$ and $a \in \mathbb{C}$. This shows that $\langle\cdot,\cdot\rangle$ is a (semi-)inner product. We also have

$$\langle v, v\rangle = \tfrac{1}{4}\|2v\|^2 + \tfrac{\mathrm{i}}{4}|1+\mathrm{i}|^2\|v\|^2 - |1-\mathrm{i}|^2\|v\|^2 = \tfrac{1}{4}\|2v\|^2 = \|v\|^2.$$

Taking square roots shows that $\|\cdot\|$ is the (semi)norm derived from the inner product $\langle\cdot,\cdot\rangle$, and so gives the theorem when $\mathbb{F} = \mathbb{C}$. ■

As a consequence of the proof we have the following formulae which relate an inner product to the norm defined by it.

4.1.10 Corollary (Polarisation identity) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-(semi-)inner product space with $\|\cdot\|$ the norm defined by $\langle\cdot,\cdot\rangle$. The following statements hold:*

(i) if $\mathbb{F} = \mathbb{R}$ then

$$\langle \mathrm{u}, \mathrm{v}\rangle = \tfrac{1}{4}\Big(\|\mathrm{u}+\mathrm{v}\|^2 - \|\mathrm{u}-\mathrm{v}\|^2\Big)$$

for all $\mathrm{u}, \mathrm{v} \in \mathsf{V}$;

(ii) if $\mathbb{F} = \mathbb{C}$ then

$$\langle \mathrm{u}, \mathrm{v}\rangle \triangleq \tfrac{1}{4}\Big(\|\mathrm{u}+\mathrm{v}\|^2 - \|\mathrm{u}-\mathrm{v}\|^2\Big) + \tfrac{\mathrm{i}}{4}\Big(\|\mathrm{u}+\mathrm{i}\mathrm{v}\|^2 - \|\mathrm{u}-\mathrm{i}\mathrm{v}\|^2\Big)$$

for all $\mathrm{u}, \mathrm{v} \in \mathsf{V}$.

The fact is that it is unusual for a norm to be derived from an inner product. However, since norms coming from inner products are so important, we will devote a great deal of effort to this special case.

With (semi-)inner product spaces now being normed vector spaces, all the norm machinery can be piled into a (semi-)inner product space. Indeed, we shall in this chapter freely refer to any part of Chapter 3. Also, we shall frequently apply the name for a (semi)normed vector space concepts directly to a (semi-)inner product space.

4.1.3 Orthogonality

One of the essential features of an inner product spaces that distinguish them from more general normed vector spaces is that one has the notion of orthogonality. We have some intuition about what orthogonality means in low dimensions (see Section II-1.2), and some of this intuition carries over to general inner product spaces. However, as is often the case when one makes the leap to infinite-dimensions, one must be careful in relying solely on intuition in making assertions about what is true or not.

Let us give the definitions. Note that the word "orthogonal" has multiple meanings, depending on context.

4.1.11 Definition (Orthogonal, orthogonal complement) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-(semi-)inner product space.

(i) Vectors $v_1, v_2 \in \mathsf{V}$ are ***orthogonal*** if $\langle v_1, v_2\rangle = 0$. We shall write $v_1 \perp v_2$ to denote v_1 and v_2 being orthogonal.

(ii) Sets $A_1, A_2 \subseteq \mathsf{V}$ are ***orthogonal*** if $\langle v_1, v_2\rangle = 0$ for every $v_1 \in A_1$ and $v_2 \in A_2$. We shall write $A_1 \perp A_2$ to denote A_1 and A_2 being orthogonal.

(iii) If $A \subseteq \mathsf{V}$ then the ***orthogonal complement*** of A is the set

$$A^\perp = \{u \in \mathsf{V} \mid \langle u, v\rangle = 0 \text{ for all } v \in A\}.$$ •

Let us give some elementary examples of orthogonal sets.

4.1.12 Examples (Orthogonality)

1. The vectors $(1, 2\mathrm{i}, -1), (1, \frac{3}{2} + \frac{\mathrm{i}}{2}, 3\mathrm{i}) \in \mathbb{C}^3$ are orthogonal.
2. In $\mathbb{F}^3$ the sets

$$A_1 = \operatorname{span}_{\mathbb{F}}((1,1,1),(0,1,1)), \quad A_2 = \operatorname{span}_{\mathbb{F}}((0,1,-1))$$

are orthogonal. Moreover, A_1 is the orthogonal complement of A_2 and A_2 is the orthogonal complement of A_1. •

In some sense, this entire chapter is about orthogonality. Let us here give a few simple consequences of the definitions.

4.1.13 Proposition (Properties of orthogonal complement) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *be an* $\mathbb{F}$*(semi-)-inner product space, and let* $\mathrm{A}, \mathrm{B} \subseteq \mathsf{V}$*. Then the following statements hold:*

(i) if $\mathrm{A} \subseteq \mathrm{B}$ *then* $\mathrm{B}^\perp \subseteq \mathrm{A}^\perp$*;*

(ii) $\mathrm{A} \subseteq (\mathrm{A}^\perp)^\perp$*;*

(iii) $\mathrm{A}^\perp$ *is a closed subspace of* V*;*

(iv) $\mathrm{A}^\perp = (\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\mathrm{A})))^\perp$*.*

If $\langle\cdot,\cdot\rangle$ *is additionally an inner product then*

(v) $\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\mathrm{A})) \cap \mathrm{A}^\perp = \{0_{\mathsf{V}}\}$*.*

Proof The proof of parts (i), (ii), and (iii) are left to the reader as Exercise 4.1.11.

(iv) Since $A \subseteq \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A))$ it follows from part (i) that

$$A^\perp \supseteq (\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)))^\perp.$$

Now let $u \in A^\perp$ so that $\langle u, v\rangle = 0$ for every $v \in A$. Next let $\hat{v} \in \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A))$ and by Proposition 3.6.8 let $(\hat{v}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathrm{span}_{\mathbb{F}}(A)$ converging to $\hat{v}$. For each $j \in \mathbb{Z}_{>0}$ we can write

$$\hat{v}_j = \sum_{r=1}^{k_j} c_{jr} v_{jr}$$

for some $k_j \in \mathbb{Z}_{>0}$ and $c_{jr} \in \mathbb{F}$ and $v_{jr} \in A$, $r \in \{1, \dots, k_j\}$. It therefore follows that

$$\left\langle u, \hat{v}_j \right\rangle = \left\langle u, \sum_{r=1}^{k_j} c_{jr} v_{jr} \right\rangle = \sum_{j=1}^{k_j} \bar{c}_{jr} \langle u, v_{jr}\rangle = 0$$

for each $j \in \mathbb{Z}_{>0}$. This allows us to deduce that

$$\left\langle u, \hat{v} \right\rangle = \lim_{j\to\infty} \left\langle u, \hat{v}_j \right\rangle = 0$$

by Proposition 4.2.1 and Theorem 3.5.2. Thus $u \in (\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)))^\perp$ as desired.

(v) If

$$v \in \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)) \cap A^\perp = \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)) \cap (\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)))^\perp$$

then $\langle v, v\rangle = 0$ which gives $v = 0_{\mathsf{V}}$ if $\langle\cdot,\cdot\rangle$ is an inner product. ■

The equality $A^\perp = (\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)))^\perp$ is an important one. It tells us that the orthogonal complement of a set is not a feature of the set, but of the closure of the subspace generated by this set. Thus there are two operations happening when taking orthogonal complements: "span" and "closure" (in that order). The appearance of the topological closure operation here is perhaps surprising at first encounter. Indeed, since all subspaces are closed in finite dimensions, closure does not make an appearance in that case.

It is fairly obviously true that $A \neq (A^\perp)^\perp$ in general, merely because A may not be a subspace but $(A^\perp)^\perp$ is a subspace. So the question of when $A = (A^\perp)^\perp$ is only interesting when A is a subspace. However, even in this case equality does not generally hold. This is something that we will explore in greater detail in , so here we merely content ourselves with a counterexample.

what?

4.1.14 Example ($\mathsf{U} \neq (\mathsf{U}^\perp)^\perp$) Let us take $\mathsf{V} = \ell^2(\mathbb{F})$ with the inner product

$$\langle (a_j)_{j\in\mathbb{Z}_{>0}}, (b_j)_{j\in\mathbb{Z}_{>0}}\rangle = \sum_{j=1}^{\infty} a_j \bar{b}_j.$$

This is a specialisation to $p = 2$ of the Banach space $\ell^p(\mathbb{F})$ considered in Section 3.8.2. We showed in Theorem 3.8.19 that this is a Banach space and in Corollary 3.8.21 that this Banach space is the completion of $\mathbb{F}_0^\infty$. Let us then take the subspace $\mathbb{F}_0^\infty$ of $\ell^2(\mathbb{F})$. We claim that $\mathbb{F}_0^\infty$ is a strict subspace of $((\ell_0^\infty)^\perp)^\perp$. To see this we first claim that $(\mathbb{F}_0^\infty)^\perp = \{0_{\ell^2(\mathbb{F})}\}$. Indeed, let $(\boldsymbol{e}_j)_{j\in\mathbb{Z}_{>0}}$ be the standard basis for $\mathbb{F}_0^\infty$. Thus, as a reminder,

$$\boldsymbol{e}_j(k) = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

Then, if $(a_j)_{j\in\mathbb{Z}_{>0}} \in (\mathbb{F}_0^\infty)^\perp$ then

$$\langle (a_j)_{j\in\mathbb{Z}_{>0}}, \boldsymbol{e}_k\rangle = a_k = 0$$

for every $k \in \mathbb{Z}_{>0}$. Thus $(\mathbb{F}_0^\infty)^\perp = \{0_{\ell^2(\mathbb{F})}\}$ as claimed. It, therefore, follows that $((\mathbb{F}_0^\infty)^\perp)^\perp = \ell^2(\mathbb{F})$ and so we have $\mathbb{F}_0^\infty$ as a strict subspace of $((\mathbb{F}_0^\infty)^\perp)^\perp$ as claimed. •

The issue with the preceding example, as we shall see in Theorem 4.1.19, is that $\mathbb{F}_0^\infty$ is not a *closed* subspace of $\ell^2(\mathbb{F})$.

Let us also record how orthogonality interacts with sums and intersections of subsets of V. For $A, B \subseteq \mathsf{V}$ we denote

$$A + B \triangleq \{u + v \mid u \in A,\ v \in B\}.$$

We now have the following assertions.

4.1.15 Proposition (Orthogonality and sum and intersection) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a $\mathbb{F}$-inner product space, and let $\mathsf{A}, \mathsf{B} \subseteq \mathsf{V}$. Then the following statements hold:*

(i) $(\mathsf{A} + \mathsf{B})^\perp = \mathsf{A}^\perp \cap \mathsf{B}^\perp$.

(ii) $(\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\mathsf{A})) \cap \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\mathsf{B})))^\perp = \mathsf{A}^\perp + \mathsf{B}^\perp$.

Proof (i) By part (iv) of Proposition 4.1.13 we have

$$(A + B)^\perp = (\mathrm{span}_{\mathbb{F}}(A + B))^\perp = (\mathrm{span}_{\mathbb{F}}(A) + \mathrm{span}_{\mathbb{F}}(B))^\perp,$$

using the easily verified identity $\mathrm{span}_{\mathbb{F}}(A+B) = \mathrm{span}_{\mathbb{F}}(A) + \mathrm{span}_{\mathbb{F}}(B)$. Let $w \in (A+B)^\perp$. Then $\langle w, u + v\rangle = 0$ for every $u \in \mathrm{span}_{\mathbb{F}}(A)$ and $v \in \mathrm{span}_{\mathbb{F}}(B)$. In particular, $\langle w, u\rangle = 0$ and $\langle w, v\rangle = 0$ for every $u \in \mathrm{span}_{\mathbb{F}}(A)$ and $v \in \mathrm{span}_{\mathbb{F}}(B)$. Thus $w \in A^\perp \cap B^\perp$. Next suppose that $w \in A^\perp \cap B^\perp$. Then, using part (iv) of Proposition 4.1.13,

$$w \in (\mathrm{span}_{\mathbb{F}}(A))^\perp \cap (\mathrm{span}_{\mathbb{F}}(B))^\perp.$$

Therefore, $\langle w, u\rangle = \langle w, v\rangle = 0$ for every $u \in \mathrm{span}_{\mathbb{F}}(A)$ and $v \in \mathrm{span}_{\mathbb{F}}(B)$. Thus $\langle w, u+v\rangle = 0$ $u \in \mathrm{span}_{\mathbb{F}}(A)$ and $v \in \mathrm{span}_{\mathbb{F}}(B)$, giving

$$w \in (\mathrm{span}_{\mathbb{F}}(A) + \mathrm{span}_{\mathbb{F}}(B))^{\perp} = (A + B)^{\perp},$$

as desired.

(ii)

Conversely, since

$$\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)) \cap \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)) \subseteq \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A))$$

we have

$$(\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)))^{\perp} \subseteq (\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)) \cap \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)))^{\perp}$$

by part (i) of Proposition 4.1.13. By part (iv) of the same result we then have

$$A^{\perp} \subseteq (\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)) \cap \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)))^{\perp}.$$

In like manner

$$B^{\perp} \subseteq (\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)) \cap \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)))^{\perp}.$$

Since $(\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)) \cap \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)))^{\perp}$ is a subspace by part (iii) of Proposition 4.1.13 it then follows that

$$A^{\perp} + B^{\perp} \subseteq (\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)) \cap \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(A)))^{\perp},$$

giving the desired conclusion. ■

4.1.4 Hilbert spaces and their subspaces

As inner-product spaces are normed vector spaces, the whole discussion of Cauchy sequences, convergent sequences, and completeness in Sections 3.2 and 3.3 can be applied to inner product spaces. The notion of a complete inner product space is important enough to have its own name.

4.1.16 Definition (Completeness, Hilbert space) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. A $\mathbb{F}$-inner product space $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ is ***complete*** if the corresponding normed vector space is complete. A ***$\mathbb{F}$-Hilbert***[1] ***space*** is a complete $\mathbb{F}$-inner product space. •

Since inner product spaces are also normed vector spaces, the construction of the completion in Theorem 3.3.6 also applies to inner product spaces. That is to say, every inner product space possesses a completion that is a Banach space. Of course, one would also like to have the completion be a Hilbert space, and this is the content of the next result.

[1] David Hilbert (1862–1943) in one of history's greatest mathematicians. At the 1900 International Congress of Mathematics in Paris, Hilbert gave a list of twenty three problems which he felt should guide mathematical research in the upcoming centuries. Many of Hilbert's problems have been solved, some to great aplomb. Hilbert's own contributions were in many fields, including geometry, analysis, logic, and algebra.

4.1.17 Theorem (Completion of an inner product space) *If* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *is an inner product space then there exists a Hilbert space* $(\overline{\mathsf{V}}, \overline{\langle\cdot,\cdot\rangle})$ *and an injective linear map* $\mathrm{i}_{\mathsf{V}}\colon \mathsf{V} \to \overline{\mathsf{V}}$ *with the following properties:*

(i) $\mathrm{image}(\mathrm{i}_{\mathsf{V}})$ *is dense in* $\overline{\mathsf{V}}$*;*

(ii) $\langle \mathrm{v}_1, \mathrm{v}_2\rangle = \overline{\langle \mathrm{i}_{\mathsf{V}}(\mathrm{v}_1), \mathrm{i}_{\mathsf{V}}(\mathrm{v}_2)\rangle}$.

Proof We let $\overline{\mathsf{V}}$ be the vector space constructed from the normed vector space associated to V as in Theorem 3.3.6, and we let i_{V} also be the linear map constructed in the proof of that result. Now let $\overline{v} \in \overline{\mathsf{V}}$ and $v \in \mathsf{V}$ and let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in V for which $\overline{v} = \lim_{j\to\infty} v_j$. We claim that the sequence $(\langle v_j, v\rangle)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{F}$ converges. We may suppose that $v \neq 0$ without loss of generality. Let $\epsilon > 0$ and choose $N \in \mathbb{Z}_{>0}$ so that $\|v_j - v_k\| < \frac{\epsilon}{\|v\|}$ for $j, k \geq N$; this is possible by continuity of the norm. By the Cauchy–Bunyakovsky–Schwarz inequality we then have

$$|\langle v_j, v\rangle - \langle v_k, v\rangle| \leq |\langle v_j - v_k, v\rangle| \leq \|v_j - v_k\|\|v\| \leq \epsilon$$

for $j, k \geq N$, showing that $(\langle v_j, v\rangle)_{j\in\mathbb{Z}_{>0}}$ is Cauchy, and so convergent. Thus we may sensibly define $\overline{\langle \overline{v}, v\rangle} = \lim_{j\to\infty}\langle v_j, v\rangle$. We may similarly, of course, define $\overline{\langle v, \overline{v}\rangle}$, thus defining $\overline{\langle\cdot,\cdot\rangle}$ on $\overline{\mathsf{V}} \times \mathsf{V}$ and $\mathsf{V} \times \overline{\mathsf{V}}$. The same sort of arguments also allow one to define $\overline{\langle \overline{v}_1, \overline{v}_2\rangle}$ for $\overline{v}_1, \overline{v}_2 \in \overline{\mathsf{V}}$. To show that the resulting map $\overline{\mathsf{V}} \times \overline{\mathsf{V}} \ni (\overline{v}_1, \overline{v}_2) \mapsto \overline{\langle \overline{v}_1, \overline{v}_2\rangle} \in \mathbb{F}$ is an inner product is a simple verification of the axioms, using the fact, for example, that if a sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ converges to v, then the sequence $(av_j)_{j\in\mathbb{Z}_{>0}}$ converges to av for $a \in \mathbb{F}$. That (i) holds is an immediate consequence of Theorem 3.3.6, and (ii) is obvious. ■

Let us consider our inner product space examples to determine which are Hilbert spaces.

4.1.18 Examples (Hilbert spaces and non-Hilbert spaces)

1. The inner product space $(\mathbb{F}^n, \langle\cdot,\cdot\rangle_2)$ is a Banach space by virtue of the fact that every finite-dimensional inner product space is complete (Theorem 3.3.3).
2. The inner product space $(\mathbb{F}_0^\infty, \langle\cdot,\cdot\rangle_2)$ is not complete. Indeed, in Corollary 3.8.21 we saw that its completion is $\ell^2(\mathbb{F})$ which contains $\mathbb{F}_0^\infty$ as a strict subset. To "by hand" show that $(\mathbb{F}_0^\infty, \langle\cdot,\cdot\rangle_2)$ is not complete can be done following the strategy of Example 3.3.1–1. We leave the working out of this to the reader as Exercise 4.1.6. The completion of $(\mathbb{F}_0^\infty, \langle\cdot,\cdot\rangle_2)$ is $(\ell^2(\mathbb{F}), \langle\cdot,\cdot\rangle_2)$ as is proved in Corollary 3.8.21.
3. The inner product space $(\mathsf{C}^0([a,b];\mathbb{F}), \langle\cdot,\cdot\rangle_2)$ is not a Hilbert space if $b > a$. In we showed that $\mathsf{L}^2([a,b];\mathbb{F})$ is the completion of $\mathsf{C}^0([a,b];\mathbb{F})$ with respect to the norm induced by the inner product $\langle\cdot,\cdot\rangle$. Since $\mathsf{C}^0([a,b];\mathbb{F})$ is a strict subset of $\mathsf{L}^2([a,b];\mathbb{F})$ this allows us to conclude that $(\mathsf{C}^0([a,b];\mathbb{F}), \langle\cdot,\cdot\rangle_2)$ is not complete. Moreover, one can show this explicitly following the arguments of Example 3.3.1–2; see Exercise 4.1.7. • what

The following conclusion for complete subspaces of inner product spaces is important. Note that definiteness of the inner product is essential here.

4.1.19 Theorem (Complete subspaces and direct sum decompositions) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *if* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *is an* $\mathbb{F}$*-inner product space, and if* U *is a complete subspace of* V*, then* $\mathsf{V} = \mathsf{U} \oplus \mathsf{U}^{\perp}$.

Proof We refer ahead to Theorem 4.1.25 for a characterisation of the minimisation of the distance from a point to a convex subset. There is nothing in that theorem that involves machinery not yet available to us.

Let $v_0 \in \mathsf{V}$ and, by Theorem 4.1.25, let $\hat{v}_0 \in \mathsf{U}$ be the unique vector such that

$$\|v_0 - \hat{v}_0\| = \inf\{\|v_0 - u\| \mid u \in \mathsf{U}\}.$$

We claim that $v_0 - \hat{v}_0 \in \mathsf{U}^{\perp}$.

First we do a little computation. Let $v \in \mathsf{V}$, let $u \in \mathsf{U} \setminus \{0_{\mathsf{V}}\}$, and let $a \in \mathbb{F}$. Then we compute

$$\begin{aligned}
\|v - au\|^2 &= \langle v - au, v - au\rangle \\
&= \|v\|^2 - a\langle u, v\rangle - \bar{a}\langle v, u\rangle + a\bar{a}\|u\|^2 \\
&= \|v\|^2 + \|u\|^2\left(a\bar{a} - a\frac{\overline{\langle v, u\rangle}}{\|u\|^2} - \bar{a}\frac{\langle v, u\rangle}{\|u\|^2}\right) \\
&= \|v\|^2 + \|u\|^2\left(a - \frac{\langle v, u\rangle}{\|u\|^2}\right)\left(\bar{a} - \frac{\overline{\langle v, u\rangle}}{\|u\|^2}\right) - \frac{|\langle v, u\rangle|^2}{\|u\|^2} \\
&= \|v\|^2 + \|u\|^2\left|a - \frac{\langle v, u\rangle}{\|u\|^2}\right|^2 - \frac{|\langle v, u\rangle|^2}{\|u\|^2}.
\end{aligned}$$

As a function of a this quantity is minimised when $a = a_0 \triangleq \frac{\langle v,u\rangle}{\|u\|^2}$ and the minimum value of the function is

$$\|v\|^2 - \frac{|\langle v, u\rangle|^2}{\|u\|^2}.$$

Now apply this to $v = v_0 - \hat{v}_0$ to give

$$\|v_0 - \hat{v}_0 - a_0 u\| = \|v_0 - \hat{v}_0\| - \frac{|\langle v_0 - \hat{v}_0, u\rangle|^2}{\|u\|^2} \tag{4.5}$$

for every $u \in \mathsf{U} \setminus \{0_{\mathsf{V}}\}$. By definition of $\hat{v}_0$ we have

$$\|v_0 - \hat{v}_0 + a_0 u\|^2 \geq \|v_0 - \hat{v}_0\|^2,$$

and from this and (4.5) we have

$$\frac{|\langle v_0 - \hat{v}_0, u\rangle|^2}{\|u\|^2} = 0 \implies |\langle v_0 - \hat{v}_0, u\rangle|^2 = 0$$

for all $u \in \mathsf{U} \setminus \{0_{\mathsf{V}}\}$. Thus $v_0 - \hat{v}_0 \in \mathsf{U}^{\perp}$, as claimed above.

Therefore, for every $v \in \mathsf{V}$ we can write $v = (v - \hat{v}) - \hat{v}$ where $\hat{v} \in \mathsf{U}$ and $v - \hat{v} \in \mathsf{U}^{\perp}$. Since $\mathsf{U} \cap \mathsf{U}^{\perp} = \{0_{\mathsf{V}}\}$ by Proposition 4.1.13 we have $\mathsf{V} = \mathsf{U} \oplus \mathsf{U}^{\perp}$, giving the theorem. ■

As concerns Hilbert spaces, we have the following result.

4.1.20 Corollary (Closed subspaces and direct sum decompositions) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, if* $(\mathsf{V}, \langle \cdot, \cdot \rangle)$ *is an* $\mathbb{F}$*-Hilbert space, and if* U *is a closed subspace of* V*, then* $\mathsf{V} = \mathsf{U} \oplus \mathsf{U}^\perp$*.*

Proof By Proposition 1.1.33 closed subspaces of Hilbert spaces are complete. Thus the result follows from Theorem 4.1.19. ∎

In finite dimensions the hypotheses of the theorem are always satisfied for any inner product.

4.1.21 Corollary (Orthogonal decompositions of finite-dimensional inner product spaces) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, if* $(\mathsf{V}, \langle \cdot, \cdot \rangle)$ *is a finite-dimensional* $\mathbb{F}$*-inner product space, and if* U *is a subspace of* V*, then* $\mathsf{V} = \mathsf{U} \oplus \mathsf{U}^\perp$*.*

Let us give an examples exploring the necessity of the hypotheses of the preceding results concerning direct sum decompositions.

4.1.22 Examples (Direct sum decomposition of inner product spaces)

1. The assumption in Theorem 4.1.19 that U is complete is essential. Indeed, consider the Hilbert space $(\ell^2(\mathbb{F}), \langle \cdot, \cdot \rangle_2)$ with

$$\langle (a_j)_{j\in\mathbb{Z}_{>0}}, (b_j)_{j\in\mathbb{Z}_{>0}} \rangle = \sum_{j=1}^\infty a_j \bar{b}_j.$$

   Take the subspace $\mathbb{F}_0^\infty$ which is not complete since its completion is $\ell^2(\mathbb{F})$ by Corollary 3.8.21. By Proposition 4.1.13 we have

$$(\mathbb{F}_0^\infty)^\perp = (\mathrm{cl}(\mathbb{F}_0^\infty))^\perp = \ell^2(\mathbb{F})^\perp = \{0_{\ell^2(\mathbb{F})}\}.$$

   Thus we have $\ell^2(\mathbb{F}) \neq \mathbb{F}_0^\infty \oplus (\mathbb{F}_0^\infty)^\perp$.

2. Let us now consider the necessity that V be a Hilbert space in Corollary 4.1.20. We consider the incomplete inner product space $(\mathbb{F}_0^\infty, \langle \cdot, \cdot \rangle_2)$ and the subspace

$$\mathsf{U} = \Big\{ (a_j)_{j\in\mathbb{Z}_{>0}} \;\Big|\; \sum_{j=1}^\infty \frac{a_j}{j} = 0 \Big\}.$$

   We leave to the reader the elementary verification that U is a proper subspace of $\mathbb{F}_0^\infty$.

   Let us verify that U is closed. Let $((a_{jl})_{j\in\mathbb{Z}_{>0}})_{l\in\mathbb{Z}_{>0}}$ be a sequence in U converging to $(a_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{F}_0^\infty$. Fix $j \in \mathbb{Z}_{>0}$ and let $\epsilon \in \mathbb{R}_{>0}$. Choose $N \in \mathbb{Z}_{>0}$ sufficiently large that

$$\|(a_j)_{j\in\mathbb{Z}_{>0}} - (a_{jl})_{j\in\mathbb{Z}_{>0}}\| < \epsilon$$

   for $l \geq N$. Then, for $l \geq N$,

$$|a_j - a_{jl}|^2 \leq \sum_{k=1}^\infty |a_k - a_{kl}|^2 = \|(a_k)_{k\in\mathbb{Z}_{>0}} - (a_{kl})_{k\in\mathbb{Z}_{>0}}\|^2 < \epsilon^2.$$

That is to say, $\lim_{l\to\infty} a_{jl} = a_j$ for each $j \in \mathbb{Z}_{>0}$. Define

$$b_{nl} = \sum_{j=1}^{n} \frac{a_{jl}}{j}.$$

We claim that the double sequence $(b_{nl})_{n,l\in\mathbb{Z}}$ converges to zero. Since $(a_j)_{j\in\mathbb{Z}_{>0}} \in \mathbb{F}_0^\infty$ there exists $N_1 \in \mathbb{Z}_{>0}$ such that $a_j = 0$ for $j > N_1$. Now let $\epsilon \in \mathbb{R}_{>0}$ and let $N_2 \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\|(a_j)_{j\in\mathbb{Z}_{>0}} - (a_{jl})_{j\in\mathbb{Z}_{>0}}\| < \frac{\epsilon}{M}$$

for $l \geq N_2$, where

$$M \triangleq \sum_{j=1}^{\infty} \frac{1}{j^2},$$

this series being summable by Example I-2.4.2–4. Then, using the Cauchy–Bunyakovsky–Schwarz inequality, for $l, n \geq \max\{N_1, N_2\}$,

$$\begin{aligned} |b_{nl}| = \Big|\sum_{j=1}^{n} \frac{a_{jl}}{j}\Big| &\leq \Big|\sum_{j=1}^{n} \frac{a_{jl} - a_j}{j}\Big| + \Big|\sum_{j=1}^{n} \frac{a_j}{j}\Big| \\ &\leq \Big(\sum_{j=1}^{n} |a_j - a_{jl}|^2\Big)^{1/2} \Big(\sum_{j=1}^{n} \frac{1}{j^2}\Big)^{1/2} \\ &\leq \Big(\sum_{j=1}^{\infty} |a_j - a_{jl}|^2\Big)^{1/2} \Big(\sum_{j=1}^{\infty} \frac{1}{j^2}\Big)^{1/2} < \epsilon, \end{aligned}$$

as desired. Then we have

$$\sum_{j=1}^{\infty} \frac{a_j}{j} = \sum_{j=1}^{\infty} \lim_{l\to\infty} \frac{a_{jl}}{j} = \lim_{n\to\infty} \lim_{l\to\infty} b_{nl} = 0,$$

using Proposition I-2.3.21. Thus we indeed have $(a_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{U}$ and so U is closed.

Now let us show that $\mathsf{U}^\perp = \{0_{\mathbb{F}_0^\infty}\}$. Let $(a_j)_{j\in\mathbb{Z}_{>0}} \in \mathsf{U}^\perp$ and let $N \in \mathbb{Z}_{>0}$ be such that $a_j = 0$ for $j > N$. Then define $(b_{jl})_{j\in\mathbb{Z}_{>0}} \in \mathsf{U}$, $l \in \{1, \dots, N+1\}$, by

$$b_{jl} = \begin{cases} -l, & j = l, \\ N+1, & j = N+1, \\ 0, & \text{otherwise.} \end{cases}$$

Then

$$\sum_{j=1}^{\infty} \frac{b_{jl}}{j} = -\frac{l}{l} + \frac{N+1}{N+1} = 0,$$

so $(b_{jl})_{j\in\mathbb{Z}_{>0}}$ is indeed in U for each $l \in \{1,\dots,N+1\}$. Moreover, for each $l \in \{1,\dots,N\}$,

$$0 = \langle (a_j)_{j\in\mathbb{Z}_{>0}}, (b_{jl})_{j\in\mathbb{Z}_{>0}}\rangle = -la_l,$$

and so $a_l = 0$ for $l \in \{1,\dots,N\}$. Thus $\mathsf{U}^\perp = \{0_{\mathbb{F}_0^\infty}\}$ as claimed. •

The preceding examples suggest that there is some sort of relationship between completeness of inner product spaces and properties of closed subspaces. Let us clarify this with the following result.

4.1.23 Theorem (Subspace characterisations of completeness of inner product spaces) *For $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ and for a $\mathbb{F}$-inner product space $(\mathsf{V},\langle\cdot,\cdot\rangle)$, the following statements are equivalent:*

(i) *V is a Hilbert space;*

(ii) *for every closed subspace U of V it holds that $\mathsf{V} = \mathsf{U}\oplus\mathsf{U}^\perp$;*

(iii) *for every closed subspace U of V it holds that $\mathsf{U} = (\mathsf{U}^\perp)^\perp$;*

(iv) *for every proper closed subspace U of V it holds that $\mathsf{U}^\perp \neq \{0_{\mathsf{V}}\}$.*

Proof (i) $\Longrightarrow$ (ii) This is Corollary 4.1.20.

(ii) $\Longrightarrow$ (iii) By Proposition 4.1.13 we have $\mathsf{U} \subseteq (\mathsf{U}^\perp)^\perp$. Now let $v \in (\mathsf{U}^\perp)^\perp$ and write $v = v_1 + v_2$ for $v_1 \in \mathsf{U}$ and $v_2 \in \mathsf{U}^\perp$. Then $v_2 = v - v_1 \in (\mathsf{U}^\perp)^\perp$ since $v \in (\mathsf{U}^\perp)^\perp$ and $v_1 \in \mathsf{U} \subseteq (\mathsf{U}^\perp)^\perp$. But this means that $v_2 \in \mathsf{U}^\perp \cap (\mathsf{U}^\perp)^\perp = \{0_{\mathsf{V}}\}$ and so $v = v_1 \in \mathsf{U}$.

(iii) $\Longrightarrow$ (iv) Let U be a subspace of V for which $\mathsf{U}^\perp = \{0_{\mathsf{V}}\}$. By assumption, $\mathsf{U} = \{0_{\mathsf{V}}\}^\perp = \mathsf{V}$. Thus U is not proper.

(iv) $\Longrightarrow$ (i) Let $\overline{\mathsf{V}}$ be a completion of V and regard V as a subspace of $\overline{\mathsf{V}}$. Let $\bar{v} \in \overline{\mathsf{V}}$. If $\bar{v} = 0_{\mathsf{V}}$ then $\bar{v} \in \mathsf{V}$. So suppose that $\bar{v} \neq 0_{\mathsf{V}}$. Define $f_{\bar{v}}\colon \mathsf{V} \to \mathbb{F}$ by $f_{\bar{v}}(u) = \langle u, \bar{v}\rangle$ noting that $f_{\bar{v}}$ is continuous by Proposition 4.2.1. Thus $\ker(f_{\bar{v}})$ is closed by Theorem 3.5.2, being the preimage of the closed set $\{0_{\mathsf{V}}\}$. We claim that $\ker(f_{\bar{v}})$ is a proper subspace. To see this, suppose that $\ker(f_{\bar{v}}) = \mathsf{V}$ and let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in V converging to $\bar{v}$. Then, by Theorem 3.5.2 and Proposition 4.2.1, we have

$$\langle \bar{v}, \bar{v}\rangle = \langle \lim_{j\to\infty} v_j, \bar{v}\rangle = \lim_{j\to\infty}\langle v_j, \bar{v}\rangle = 0,$$

contradicting the definiteness of the inner product. Thus we have $\ker(f_{\bar{v}}) \subset \mathsf{V}$. By assumption there exists $v' \in \ker(f_{\bar{v}})^\perp$ such that $\|v\| = 1$. One can verify, cf. the proof of Theorem 4.2.2 below, that if we take $v = \overline{f_v(v')}v'$ then $\langle u, \bar{v}\rangle = \langle u, v\rangle$ for every $u \in \mathsf{V}$. Thus $\langle u, \bar{v} - v\rangle = 0$ for every $u \in \mathsf{V}$ and so $\bar{v} = v$. Thus $\overline{\mathsf{V}} = \mathsf{V}$. ■

For other conditions equivalent to completeness we refer to Theorems 4.2.4 and 4.4.10.

4.1.5 Minimising distance to a set

One of the very interesting and useful features of inner product spaces is that they allow one to solve certain sorts of problems. In this section we consider the following problem.

4.1.24 Problem (Distance minimisation problem) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \|\cdot\|)$ be a $\mathbb{F}$-normed vector space. For $v_0 \in \mathsf{V}$ and for a subset $S \subseteq \mathsf{V}$ do the following:

(i) determine $\mathrm{dist}(v_0, S) \triangleq \inf\{\|v_0 - v\| \mid v \in S\}$;

(ii) ascertain whether there exists $\hat{v}_0 \in S$ such that $\|v_0 - \hat{v}_0\| = \mathrm{dist}(v_0, S)$. •

In general, the previous problem is too difficult to be approachable. There are a couple of reasons for this. First of all, by stating the problem for arbitrary subsets the problem is simply unreasonable. One really must place some additional structure on the set S. Below we will consider the case when S is convex. However, even if one restricts the set S to be something "reasonable," the problem can still be too difficult to solve. One of the reasons this may be so is that general norms are difficult to understand. The reader can explore this a little in the finite-dimensional situation in Exercise 4.1.14. However, if one restricts the norm to come from an inner product it turns out that it is possible to characterise the solutions to some distance minimisation problems in a useful way. Thus we restrict our attention in this section to the distance minimisation problem for inner product spaces.

The most accessible sufficiently interesting result concerns the minimisation of the distance from a point to a convex set. We dealt with convexity in $\mathbb{R}^n$ in detail in Section II-1.9 and in general vector spaces in Chapter 5. Here we simply recall that a convex subset of a $\mathbb{F}$-vector space V is a subset C for which

$$u, v \in C \quad \Longrightarrow \quad \{(1-s)u + sv \mid s \in [0,1]\} \subseteq C.$$

We then have the following result which gives a case where the distance minimisation problem possesses a unique solution.

4.1.25 Theorem (Minimisation of distance to convex subsets) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space, and let $\mathrm{v}_0 \in \mathsf{V}$. If $\mathsf{C} \subseteq \mathsf{V}$ is a complete convex set then there exists a unique vector $\hat{\mathrm{v}}_0 \in \mathsf{C}$ for which*

$$\|\mathrm{v}_0 - \hat{\mathrm{v}}_0\| = \mathrm{dist}(\mathrm{v}_0, \mathsf{C}).$$

Proof Denote $m = \mathrm{dist}(v, C)$ and let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in C such that $\|v_0 - v_j\|^2 < m^2 + \frac{1}{j}$. We claim that the set

$$\{v_0\} + \mathsf{C} = \{v_0 + v \mid v \in C\}$$

is convex. Indeed, if $v_0 + v_1, v_0 + v_2 \in \{v_0\} + \mathsf{C}$ for $v_1, v_2 \in \mathsf{C}$ and if $s \in [0,1]$ then

$$(1-s)(v_0 + v_1) + s(v_0 + v_2) = v_0 + (1-s)v_1 + sv_2 \in \{v_0\} + C.$$

Now, since $\{v_0\} + C$ is convex, for each $j, k \in \mathbb{Z}_{>0}$ we have $\left\|\frac{1}{2}((v_0 + v_j) + (v_0 + v_k))\right\|^2 \geq m^2$. Now let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that $\frac{4}{N} < \epsilon^2$. For $j, k \geq N$, using the parallelogram law we then have

$$\begin{aligned}\|v_j - v_k\|^2 &= \|(v_0 - v_j) - (v_0 - v_k)\|^2\\ &= 2\|v_0 - v_j\|^2 + 2\|v_0 - v_k\|^2 - 4\left\|\tfrac{1}{2}((v_0 + v_j) + (v_0 + v_k))\right\|^2\\ &< 2m^2 + \tfrac{2}{j} + 2m^2 + \tfrac{2}{k} - 4m^2 < \tfrac{4}{N} < \epsilon^2.\end{aligned}$$

Thus $\|v_j - v_k\| < \epsilon$ for $j, k \geq N$ and so $(v_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence. Since C is complete there exists $\hat{v}_0 \in C$ such that $(v_j)_{j\in\mathbb{Z}_{>0}}$ converges to $\hat{v}_0$. This gives the existence part of the lemma.

If $\hat{u}_0 \in C$ has the property that $\|v_0 - \hat{u}_0\| = m$ then, using the parallelogram law,

$$\|\hat{u}_0 - \hat{v}_0\|^2 = 2\|v_0 - \hat{u}_0\|^2 + 2\|v_0 - \hat{v}_0\|^2 \\ - 4\left\|\tfrac{1}{2}((v_0 + \hat{u}_0) + (v_0 + \hat{v}_0))\right\| \leq 2m^2 + 2m^2 - 4m^2 = 0.$$

Thus $\|\hat{u}_0 - \hat{v}_0\| = 0$ and so $\hat{u}_0 = \hat{v}_0$. ■

Since a subspace of a vector space is obviously convex we can immediately apply the preceding result to the case when C is a subspace. For subspaces, however, there is more that can be said about the character of the points that solve the distance minimisation problem: they are orthogonal to the subspace.

4.1.26 Theorem (Minimisation of distance to subspaces) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *be an* $\mathbb{F}$*-inner product with* $\mathrm{v}_0 \in \mathsf{V}$ *and* $\mathsf{U} \subseteq \mathsf{V}$ *a subspace. Then* $\hat{\mathrm{v}}_0 \in \mathsf{U}$ *satisfies*

$$\|\mathrm{v}_0 - \hat{\mathrm{v}}_0\| = \mathrm{dist}(\mathrm{v}_0, \mathsf{U}) \tag{4.6}$$

if and only if $\mathrm{v}_0 - \hat{\mathrm{v}}_0 \in \mathsf{U}^\perp$*. Furthermore, if* U *is complete then there exists a unique vector* $\hat{\mathrm{v}}_0 \in \mathsf{U}$ *such that (4.6) holds.*

Proof First suppose that $v_0 - \hat{v}_0 \in \mathsf{U}^\perp$. Then, since $\hat{v}_0 - u \in \mathsf{U}$ for any $u \in \mathsf{U}$, $v_0 - \hat{v}_0$ and $\hat{v}_0 - u$ are orthogonal. The Pythagorean identity (Exercise 4.1.12) then gives

$$\|v_0 - u\|^2 = \|v_0 - \hat{v}_0\|^2 + \|\hat{v}_0 - u\|^2$$

for any $u \in \mathsf{U}$. From this we conclude that $\|v_0 - \hat{v}_0\|^2 \leq \|v_0 - u\|^2$ for every $u \in \mathsf{U}$. This exactly means that $\hat{v}_0$ satisfies (4.6).

Now suppose that $\hat{v}_0$ satisfies (4.6). Let $\alpha \in \mathbb{F} \setminus \{0\}$ and define $f_\alpha \colon \mathsf{U} \to \mathsf{U}$ by $f_\alpha(u) = \hat{v}_0 + \alpha(u - \hat{v}_0)$. Since $\hat{v}_0$ satisfies (4.6) we have

$$\begin{aligned} \|v_0 - \hat{v}_0\|^2 &\leq \|v_0 - f_\alpha(u)\|^2 \\ &= \|(v_0 - \hat{v}_0) - \alpha(u - \hat{v}_0)\|^2 \\ &= \|v_0 - \hat{v}_0\|^2 + |\alpha|^2\|u - \hat{v}_0\| - \alpha\langle u - \hat{v}_0, v_0 - \hat{v}_0\rangle - \bar{\alpha}\langle v_0 - \hat{v}_0, u - \hat{v}_0\rangle. \end{aligned}$$

From this we conclude that

$$\alpha\langle u - \hat{v}_0, v_0 - \hat{v}_0\rangle + \bar{\alpha}\overline{\langle u - \hat{v}_0, v_0 - \hat{v}_0\rangle} \leq |\alpha|^2\|u - \hat{v}_0\| \tag{4.7}$$

for every $u \in \mathsf{U}$. Now we write $\alpha = |\alpha|e^{\mathrm{i}\theta}$ for $\theta \in (-\pi, \pi]$. If $\mathbb{F} = \mathbb{R}$ we restrict to $\theta \in \{0, \pi\}$. Now divide (4.7) by $|\alpha|$ and take the limit as $|\alpha| \to 0$. Also note that

$$\{u - \hat{v}_0 \mid u \in \mathsf{U}\} = \mathsf{U}.$$

Putting this all together gives

$$e^{\mathrm{i}\theta}\langle u, v_0 - \hat{v}_0\rangle + e^{-\mathrm{i}\theta}\overline{\langle u, v_0 - \hat{v}_0\rangle} \leq 0,$$

which again holds for all $u \in \mathsf{U}$ and $\theta \in (-\pi, \pi]$. Taking $\theta = 0$ gives

$$2\,\mathrm{Re}(\langle u, v_0 - \hat{v}_0\rangle) \le 0,$$

and taking $\theta = \pi$ gives

$$-2\,\mathrm{Re}(\langle u, v_0 - \hat{v}_0\rangle) \le 0$$

for all $u \in \mathsf{U}$. From this we conclude that $\mathrm{Re}(\langle u, v_0 - \hat{v}_0\rangle) = 0$ for all $u \in \mathsf{U}$. A similar argument, using $\theta = \frac{\pi}{2}$ and $\theta = -\frac{\pi}{2}$, gives $\mathrm{Im}(\langle u, v_0 - \hat{v}_0\rangle) = 0$. Thus $v_0 - \hat{v}_0 \in \mathsf{U}^\perp$, as desired.

The final assertion of the theorem follows directly from Theorem 4.1.25. ■

The preceding result is insightful as it gives us a concrete description of the set of points that minimise the distance from a vector v_0 to a subspace U. This description will be important for us in Section 4.4.4 subsequently for applications of the ideas in Section 4.4.4. You will observe that the most difficult part of Theorem 4.1.26 is showing that the set of points minimising the distance is nonempty, and in fact contains a single point, at least when U is complete. In finite-dimensions, these issues are not so complicated, as can be seen in Exercise 4.1.15.

nonexistence for nonclosed subspaces

4.1.6 Norms

Theorem 4.1.9 was proved by John von Neumann.

Example 4.1.22–2 is taken from [Gudder 1974], as are the characterisations of completeness in Theorem 4.1.23.

Exercises

4.1.1 Prove Proposition 4.1.2.

4.1.2 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a $\mathbb{F}$-(semi-)inner product space. For a subspace $\mathsf{U} \subseteq \mathsf{V}$, define a map from $\mathsf{U}\times\mathsf{U}$ to $\mathbb{F}$ by $(u_1, u_2) \mapsto \langle u_1, u_2\rangle \triangleq \langle u_1, u_2\rangle_{\mathsf{U}}$. Show that $(\mathsf{U}, \langle\cdot,\cdot\rangle_{\mathsf{U}})$ is an $\mathbb{F}$-(semi-)inner product space.

4.1.3 Show that a $\mathbb{C}$-inner product space is always a $\mathbb{R}$-inner product space, using the fact that a $\mathbb{C}$-vector space is always a $\mathbb{R}$-vector space.

4.1.4 Answer the following three questions.

(a) Show that the norm defined by an inner product satisfies the parallelogram law.

(b) Show that the norm defined in Example 3.1.3–4 does not come from an inner product.

(c) Give an interpretation of the parallelogram law in $\mathbb{R}^2$ with the standard inner product.

4.1.5 Show using the parallelogram law that the norms $\|\cdot\|_1$ and $\|\cdot\|_\infty$ on $\mathbb{F}^n$ are not derived from an inner product if $n \ge 2$.

4.1.6 Show explicitly (i.e., as is done in Example 3.3.1–1) that $(\mathbb{F}_0^\infty, \langle\cdot,\cdot\rangle_2)$ is not complete.

4.1.7 Show explicitly (i.e., as is done in Example 3.3.1–2) that $(\mathsf{C}^0([a,b],\mathbb{F}),\langle\cdot,\cdot\rangle_2)$ is not complete.

4.1.8 Let $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and let $(\mathsf{V},\|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Show that the following two assertions are equivalent:

(i) there exists a (semi-)inner product $\langle\cdot,\cdot\rangle$ on V such that $\|v\|=\sqrt{\langle v,v\rangle}$ for every $v\in\mathsf{V}$;

(ii) the expression

$$\|u+v+w\|^2+\|u+v-w\|^2-\|u-v-w\|^2-\|u-v+w\|^2$$

is independent of w.

Hint: *Use Theorem 4.1.9.*

4.1.9 Let $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and let $(\mathsf{V},\|\cdot\|)$ be a (semi)normed $\mathbb{F}$-vector space. Show that the following two assertions are equivalent:

(i) there exists a (semi-)inner product $\|\cdot\|\cdot$ on V such that $\|v\|=\sqrt{\langle v,v\rangle}$ for every $v\in\mathsf{V}$;

(ii) the function $s\mapsto\|u+sv\|^2$ is a polynomial function of degree 2 for every $u,v\in\mathsf{V}$.

Hint: *Use Theorem 4.1.9.*

4.1.10 Let $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and let $(\mathsf{V},\langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space. Show that if subsets $A,B\subseteq\mathsf{V}$ are orthogonal then so too are the subsets $\operatorname{span}_{\mathbb{F}}(A)$ and $\operatorname{span}_{\mathbb{F}}(B)$.

4.1.11 Prove parts (i), (ii), and (iii) of Proposition 4.1.13.

4.1.12 Let $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$ and let $(\mathsf{V},\langle\cdot,\cdot\rangle)$ be a $\mathbb{F}$-semi-inner product space.

(a) Prove the ***Pythagorean identity***:

$$\|v_1+v_2\|^2=\|v_1\|^2+\|v_2\|^2$$

if v_1 and v_2 are orthogonal.

(b) Show that if $\mathbb{F}=\mathbb{R}$ then the Pythagorean identity for v_1 and v_2 implies that v_1 and v_2 are orthogonal.

(c) Give an example showing that the assertion in part (b) is generally false if $\mathbb{F}=\mathbb{C}$.

4.1.13 For $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$, for an $\mathbb{F}$-inner product space $(\mathsf{V},\langle\cdot,\cdot\rangle)$, and for a subspace $\mathsf{U}\subseteq\mathsf{V}$, answer the following two questions.

(a) Show that $\mathsf{U}\cap\mathsf{U}^\perp=\{0\}$.

(b) Show that if U is closed then for every $v\in\mathsf{V}$ there exists unique vectors $u_1\in\mathsf{U}$ and $u_2\in\mathsf{U}^\perp$ so that $v=u_1+u_2$.

4.1.14 Consider the Banach space $(\mathbb{R}^2,\|\cdot\|_2)$ of Example 3.1.3–2 and the Banach space $(\mathbb{R}^2,\|\cdot\|_\infty)$ of Example 3.1.3–4. For each of these norms, and for the subsets S and the points $\boldsymbol{v}_0$ given below, determine $\operatorname{dist}(\boldsymbol{v}_0,S)$ and determine the set of points $\hat{\boldsymbol{v}}_0\in S$ such that $\|\boldsymbol{v}_0-\hat{\boldsymbol{v}}_0\|=\operatorname{dist}(\boldsymbol{v}_0,S)$.

(a) $v_0 = (0, 1)$ and $S = \{(v_1, 0) \mid v_1 \in [-1, 1]\}$.
(b) $v_0 = (0, 1)$ and $S = \mathrm{span}_{\mathbb{R}}((1, 0))$.
(c) $v_0 = (0, 1)$ and $S = \mathrm{span}_{\mathbb{R}}((1, 1))$.
(d) $v_0 = (0, 0)$ and $S = \{(v_1, v_2) \mid v_1^2 + v_2^2 \geq 1\}$.
(e) $v_0 = (0, 0)$ and $S = \{(v_1, v_2) \mid v_1^2 + v_2^2 > 1\}$.

In the next exercise you will prove Theorem 4.1.26 when V is finite-dimensional. As you will see, it is possible to be somewhat more concrete in this case, making you appreciate that there is something real happening in the proof of Theorem 4.1.26.

4.1.15 Let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a finite-dimensional inner product space, and let $v_0 \in \mathsf{V}$ with $\mathsf{U} \subseteq \mathsf{V}$ a subspace. Provide a proof of Theorem 4.1.26 in this case along the following lines.

1. Argue that the result is trivial unless $v_0 \notin \mathsf{U}$. Thus assume this for the remainder of the proof.
2. For a subspace $\mathsf{U} \subseteq \mathsf{V}$ let $\{u_1, \dots, u_m\}$ be an orthonormal basis for U. Can this always be done?
3. Extend the basis from the previous part of the question to an orthonormal basis $\{v_1 = u_1, \dots, v_m = u_m, v_{m+1}, \dots, v_n\}$ for V. Can this always be done?
4. As a function on U, use the above basis to explicitly write down the function defining the distance from U to v_0.
5. Show that the unique point in U that minimises the distance function is

$$\hat{v}_0 = \sum_{j=1}^{m} \langle v, u_j\rangle u_j.$$

Section 4.2

Continuous maps between inner product spaces

Inner product spaces, being normed vector spaces, are of course subject to all the definitions and results concerning maps between normed vector spaces as stated in Section 3.5. We shall take all of these definitions and results for granted, and instead emphasise the things that are distinctive for inner product spaces.

Do I need to read this section? The results in this section complement those of Section 3.5, and so should be absorbed if one is in the business of understanding continuous maps between infinite-dimensional spaces. •

4.2.1 The dual of an inner product space

Much of the special character of inner product spaces, as opposed to more general normed vector spaces, is reflected in the structure of the topological dual of an inner product space. In order to understand this it is useful to first record some elementary properties of inner products.

4.2.1 Proposition (Continuity properties of operations in an inner product space) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an inner product space. Then the following maps are uniformly continuous:*

(i) $\mathsf{V} \ni \mathrm{v} \mapsto \langle \mathrm{v}, \mathrm{v}_0 \rangle \in \mathbb{F}$ *for* $\mathrm{v}_0 \in \mathsf{V}$;

(ii) $\mathsf{V} \ni \mathrm{v} \mapsto \langle \mathrm{v}_0, \mathrm{v} \rangle \in \mathbb{F}$ *for* $\mathrm{v}_0 \in \mathsf{V}$;

(iii) $\mathbb{F} \ni \mathrm{a} \mapsto \langle \mathrm{a}\mathrm{v}_1, \mathrm{v}_2 \rangle \in \mathbb{F}$ *for* $\mathrm{v}_1, \mathrm{v}_2 \in \mathsf{V}$;

(iv) $\mathbb{F} \ni \mathrm{a} \mapsto \langle \mathrm{v}_1, \mathrm{a}\mathrm{v}_2 \rangle \in \mathbb{F}$ *for* $\mathrm{v}_1, \mathrm{v}_2 \in \mathsf{V}$.

Proof (i) If $v_0 = 0_{\mathsf{V}}$ the assertion is clearly true as the map is the constant map with value zero. Thus consider $v_0 \neq 0_{\mathsf{V}}$. Let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta = \frac{\epsilon}{\|v_0\|}$. Then, using the Cauchy–Bunyakovsky–Schwarz inequality,

$$|\langle v_1, v_0 \rangle - \langle v_2, v_0 \rangle| = |\langle v_1 - v_2, v_0 \rangle| \leq \|v_1 - v_2\| \|v_0\| \leq \epsilon$$

for $\|v_1 - v_2\| < \delta$.

(ii) Conjugation $a \mapsto \bar{a}$ is clearly uniformly continuous. Therefore, $v \mapsto \langle v_0, v \rangle = \overline{\langle v, v_0 \rangle}$ is uniformly continuous, being a composition of uniformly continuous maps.

(iii) If $\langle v_1, v_2 \rangle = 0$ then clearly the given map is continuous since it is the constant map with value zero. So suppose that $\langle v_1, v_2 \rangle$ is nonzero. Let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta = \frac{\epsilon}{|\langle v_1, v_2 \rangle|}$. Then

$$|\langle a_1 v_1, v_2 \rangle - \langle a_2 v_1, v_2 \rangle| = |\langle (a_1 - a_2) v_1, v_2 \rangle| = |a_1 - a_2| |\langle v_1, v_2 \rangle| \leq \epsilon$$

for $|a_1 - a_2| < \delta$.

(iv) This follows from part (iii) as part (ii) follows from (i). ■

The central result concerning the dual of an inner product space is then the following.

4.2.2 Theorem (Riesz Representation Theorem) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a Hilbert space with topological dual V^*. If $\alpha \in \mathsf{V}^*$ then there exists a unique $\mathrm{v}_\alpha \in \mathsf{V}$ such that $\langle \mathrm{u}, \mathrm{v}_\alpha\rangle = \alpha(\mathrm{u})$ for every $\mathrm{u} \in \mathsf{V}$.*

Proof If $\alpha = 0$ then we can take $v_\alpha = 0$. So let $\alpha \in \mathsf{V}^* \setminus \{0\}$. We claim that $\ker(\alpha)$ is a closed subspace of V. It is certainly a subspace. To show that it is closed, let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\ker(\alpha)$ converging to $v_0 \in \mathsf{V}$. Then, by continuity of α and Theorem 3.5.2 we have

$$\alpha(v_0) = \alpha\Big(\lim_{j\to\infty} v_j\Big) = \lim_{j\to\infty} \alpha(v_j) = 0.$$

Thus $v_0 \in \ker(\alpha)$ and so $\ker(\alpha)$ is closed by Proposition 3.6.8. Since $\alpha \neq 0$, $\ker(\alpha) \neq \mathsf{V}$. By Theorem 4.1.19, since $\ker(\alpha)$ is closed we can choose a nonzero vector $v_0 \in \ker(\alpha)^\perp$, supposing this vector to further have length 1. We claim that we can take $v_\alpha = \bar{\alpha}(v_0)v_0$, where $\bar{\alpha}\colon \mathsf{V} \to \mathbb{F}$ is defined by $\bar{\alpha}(v) = \overline{\alpha(v)}$. Indeed note that for $u \in \mathsf{V}$ the vector $\alpha(u)v_0 - \alpha(v_0)u$ is in $\ker(\alpha)$. Therefore

$$0 = \langle \alpha(u)v_0 - \alpha(v_0)u, v_0\rangle = \alpha(u) - \alpha(v_0)\langle u, v_0\rangle.$$

Thus

$$\alpha(u) = \langle u, \bar{\alpha}(v_0)v_0\rangle = \langle u, v_\alpha\rangle.$$

Thus v_α as defined meets the desired criterion. Let us show that this is the only vector satisfying the conditions of the theorem. Suppose that $v_1, v_2 \in \mathsf{V}$ have the property that $\alpha(u) = \langle u, v_1\rangle = \langle u, v_2\rangle$ for all $u \in \mathsf{V}$. Then $\langle u, v_1 - v_2\rangle = 0$ for all $u \in \mathsf{V}$. In particular, taking $u = v_1 - v_2$ we have $\|v_1 - v_2\|^2 = 0$, giving $v_1 = v_2$. ■

The assumption that V is a Hilbert space is essential as the following example shows.

4.2.3 Example (The dual of an incomplete inner product space) Let us consider the $\mathbb{F}$-inner product space $(\mathbb{F}_0^\infty, \langle\cdot,\cdot\rangle_2)$ where, we recall, that

$$\langle (a_j)_{j\in\mathbb{Z}_{>0}}, (b_j)_{j\in\mathbb{Z}_{>0}}\rangle = \sum_{j=1}^\infty a_j\bar{b}_j;$$

the sum is finite. Recall from Proposition I-5.7.5 that $(\mathbb{F}_0^\infty)' = \mathbb{F}^\infty$ and so $(\mathbb{F}_0^\infty)^*$ is a subspace of $\mathbb{F}^\infty$. Define $\boldsymbol{\alpha} \in \mathbb{F}^\infty$ by $\boldsymbol{\alpha}(j) = \frac{1}{j}$ for each $j \in \mathbb{Z}_{>0}$. By Example I-2.4.2–4 note that

$$(\tfrac{1}{j})_{j\in\mathbb{Z}_{>0}} \in \ell^2(\mathbb{F}) \subseteq \mathbb{F}^\infty \quad\implies\quad M^2 \triangleq \sum_{j=1}^\infty \frac{1}{j^2} < \infty.$$

(In fact, $M^2 = \frac{\pi^2}{6}$ but this precise number is not important for us, only that it is finite.)

We claim that $\boldsymbol{\alpha}$ is a continuous linear function on $\mathbb{F}_0^\infty$. Indeed, let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta = \frac{\epsilon}{M}$. Let $\boldsymbol{a} = (a_j)_{j\in\mathbb{Z}_{>0}}, \boldsymbol{b} = (b_j)_{j\in\mathbb{Z}_{>0}}$ be such that

$$\|(a_j)_{j\in\mathbb{Z}_{>0}} - (b_j)_{j\in\mathbb{Z}_{>0}}\|_2 < \delta.$$

Then, using the Cauchy–Bunyakovsky–Schwarz inequality,

$$|\boldsymbol{\alpha}(\boldsymbol{a}) - \boldsymbol{\alpha}(\boldsymbol{b})| = |\boldsymbol{\alpha}(\boldsymbol{a} - \boldsymbol{b})| = \Big|\sum_{j=1}^{\infty} \frac{a_j - b_j}{j^2}\Big| \leq \Big(\sum_{j=1}^{\infty} |a_j - b_j|^2\Big)^{1/2} \Big(\sum_{j=1}^{\infty} \frac{1}{j^2}\Big)^{1/2} < \epsilon.$$

Thus $\boldsymbol{\alpha}$ is indeed continuous.

We next claim that there exists no $\boldsymbol{f}_{\boldsymbol{\alpha}} \in \mathbb{F}_0^\infty$ such that $\langle \boldsymbol{f}_{\boldsymbol{\alpha}}, \boldsymbol{a}\rangle = \boldsymbol{\alpha}(\boldsymbol{a})$ for every $\boldsymbol{a} \in \mathbb{F}_0^\infty$. To see this, let $(\boldsymbol{e}_j)_{j\in\mathbb{Z}_{>0}}$ be the standard basis for $\mathbb{F}_0^\infty$ so that $\boldsymbol{e}_j(k) = 1$ for $j = k$ and 0 otherwise. Then, if $\boldsymbol{f}_{\boldsymbol{\alpha}} \in \mathbb{F}_0^\infty$ we have $\langle \boldsymbol{f}_{\boldsymbol{\alpha}}, \boldsymbol{e}_j\rangle = \boldsymbol{f}_{\boldsymbol{\alpha}}(j)$. Also, $\boldsymbol{\alpha}(\boldsymbol{e}_j) = \frac{1}{j}$ for each $j \in \mathbb{Z}_{>0}$. Thus if $\boldsymbol{f}_{\boldsymbol{\alpha}} \in \mathbb{F}^\infty$ has the property that $\langle \boldsymbol{f}_{\boldsymbol{\alpha}}, \boldsymbol{e}_j\rangle = \boldsymbol{\alpha}(\boldsymbol{e}_j)$ for every $j \in \mathbb{Z}_{>0}$ then it follows that $\boldsymbol{f}_{\boldsymbol{\alpha}}(j) = \frac{1}{j}$ for each $j \in \mathbb{Z}_{>0}$. But this means that $\boldsymbol{f}_{\boldsymbol{\alpha}}(j) \notin \mathbb{F}_0^\infty$. •

The preceding example is, actually, representative of the general situation in the sense of the following result which states that the assumption that V be a Hilbert space is essential in the Riesz Representation Theorem.

4.2.4 Theorem (The Riesz Representation Theorem does not hold for non-Hilbert spaces) *For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and for a $\mathbb{F}$-inner product space $(\mathsf{V}, \langle\cdot,\cdot\rangle)$, the following statements are equivalent:*

(i) V is a Hilbert space;

(ii) for every $\alpha \in \mathsf{V}^*$ there exists $\mathrm{v}_\alpha \in \mathsf{V}$ such that $\langle \mathrm{u}, \mathrm{v}_\alpha\rangle = \alpha(\mathrm{u})$ for every $\mathrm{u} \in \mathsf{V}$.

Proof That (i) $\implies$ (ii) is simply Theorem 4.2.2, so we need only prove the converse. Thus we let $\overline{\mathsf{V}}$ be a completion of V, let $\bar{v} \in \overline{\mathsf{V}}$, and define $f_{\bar{v}}\colon \mathsf{V} \to \mathbb{F}$ by $f_{\bar{v}}(u) = \langle u, \bar{v}\rangle$. By Proposition 4.2.1 it follows that $f_{\bar{v}}$ is continuous. By assumption there exists $v \in \mathsf{V}$ such that $\langle u, v\rangle = f_{\bar{v}}(u) = \langle u, \bar{v}\rangle$ for every $u \in \mathsf{V}$. Thus $v = \bar{v}$ and so $\overline{\mathsf{V}} = \mathsf{V}$. ■

Let us examine a consequence of the Riesz Representation Theorem.

4.2.5 Corollary (The dual of a Hilbert space) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-Hilbert space. Then the map $\alpha \mapsto \mathrm{v}_\alpha$ from V^* to V is an isomorphism of $\mathbb{R}$-normed vector spaces that further satisfies $\mathrm{v}_{a\alpha} = \bar{a}\mathrm{v}_\alpha$.*

Proof According to the proof of Theorem 4.2.2 we have $v_\alpha = \bar{\alpha}(v_0)v_0$, where $\bar{\alpha} \in \mathsf{V}^*$ is defined by $\bar{\alpha}(v) = \overline{\alpha(v)}$ and where v_0 is a fixed vector of unit length in $\ker(\alpha)^\perp$. The conclusions of the corollary are directly verified. ■

Note that V^* are V are not isomorphic as $\mathbb{F}$-vector spaces in the case when $\mathbb{F} = \mathbb{C}$. Sometimes the property of a linear map $\mathsf{L}\colon \mathsf{U} \to \mathsf{V}$ that

1. $\mathsf{L}(u_1 + u_2) = \mathsf{L}(u_1) + \mathsf{L}(u_2)$, $u_1, u_2 \in \mathsf{U}$, and
2. $\mathsf{L}(au) = \bar{a}\mathsf{L}(u)$, $a \in \mathbb{F}$, $u \in \mathsf{U}$,

is called ***conjugate linearity*** and agree with the property of linearity if and only if $\mathbb{F} = \mathbb{R}$.

Let us examine the Riesz Representation Theorem in a few special cases.

4.2.6 Examples (Riesz Representation Theorem)

1. Let us consider the inner product space $(\mathbb{F}^n, \langle\cdot,\cdot\rangle_2)$. We represent an element $\boldsymbol{\alpha} \in (\mathbb{F}^n)^*$ by a $1 \times n$ matrix:

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}(1) & \cdots & \boldsymbol{\alpha}(n) \end{bmatrix}.$$

The vector $\boldsymbol{v}_{\boldsymbol{\alpha}} \in \mathbb{F}^n$ corresponding to $\boldsymbol{\alpha}$ must then satisfy

$$\begin{aligned} & \boldsymbol{\alpha}(\boldsymbol{u}) = \langle \boldsymbol{u}, \boldsymbol{v}_{\boldsymbol{\alpha}} \rangle, \qquad \boldsymbol{u} \in \mathbb{F}^n \\ \implies \quad & \sum_{j=1}^n \boldsymbol{\alpha}(j)\boldsymbol{u}(j) = \sum_{j=1}^n \boldsymbol{u}(j)\overline{\boldsymbol{v}_{\boldsymbol{\alpha}}(j)}, \qquad \boldsymbol{u} \in \mathbb{F}^n \\ \implies \quad & \boldsymbol{v}_{\boldsymbol{\alpha}}(j) = \overline{\boldsymbol{\alpha}(j)}, \qquad j \in \{1,\dots,n\}. \end{aligned}$$

2. Next we consider the Hilbert space $(\ell^2(\mathbb{F}), \langle\cdot,\cdot\rangle_2)$ and let $\alpha \in \ell^2(\mathbb{F})^*$. Then Corollary 4.2.5 ensures that there exists $\boldsymbol{v}_\alpha \in \ell^2(\mathbb{F})$ such that

$$\alpha(\boldsymbol{u}) = \sum_{j=1}^\infty \boldsymbol{u}(j)\overline{\boldsymbol{v}_\alpha(j)}$$

for every $\boldsymbol{u} \in \ell^2(\mathbb{F})$. From this expression we easily see that $\boldsymbol{v}_\alpha(j) = \overline{\alpha(\boldsymbol{e}_j)}$, $j \in \mathbb{Z}_{>0}$, where $\{\boldsymbol{e}_j\}_{j\in\mathbb{Z}_{>0}}$ is the standard basis for $\mathbb{F}_0^\infty$.

3. Finally, we consider the Hilbert space $(\mathsf{L}^2([a,b];\mathbb{F}), \langle\cdot,\cdot\rangle_2)$. If $\alpha \in \mathsf{L}^2([a,b];\mathbb{F})^*$ then Corollary 4.2.5 ensures that there exists $f_\alpha \in \mathsf{L}^2([a,b];\mathbb{F})$ such that

$$\alpha(g) = \int_a^b g(x) f_\alpha(x)\,\mathrm{d}x$$

for every $g \in \mathsf{L}^2([a,b];\mathbb{F})$. To extract a more explicit characterisation of f_α is possible once one has on hand the notion of a maximal orthonormal family. We refer to Exercise 4.4.8 for a working out of this characterisation. •

4.2.2 Particular aspects of continuity for inner product spaces

To get started we give a few constructions concerning linear maps between inner product spaces that are specific to the inner product structure. We begin with the notion of the adjoint of a continuous linear map.

4.2.7 Definition

4.2.8 Remark (Self-adjointness in Sturm–Liouville[2] theory) One of the important areas of application of inner product spaces is in so-called "Sturm–Liouville theory," which deals with a certain sort of ordinary differential equation. In this subject one is interested in linear maps that are self-adjoint. The sort of maps that arise in Sturm–Liouville theory are *not* of the sort coming from the preceding definition. There are many reasons why this is so, and we refer the reader to for details. We mention this here because in reading some elementary treatments of Sturm–Liouville theory one might be led to believe that the theory has to do with the more or less simple situation of Definition 4.2.7. • where?

4.2.3 The adjoint of a continuous linear map

Consider two $\mathbb{F}$-inner product spaces $(\mathsf{U}, \langle\cdot,\cdot\rangle_{\mathsf{U}})$ and $(\mathsf{V}, \langle\cdot,\cdot\rangle_{\mathsf{V}})$. For fixed $v \in \mathsf{V}$ consider the map from U to $\mathbb{F}$ given by $u \mapsto \langle \mathsf{L}(u), v\rangle$. Since the inner product is continuous (Proposition 4.2.1) this map is an element of U^*. In this way we assign to each $v \in \mathsf{V}$ an element $\alpha_v \in \mathsf{U}^*$. By the Riesz representation theorem this therefore defines an element $u_v \in \mathsf{U}$. In other words, we have defined a map $\mathsf{L}^* \colon \mathsf{V} \to \mathsf{U}$. It is a straightforward exercise, given as Exercise 4.2.1, to show that L^* is linear. We call L^* the ***adjoint*** of L in this case.

Let us consider an example of an adjoint defined on an infinite-dimensional vector space.

4.2.9 Example On $\mathsf{L}^2([0,1];\mathbb{F})$ we consider the linear transformation defined by $\mathsf{L}(f)(t) = tf(t)$, as in Example 3.5.28–3. We showed in that preceding example that L is continuous, so it certainly possesses an adjoint as we describe here. Let $f, g \in \mathsf{V}$ and compute

$$\langle \mathsf{L}(f), g\rangle = \int_0^1 tf(t)\overline{g(t)}\,\mathrm{d}t = \int_0^1 f(t)\overline{tg(t)}\,\mathrm{d}t = \langle f, \mathsf{L}^*(g)\rangle$$

where $\mathsf{L}^*(g)(t) = tg(t)$. Thus we see in this case that $\mathsf{L}^* = \mathsf{L}$. •

The following results might help in understanding the adjoint, telling us what it looks like in $\mathbb{F}^n$ with the inner product being the dot product.

4.2.10 Proposition *Consider the inner product on* $\mathbb{F}^{\mathrm{n}}$ *given by the dot product:*

$$\langle \mathbf{x}, \mathbf{y}\rangle = \mathbf{x}\cdot\bar{\mathbf{y}} = \sum_{\mathrm{j}=1}^{\mathrm{n}} \mathrm{x}_{\mathrm{i}}\bar{\mathrm{y}}_{\mathrm{i}}.$$

If $\mathbf{L} \in \mathrm{L}(\mathbb{F}^{\mathrm{n}};\mathbb{F}^{\mathrm{m}})$ *is a linear map (i.e., an* $\mathrm{m}\times\mathrm{n}$ *matrix with entries in* $\mathbb{F}$*), then* $\mathbf{L}^* = \bar{\mathbf{L}}^{\mathrm{T}}$*. That is, the matrix corresponding to the linear map* $\mathbf{L}^*$ *is obtained by taking the conjugate of all entries in the transpose* $\mathbf{L}^{\mathrm{T}}$*.*

[2]Friedrich Otto Rudolf Sturm (1841–1919) was a German mathematician whose contributions were mainly in the area of geometry. Joseph Liouville (1809–1882) was a French mathematician who made contributions to many areas of mathematics and its applications. These areas include mathematical physics, differential equations, number theory, and analysis.

Proof One may write the dot product in terms of matrix multiplication like this:

$$\boldsymbol{x}\cdot\bar{\boldsymbol{y}} = \boldsymbol{x}^T\bar{\boldsymbol{y}}.$$

The definition of adjoint is then as follows. For $\boldsymbol{x}\in\mathbb{F}^m$, $\mathbf{L}^*\boldsymbol{x}\in\mathbb{F}^n$ satisfies

$$(\mathbf{L}^*\boldsymbol{x})\cdot\bar{\boldsymbol{y}} = \boldsymbol{x}\cdot\left(\overline{\mathbf{L}\boldsymbol{y}}\right)$$

for every $\boldsymbol{y}\in\mathbb{F}^n$. Using the matrix multiplication characterisation of the dot product, this gives, for every $\boldsymbol{y}\in\mathbb{F}^n$,

$$\begin{aligned} & (\mathbf{L}^*\boldsymbol{x})^T\bar{\boldsymbol{y}} = \boldsymbol{x}^T\left(\overline{\mathbf{L}\boldsymbol{y}}\right)\\ \implies\quad & \left(\boldsymbol{x}^T(\mathbf{L}^*)^T\right)\bar{\boldsymbol{y}} = \boldsymbol{x}^T\left(\bar{\mathbf{L}}\bar{\boldsymbol{y}}\right)\\ \implies\quad & \left(\boldsymbol{x}^T(\mathbf{L}^*)^T\right)\bar{\boldsymbol{y}} = \left(\boldsymbol{x}^T\bar{\mathbf{L}}\right)\bar{\boldsymbol{y}}. \end{aligned}$$

Since this must be true for every $\boldsymbol{y}\in\mathbb{F}^n$ we can assert that

$$\begin{aligned} & \boldsymbol{x}^T(\mathbf{L}^*)^T = \boldsymbol{x}^T\bar{\mathbf{L}}\\ \implies\quad & \mathbf{L}^*\boldsymbol{x} = \bar{\mathbf{L}}^T\boldsymbol{x}. \end{aligned}$$

Thus we have shown that $\mathbf{L}^* = \bar{\mathbf{L}}^T$, as desired. ■

Thus, if $\mathbb{F}=\mathbb{R}$, a self-adjoint linear on $\mathbb{R}^n$ is simply a symmetric matrix. However, our principal interest is in understanding self-adjoint maps in the case when V is infinite-dimensional.

4.2.4 Spectral properties for operators on Hilbert spaces

The eigenvalues and eigenvectors of a self-adjoint or symmetric linear operator have some useful properties. Let us first consider eigenvalues for symmetric linear operators. Note that the following result does *not* say that a symmetric linear *has* eigenvalues.

4.2.11 Theorem *Let* $(\mathsf{V},\langle\cdot,\cdot\rangle)$ *be an* $\mathbb{F}$*-inner product space and let* $(\mathsf{L},\mathrm{dom}(\mathsf{L}))$ *be a symmetric linear transformation on* V*. The following statements hold:*

(i) $\langle\mathsf{L}(\mathrm{v}),\mathrm{v}\rangle$ *is real for each* $\mathrm{v}\in\mathrm{dom}(\mathsf{L})$*;*

(ii) $\mathrm{spec}_0(\mathsf{L})\subseteq\mathbb{R}$*;*

(iii) $\mathrm{spec}_1(\mathsf{L})\subseteq\mathbb{R}$*;*

(iv) *if* λ_1 *and* λ_2 *are distinct eigenvalues for* L*, and if* v_i *is an eigenvector for* λ_i*,* $\mathrm{i}=1,2$*, then* $\langle\mathrm{v}_1,\mathrm{v}_2\rangle=0$*.*

Proof (i) We have

$$\langle\mathsf{L}(v),v\rangle = \langle v,\mathsf{L}(v)\rangle = \overline{\langle\mathsf{L}(v),v\rangle},$$

using the fact that L is symmetric.

(ii) Suppose that $\lambda \in \mathrm{spec}_0(\mathsf{L})$ and that $\lambda \neq 0$, otherwise the result is trivial. Let v be an eigenvector for λ and note that

$$\langle \mathsf{L}(v), v\rangle = \langle \lambda v, v\rangle = \lambda\langle v, v\rangle.$$

We also have, using the properties of the inner product,

$$\begin{aligned}\langle \mathsf{L}(v), v\rangle &= \langle v, \mathsf{L}(v)\rangle \\ &= \overline{\langle \mathsf{L}(v), v\rangle} \\ &= \overline{\lambda\langle v, v\rangle} \\ &= \bar{\lambda}\langle v, v\rangle,\end{aligned}$$

since $\langle v, v\rangle$ is real. This shows that

$$\lambda\langle v, v\rangle = \bar{\lambda}\langle v, v\rangle,$$

giving $\bar{\lambda} = \lambda$ as $\langle v, v\rangle \neq 0$.

(iii) This is the most difficult part of the theorem, and to prove it we use two technical lemmas.

1 **Lemma** *If* $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ *is an invertible linear operator on a normed vector space* $(\mathsf{V}, \|\cdot\|)$ *for which* $(\mathsf{L}^{-1}, \mathrm{image}(\mathsf{L}))$ *is unbounded, then there exists a sequence* $(\mathrm{v_j})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *with the following properties:*

(i) $\|\mathrm{v_j}\| = 1$, $\mathrm{j} \in \mathbb{Z}_{>0}$;

(ii) $\|\mathsf{L}(\mathrm{v_j})\| < \frac{1}{\mathrm{j}}$, $\mathrm{j} \in \mathbb{Z}_{>0}$.

Proof Let $\mathbb{S} = \{v \in \mathrm{image}(\mathsf{L}) \mid \|v\| = 1\}$. We claim that $\mathsf{L}^{-1}(\mathbb{S}) \subseteq \mathsf{V}$ is unbounded. To see this, suppose that $\mathsf{L}^{-1}(\mathbb{S})$ is bounded. Denote by $\overline{\mathsf{B}}(r, 0) = \{v \in \mathsf{V} \mid \|v\| \leq r\}$ the closed ball of radius r centred at $0 \in \mathsf{V}$. Since $\mathsf{L}^{-1}(\mathbb{S})$ is bounded there exists $M > 0$ with the property that $\mathsf{L}^{-1}(\overline{\mathsf{B}}(1, 0)) \subseteq \overline{\mathsf{B}}(M, 0)$. Now let $\epsilon > 0$. Choosing $\delta = \frac{\epsilon}{M}$ we see that $\mathsf{L}^{-1}(\overline{\mathsf{B}}(\delta, 0)) \subseteq \overline{\mathsf{B}}(\epsilon, 0)$ by linearity of L^{-1}. This shows that if $\mathsf{L}^{-1}(\mathbb{S})$ is bounded then L^{-1} is bounded.

Now, since $\mathsf{L}^{-1}(\mathbb{S})$ is unbounded there exists a sequence $(u_k)_{k\in\mathbb{Z}_{>0}}$ in $\mathbb{S}$ so that $\lim_{k\to\infty}\|\mathsf{L}^{-1}(u_k)\| = \infty$. Since $u_j \in \mathrm{image}(\mathsf{L})$ there exists a sequence $(\tilde{v}_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathrm{dom}(\mathsf{L})$ so that $\mathsf{L}(\tilde{v}_j) = u_j$, $j \in \mathbb{Z}_{>0}$. Therefore $\mathsf{L}^{-1} \circ \mathsf{L}(\tilde{v}_j) = \tilde{v}_j = \mathsf{L}^{-1}(u_j)$. Thus $\lim_{j\to\infty}\|\tilde{v}_j\| = \infty$. Defining $\left(v_j = \frac{\tilde{v}_j}{\|v_j\|}\right)_{j\in\mathbb{Z}_{>0}}$ we see that $\lim_{j\to\infty}\|\mathsf{L}(v_j)\| = \lim_{j\to\infty}\frac{\|u_j\|}{\|\tilde{v}_j\|} = 0$. Thus there exists a subsequence $(v_{j_k})_{k\in\mathbb{Z}_{>0}}$ of $(v_j)_{j\in\mathbb{Z}_{>0}}$ having the property as asserted in the lemma. ▼

2 **Lemma** *If* $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ *is a symmetric linear operator on an inner product space* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *and if* $\lambda = \xi + \mathrm{i}\eta \in \mathbb{C}$ *then* $\|(\mathsf{L} - \lambda\,\mathrm{id}_{\mathsf{V}})(\mathrm{v})\|^2 \geq \eta^2\|\mathrm{v}\|^2$ *for each* $\mathrm{v} \in \mathrm{dom}(\mathsf{L})$.

Proof We compute

$$\begin{aligned}\|(\mathsf{L}-\lambda\,\mathrm{id}_{\mathsf{V}})(v)\|^2 &= \langle(\mathsf{L}-\lambda\,\mathrm{id}_{\mathsf{V}})(v),(\mathsf{L}-\lambda\,\mathrm{id}_{\mathsf{V}})(v)\rangle\\ &= \|\mathsf{L}(v)\|^2-\langle\mathsf{L}(v),\lambda v\rangle-\langle\lambda v,\mathsf{L}(v)\rangle+\|\lambda v\|^2\\ &= \|\mathsf{L}(v)\|^2-\overline{\lambda}\langle\mathsf{L}(v),v\rangle-\lambda\langle\mathsf{L}(v),v\rangle+(\xi^2+\eta^2)\|v\|^2\\ &= \|\mathsf{L}(v)\|^2-2\xi\langle\mathsf{L}(v),v\rangle+\xi^2\|v\|^2+\eta^2\|v\|^2\\ &= \|(\mathsf{L}(v)-\xi\,\mathrm{id}_{\mathsf{V}})(v)\|^2+\eta^2\|v\|^2\\ &\geq \eta^2\|v\|^2,\end{aligned}$$

as desired. ▼

We now proceed with the proof by showing that if $\mathrm{Im}(\lambda)\neq 0$ then $\mathsf{L}-\lambda\,\mathrm{id}_{\mathsf{V}}$ is bounded. Let us write $\lambda=\xi+i\eta$. For any sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ with the property that $\|v_j\|=1$, $j\in\mathbb{Z}_{>0}$, by Lemma 2 we have $\|(\mathsf{L}-\lambda\,\mathrm{id}_{\mathsf{V}})(v_j)\|\geq|\eta|$ for $j\in\mathbb{Z}_{>0}$. Thus there exists $N\in\mathbb{Z}_{>0}$ so that $\|(\mathsf{L}-\lambda\,\mathrm{id}_{\mathsf{V}})(v_j)\|>\frac{1}{j}$ provided that $j\geq N$. By Lemma 1 this means that $(\mathsf{L}-\lambda\,\mathrm{id}_{\mathsf{V}})^{-1}$ must be bounded if $\mathrm{Im}(\lambda)\neq 0$, meaning that no such λ can lie in the continuous spectrum of L.

(iv) Let $\lambda_1,\lambda_2\in\mathbb{R}$ and $v_1,v_2\in\mathrm{dom}(\mathsf{L})$ be as specified. Then we compute

$$\begin{aligned}(\lambda_1-\lambda_2)\langle v_1,v_2\rangle &= \langle\lambda_1 v_1,v_2\rangle-\langle v_1,\lambda_2 v_2\rangle\\ &= \langle\mathsf{L}(v_1),v_2\rangle-\langle v_1,\mathsf{L}(v_2)\rangle\\ &= 0,\end{aligned}$$

using properties of the inner product and self-adjointness of L. Since $\lambda_1\neq\lambda_2$, it follows that v_1 and v_2 are orthogonal as stated. ■

With part (i) of the theorem at hand, the following definition makes sense.

4.2.12 Definition Suppose that $(\mathsf{L},\mathrm{dom}(\mathsf{L}))$ is a symmetric linear operator on $(\mathsf{V},\langle\cdot,\cdot\rangle)$.

(i) $(\mathsf{L},\mathrm{dom}(\mathsf{L}))$ is ***positive-definite*** if $\langle\mathsf{L}(v),v\rangle\geq 0$ for each $v\in\mathsf{V}$ and $\langle\mathsf{L}(v),v\rangle=0$ only if $v=0$.

(ii) $(\mathsf{L},\mathrm{dom}(\mathsf{L}))$ is ***negative-definite*** if $(-\mathsf{L},\mathrm{dom}(\mathsf{L}))$ is positive-definite. •

Now let us consider a further refinement that can be made for linear operators that are not only symmetric, but self-adjoint.

4.2.13 Theorem *If* $(\mathsf{L},\mathrm{dom}(\mathsf{L}))$ *is a self-adjoint linear operator on a Hilbert space* $(\mathsf{V},\langle\cdot,\cdot\rangle)$ *then* $\mathrm{spec}(\mathsf{L})\subseteq\mathbb{R}$ *and* $\mathrm{spec}_{-1}(\mathsf{L})=\varnothing$.

Proof Since a self-adjoint linear operator is symmetric, from Theorem 4.2.11 we need only show that $\mathrm{spec}_{-1}(\mathsf{L})=\varnothing$. We begin with a lemma that is of interest in its own right.

1 Lemma *Let* $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ *be a linear operator, not necessarily self-adjoint, on a Hilbert space* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *with* $\mathrm{dom}(\mathsf{L})$ *dense in* V*, and let* $\lambda \in \mathrm{spec}_{-1}(\mathsf{L})$ *have deficiency* m*. Then* $\bar{\lambda}$ *is an eigenvalue of* L^* *with multiplicity* m*.*

Proof Note that $\dim(\mathsf{V}/\mathrm{cl}(\mathrm{image}(\mathsf{L}_\lambda))) = \dim(\mathrm{image}(\mathsf{L}_\lambda)^\perp)$. Indeed, for those familiar with the notation involved with quotient spaces, the map sending $v + \mathrm{cl}(\mathrm{image}(\mathsf{L}_\lambda)) \in \mathsf{V}/\mathrm{cl}(\mathrm{image}(\mathsf{L}_\lambda))$ to the orthogonal projection of v onto $\mathrm{image}(\mathsf{L}_\lambda)^\perp$ is an isomorphism of $\mathsf{V}/\mathrm{cl}(\mathrm{image}(\mathsf{L}_\lambda))$ with $\mathrm{image}(\mathsf{L}_\lambda)^\perp$. For $v \in \mathrm{dom}(\mathsf{L})$ and $u \in \mathrm{image}(\mathsf{L}_\lambda)^\perp$ we have $\langle(\mathsf{L} - \lambda\,\mathrm{id}_\mathsf{V})(v), u\rangle = 0 = \langle v, 0\rangle$. This shows that $0 \in \mathrm{dom}((\mathsf{L} - \lambda\,\mathrm{id}_\mathsf{V})^*)$ and that $(\mathsf{L} - \lambda\,\mathrm{id}_\mathsf{V})^*(u) = 0$. Now note that $(\mathsf{L} - \lambda\,\mathrm{id}_\mathsf{V})^* = \mathsf{L}^* - \bar{\lambda}\,\mathrm{id}_\mathsf{V}$ (this is Exercise 4.2.2). This shows that $u \in \mathrm{image}(\mathsf{L}_\lambda)^\perp$ is an eigenvector for L^* with eigenvalue $\bar{\lambda}$, as desired. ▼

Now we proceed with the proof. If $\lambda \in \mathrm{spec}_{-1}(\mathsf{L})$ then, since $(\mathsf{L}, \mathrm{dom}(\mathsf{L}))$ is self-adjoint and by Lemma 1, we know that $\lambda \in \mathrm{spec}_0(\mathsf{L})$. Thus $\mathrm{spec}_{-1}(\mathsf{L}) \subseteq \mathbb{R}$. However, if $\lambda \in \mathbb{R}$ is in $\mathrm{spec}_{-1}(\mathsf{L})$ then Lemma 1 implies that λ is an eigenvalue of L^* and so an eigenvalue of L. However, points in $\mathrm{spec}_{-1}(\mathsf{L})$ cannot be eigenvalues, so the result follows. ■

The reader might recall that if V is a *finite-dimensional* inner product space, then there is always a basis of orthogonal eigenvectors for a self-adjoint linear transformation. The reader is led through a proof of this in Exercise 4.2.5. In infinite-dimensions, things are more subtle. Indeed, in infinite dimensions it is possible that there be no eigenvalues, that there be finitely many eigenvalues, or that there be infinitely many eigenvalues. The first two of these possibilities is exhibited in Exercises 4.2.7 and 4.2.8.

Orthogonal transformations and representations in finite-dimensions

4.2.5 Notes

The Riesz Representation Theorem is frequently attributed to Riesz [1907c] and Riesz [1909] and also to Fréchet [1907].

Exercises

4.2.1 Let L be a continuous linear transformation of an inner product space $(\mathsf{V}, \langle\cdot,\cdot\rangle)$. Show that the resulting map L^* is linear.

4.2.2 Let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a Hilbert space and let $\lambda \in \mathbb{F}$. What is the adjoint of the linear operator $(\mathsf{I}_\lambda, \mathsf{V})$ defined by $\mathsf{I}_\lambda(v) = \lambda v$?

4.2.3 On $(\mathsf{V} = \mathsf{L}^2([0,1];\mathbb{F}), \langle\cdot,\cdot\rangle)$ consider the linear operator $(\mathsf{L}, \mathrm{dom}(\mathsf{L}) = \mathsf{V})$ defined by

$$\mathsf{L}(f)(t) = \int_0^t f(\xi)\,\mathrm{d}\xi.$$

Show that $\mathrm{dom}(\mathsf{L}^*) = \mathsf{V}$ and that

$$\mathsf{L}^*(f)(t) = \int_t f(\xi)\,\mathrm{d}\xi.$$

4.2.4 If $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ is an inner product space and if $\mathsf{L} \in L_c(\mathsf{V};\mathsf{V})$ is a self-adjoint linear map, continuous with respect to the norm defined by the inner product, show that the operator norm, which we denote by $|||\cdot|||$, satisfies

$$|||\mathsf{L}||| = \sup_{\|v\|=1} \langle \mathsf{L}(v), v\rangle.$$

In the following exercise you will be led through an unconventional proof that a self-adjoint linear transformation on a finite-dimensional vector space possesses a basis of eigenvectors.

4.2.5 Let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a finite-dimensional $\mathbb{R}$-inner product space and let $\mathsf{L}\colon \mathsf{V} \to \mathsf{V}$ be a self-adjoint linear map. Define

$$\mathbb{S}_1 = \{v \in v \mid \ \|v\| = 1\}$$

to be the sphere of radius r in V in the norm $\|\cdot\|$ defined by the inner product $\langle\cdot,\cdot\rangle$.

(a) Show that $\mathbb{S}$ is closed and bounded.

Consider the function $\phi_1\colon \mathsf{V} \to \mathbb{R}$ defined by $\phi_1(v) = |\langle \mathsf{L}(v), v\rangle|$.

(b) Argue that the restriction of ϕ_1 to $\mathbb{S}_1$ attains its maximum on $\mathbb{S}_1$.

Now recall the Lagrange multiplier theorem.

Lagrange multiplier theorem *On the finite-dimensional $\mathbb{R}$-inner product space $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ let $\mathrm{f}, \mathrm{g}\colon \mathsf{V} \to \mathbb{R}$ be functions with g having the property that for every $\mathrm{v} \in \mathrm{g}^{-1}(0)$, $\mathrm{g}'(\mathrm{v}) \neq 0$. Then the derivative of the restriction of f to $\mathrm{g}^{-1}(0)$ vanishes at a point $\mathrm{v}_0 \in \mathrm{g}^{-1}(0)$ if and only if there exists $\lambda \in \mathbb{R}$ so that the derivative of the function*

$$\mathsf{V} \ni \mathrm{v} \mapsto \mathrm{f}(\mathrm{v}) + \lambda \mathrm{g}(\mathrm{v}) \in \mathbb{R}$$

vanishes at v_0.

Motivated by this, define $\psi_1\colon \mathsf{V} \to \mathbb{R}$ by $\psi_1(v) = \langle v, v\rangle - 1$ so that $\mathbb{S}_1 = \psi_1^{-1}(0)$.

(c) Show that $\psi_1'(v) \neq 0$ for all $v \in \psi_1^{-1}(0)$.
Hint: *To differentiate a function on V, use an orthonormal basis for V to write the function in terms of the components of a point $\mathrm{v} \in \mathsf{V}$, and then differentiate in the usual manner (you may have seen before the derivative of a function on a vector space as the "gradient" of the function).*

Now note that by (b), the restriction of the function ϕ_1 to $\mathbb{S}_1$ attains its maximum on $\mathbb{S}_1$. Thus, at the point $v_1 \in \mathbb{S}_1$ where the restriction of ϕ_1 attains its maximum, the derivative of the restriction must vanish.

(d) Show that the point $v_1 \in \mathbb{S}_1$ where the restriction of ϕ_1 attains its maximum is an eigenvector for L.
Hint: *Use the Lagrange multiplier theorem, this being valid by (c).*
Hint: *There are two cases to consider: (1) $\langle \mathsf{L}(\mathrm{v}_1), \mathrm{v}_1\rangle > 0$ and $\langle \mathsf{L}(\mathrm{v}_1), \mathrm{v}_1\rangle < 0$.*

Let v_1 be as in part (d) and consider the subspace $\mathsf{V}_2 = v_1^\perp$ which is the orthogonal complement to $\operatorname{span}(v_1)$. Define $\phi_2\colon \mathsf{V} \to \mathbb{R}$ by

$$\phi_2(v) = \phi_1(v - \langle v, v_1\rangle v_1),$$

let $\mathbb{S}_2 = \mathbb{S}_1 \cap \mathsf{V}_2$, and define $\psi_2\colon \mathsf{V}_2 \to \mathbb{R}$ by

$$\psi_2(v) = \langle v, v\rangle - 1.$$

(e) Show that there exists a linear map $\mathsf{L}_2\colon \mathsf{V}_2 \to \mathbb{R}$ so that $\phi_2(v) = |\langle \mathsf{L}_2(v), v\rangle|$ for $v \in \mathsf{V}_2$.

(f) Using part (e), argue that the above procedure can be emulated to show that the point at which ϕ_2 attains its maximum on $\mathbb{S}_2$ is an eigenvector v_2 for L_2.

(g) Show that $v_2 \in \mathsf{V}_2$ is an eigenvector for L, as well as being an eigenvector for L_2.

(h) Show that this process terminates after at most $n = \dim(\mathsf{V})$ applications of the above procedure.
Hint: *Determine what causes the process to terminate?*

(i) Show that the procedure produces a collection, $\{v_1, \dots, v_n\}$ of orthonormal eigenvectors for L. Be careful that you handle properly the case when the above process terminates *before* n steps.

4.2.6 Come to grips with Exercise 4.2.5 in the case when $\mathsf{V} = \mathbb{R}^2$, $\langle\cdot,\cdot\rangle$ is the "dot product," and for each of the following three self-adjoint linear maps.

(a) $\mathsf{L}(x, y) = (2x, y)$.

(b) $\mathsf{L}(x, y) = (-x, 2y)$.

(c) $\mathsf{L}(x, y) = (x, 0)$.

Thus you should in each case identify the maps ϕ_1 and ϕ_2, and show geometrically why maximising these functions picks off the eigenvectors as stated in Exercise 4.2.5.

4.2.7 Consider again the inner product space $(\mathsf{L}^2([0,1],\mathbb{R}), \langle\cdot,\cdot\rangle_2)$, and define a function $k\colon [0,1] \to \mathbb{R}$ by $k(t) = t$. Now define a linear transformation L_k by $(\mathsf{L}_k(f))(t) = k(t)f(t)$. Show that L_k is self-adjoint, but has no eigenvalues.

4.2.8 Consider the inner product space $(\mathsf{L}^2([0,1],\mathbb{R}), \langle\cdot,\cdot\rangle_2)$, and define a function $k\colon [0,1] \to \mathbb{R}$ by

$$k(t) = \begin{cases} 0, & t \in [0, \frac{1}{2}) \\ 1, & t \in [\frac{1}{2}, 1]. \end{cases}$$

Now define a linear transformation L_k as in Exercise 4.2.7. Show that the only eigenvalues for L_k are $\lambda_1 = 0$ and $\lambda_2 = 1$, and characterise all eigenvectors for each eigenvalue.

Section 4.3

Examples of Hilbert spaces

That $(\mathsf{L}^p(\mathbb{T};\mathbb{F}), \|\cdot\|_p)$ is a Hilbert space if and only if $p = 2$ can be shown using the parallelogram law, Theorem 4.1.9, and the reader is encouraged to do this (Exercise 4.3.1).

Exercises

4.3.1 Using the parallelogram law, show that $(\mathsf{L}^p(\mathbb{T};\mathbb{F}), \|\cdot\|_p)$ is a Hilbert space if and only if $p = 2$.

Section 4.4

Orthonormal bases in Hilbert spaces

One of the features distinguishing Hilbert spaces from their more general Banach space brethren is that Hilbert spaces always possess a Schauder basis. In the theory of Hilbert space these bases go by various names, including maximal orthonormal set or complete orthonormal families; we use the former convention. The idea that every vector in a Hilbert space can be written as a (possibly infinite) sum of distinguished basis vectors is an important one, and plays an important rôle in the theory of, for example, Fourier series; see Chapter IV-5. Our presentation in this section begins with the finite-dimensional case in order to build some important intuition. We then progress to enumerable then general bases.

Do I need to read this section? This chapter, at least that part dealing with countable maximal orthonormal sets, is important in our study of Fourier series in Chapter IV-5. Moreover, understanding the "geometry" of Hilbert spaces will be facilitated by understanding the notion of a maximal orthonormal set. •

4.4.1 General definitions and results

Before we proceed with our incremental treatment of orthonormal bases, let us give the definitions that apply to all inner product spaces.

4.4.1 Definition (Orthonormal set) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space.

(i) An ***orthogonal set*** is a collection $\{e_i\}_{i\in I}$ of nonzero vectors in V such that $\langle e_{i_1}, e_{i_2}\rangle = 0$ for all distinct $i_1, i_2 \in I$.

(ii) An ***orthonormal set*** is an orthogonal set $\{e_i\}_{i\in I}$ such that $\|e_i\| = 1$ for all $i \in I$. •

Sometimes we will talk about orthonormal and orthogonal families rather than sets. In this case we shall use the notation $(e_i)_{i\in I}$. The idea is the same, however.

Let us first indicate a useful construction for constructing orthonormal sets from linearly independent sets.

4.4.2 Theorem (Gram–Schmidt[3] orthonormalisation) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *be an* $\mathbb{F}$*-inner product space, and let* J *be either the set* $\{1, \dots, \mathrm{n}\}$ *for some* $\mathrm{n} \in \mathbb{Z}_{>0}$ *or the set*

[3]Jorgen Pedersen Gram (1850–1916) was a Danish mathematician whose principal employer was the Hafnia Insurance Company. Much of Gram's mathematical work was devoted to using mathematical and statistical methods in forestry management. Despite being somewhat outside the main circle of activity in mathematics, Gram made real contributions to algebra, number theory, probability theory, and numerical analysis. Erhard Schmidt (1876–1959) was born in what is now Estonia. His principal mathematical contributions were to the areas of integral equations and functional analysis.

$\mathbb{Z}_{>0}$. *For a family* $(v_j)_{j\in J}$ *of nonzero vectors in* V *define a family* $(u'_j)_{j\in J}$ *in* V *recursively by* $u'_1 = v_1$ *and*

$$u'_j = v_j - \sum_{k=1}^{j-1} \frac{\langle v_j, u'_k\rangle}{\|u'_k\|^2} u'_k, \qquad j \in J \setminus \{1\}.$$

If the family $(v_j)_{j\in J}$ *is linearly independent then the family* $(u'_j)_{j\in J}$ *is orthogonal. Moreover, if we additionally define* $u_j = \frac{u'_j}{\|u'_j\|}$, $j \in J$, *then* $(u_j)_{j\in J}$ *is orthonormal.*

Proof Let us prove that for any $m \in J$ the set $\{u'_1, \dots, u'_m\}$ is orthogonal. We prove this by induction on m. The claim is clearly true for $m = 1$. Suppose that the claim is true for $m = r$ so that $\{u'_1, \dots, u'_r\}$ is orthogonal. If $J = \{1, \dots, n\}$ and if $r = n$ then the claim is established. Otherwise we can carry on to show that $\{u'_1, \dots, u'_{r+1}\}$ is orthogonal as follows. For any $j \in \{1, \dots, r\}$,

$$\langle u'_{r+1}, u'_j\rangle = \Big\langle v_{r+1} - \sum_{k=1}^{r} \frac{\langle v_{r+1}, u'_k\rangle}{\|u'_k\|^2} u'_k, u'_j\Big\rangle = \langle v_{r+1}, u'_j\rangle - \langle v_{r+1}, u'_j\rangle = 0.$$

Thus u'_{r+1} is orthogonal to the set $\{u'_1, \dots, u'_r\}$. We claim that u'_{r+1} is nonzero. Indeed, by Exercise 4.4.1 we know that $\{u'_1, \dots, u'_r\}$ is linearly independent. Therefore,

$$\operatorname{span}_{\mathbb{F}}(v_1, \dots, v_r) = \operatorname{span}_{\mathbb{F}}(u'_1, \dots, u'_r).$$

Therefore, we have

$$u'_{r+1} = v_{r+1} + c_1 v_1 + \dots + c_r v_r$$

for $c_1, \dots, c_r \in \mathbb{F}$. If $u'_{r+1} = 0_V$ then linear independence of $\{v_1, \dots, v_{r+1}\}$ gives $c_1 = \dots = c_r = 0$ and $1 = 0$. This last assertion is absurd, and so we must have $u'_{r+1} \neq 0_V$. This shows that $\{u'_1, \dots, u'_{r+1}\}$ is indeed orthogonal.

Next we claim that orthogonality of $\{u'_1, \dots, u'_m\}$ for any $m \in J$ suffices to establish orthogonality of $(u'_j)_{j\in J}$. If J is finite this is obvious, so we consider the case where $J = \mathbb{Z}_{>0}$. In this case the family $(u'_j)_{j\in\mathbb{Z}_{>0}}$ could not be orthonormal in two ways.

1. One of the vectors u'_j, $j \in \mathbb{Z}_{>0}$, could be nonzero. This cannot happen, however, since for any $j \in \mathbb{Z}_{>0}$ the set $\{u_1, \dots, u_j\}$ is orthogonal.
2. For distinct $j_1, j_2 \in \mathbb{Z}_{>0}$ it could hold that $\langle u_{j_1}, u_{j_2}\rangle \neq 0$. This cannot happen, however, since for any distinct $j_1, j_2 \in \mathbb{Z}_{>0}$ the set $\{u_1, \dots, u_m\}$ is orthogonal for $m > \max\{j_1, j_2\}$.

The last assertion of the theorem is obvious. ■

4.4.3 Notation (Orthogonal sets) Generally we will use the notion of orthonormal set and not of an orthogonal set. However, in practice it is sometimes convenient to be able to talk about orthogonal sets as the objects which naturally present themselves are orthogonal, but not orthonormal. Note, however, that the two notions differ only in the trivial (but sometimes annoying) manner of nonzero constants. •

The following properties of orthonormal sets will be important to us in this section, and indeed in the study of inner product spaces in general.

4.4.4 Definition (Maximal, total, and basic orthonormal sets) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space, and let $\{e_i\}_{i\in I}$ be an orthogonal (resp. orthonormal) set.

(i) The orthogonal (resp. orthonormal) set $\{e_i\}_{i\in I}$ is ***maximal*** if, for any orthogonal (resp. orthonormal) set $\{f_j\}_{j\in J}$ such that $\{e_i\}_{i\in I} \subseteq \{f_j\}_{j\in J}$, $\{f_j\}_{j\in J} \subseteq \{e_i\}_{i\in I}$.

(ii) The orthogonal (resp. orthonormal) set $\{e_i\}_{i\in I}$ is ***total*** if $\mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\{e_i\}_{i\in I})) = \mathsf{V}$.

(iii) An orthogonal (resp. orthonormal) set $\{e_i\}_{i\in I}$ is ***basic*** if, for any $v \in \mathsf{V}$, there exist constants $c_i \in \mathbb{F}$, $i \in I$, for which the series

$$\sum_{i\in I} c_i e_i$$

converges to v in the sense of Definition 3.4.16. •

For convenience, let us recall here the definition of convergence used in the above definition for basic orthonormal sets. Convergence of the series

$$\sum_{i\in I} c_i e_i \tag{4.8}$$

to v means that, for every $\epsilon \in \mathbb{R}_{>0}$, there exists a finite set $J \subseteq I$ such that

$$\Big\|\sum_{j\in J} c_j e_j - v\Big\| < \epsilon.$$

By Proposition 3.4.18 it follows that a convergent sum of the form (4.8) is such that only countably many of the coefficients c_i, $i \in I$, are nonzero. Moreover, by Theorem 3.4.20, if the index set I is countable, say $I = \mathbb{Z}_{>0}$, then a sum

$$\sum_{j\in\mathbb{Z}_{>0}} c_j e_j$$

converges to v in the sense of Definition 3.4.16 if and only if it converges unconditionally to v. In particular, if this series converges to v in the sense of Definition 3.4.16 then it converges in the usual sense. It is usually the case that one deals with countable orthonormal sets.

Before we begin to explore properties of orthonormal sets of various flavours, let us give a few useful general results. First let us give the character of coefficients in any convergent series of orthonormal vectors.

4.4.5 Proposition (Coefficients in a convergent series of orthonormal vectors) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space, and let $\{e_i\}_{i\in I}$ be an orthonormal set. If the series*

$$\sum_{i\in I} c_i e_i$$

converges to $v \in \mathsf{V}$ then the coefficients must satisfy $c_i = \langle v, e_i\rangle$, $i \in I$.

Proof If I is finite then this is Exercise 4.4.4. Let us suppose, therefore, that I is infinite. Since the series converges, by Proposition 3.4.18 it follows that there exists an injection $\phi\colon \mathbb{Z}_{>0} \to I$ such that $c_i = 0$ for $i \notin \mathrm{image}(\phi)$ and such that

$$v = \sum_{j=1}^{\infty} c_{\phi(j)} e_{\phi(j)}.$$

Then, using Proposition 4.2.1 and Theorem 3.5.2, we deduce that for $j_0 \in \mathbb{Z}_{>0}$ we have

$$\langle v, e_{\phi(j_0)} \rangle = \Big\langle \sum_{j=1}^{\infty} c_j e_{\phi(j)}, e_{\phi(j_0)} \Big\rangle = \sum_{j=1}^{\infty} c_j \langle e_{\phi(j)}, e_{\phi(j_0)} \rangle = c_{\phi(j_0)},$$

giving $c_i = \langle v, e_i \rangle$ for $i \in \mathrm{image}(\phi)$. For $i \notin \mathrm{image}(\phi)$ a similar computation gives $\langle v, e_i \rangle = 0$; and so gives the result. ■

The following result is also useful.

4.4.6 Theorem (Bessel's[4] inequality) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *if* $(\mathsf{V}, \langle \cdot, \cdot \rangle)$ *is an* $\mathbb{F}$*-inner product space, if* $\{\mathrm{e}_i\}_{i \in I}$ *is an orthonormal set, and if* $\mathrm{v} \in \mathsf{V}$, *then*

$$\sum_{i \in I} |\langle \mathrm{v}, \mathrm{e}_i \rangle|^2 \le \|\mathrm{v}\|^2;$$

in particular, the sum on the left converges.

Proof By Exercise 4.4.5 we have

$$\sum_{j=1}^{n} |\langle v, e_{i_j} \rangle|^2 \le \|v\|^2$$

for every finite subset $\{i_1, \dots, i_n\} \subseteq I$. If I is finite this immediately gives the result. Let us consider the case where I is not finite. We claim that in this case $\langle v, e_i \rangle = 0$ for all but countably many $i \in I$. To see this, define

$$I_0 = \{i \in I \mid |\langle v, e_i \rangle| > 0\}$$

and suppose that I_0 is not countable. For $k \in \mathbb{Z}_{>0}$ define

$$I_k = \{i \in I \mid |\langle v, e_i \rangle|^2 \ge \tfrac{1}{k}\}.$$

Note that $I_0 = \cup_{k \in \mathbb{Z}_{>0}} I_k$, implying by Proposition I-1.7.16 that for at least one $k \in \mathbb{Z}_{>0}$ the set I_k must be infinite (uncountable, actually, although this is not necessary). Let $N \in \mathbb{Z}_{>0}$ be such that $N > k\|v\|^2$. Then, for any finite subset $\{i_1, \dots, i_N\} \subseteq I_k$ we have

$$\sum_{j=1}^{N} |\langle v, e_{i_j} \rangle|^2 \ge \sum_{j=1}^{N} \frac{1}{k} = \frac{N}{k} > \|v\|^2,$$

[4] Friedrich Wilhelm Bessel (1784–1846) was born in what is now Germany and made mathematical contributions to analysis. His primary scientific activities were directed towards astronomy.

which gives a contradiction. Thus I_0 must indeed be countable.

Thus we have an injection $\phi\colon \mathbb{Z}_{>0} \to I$ such that $\langle v, e_i\rangle = 0$ for $i \notin \operatorname{image}(\phi)$ and such that

$$\sum_{j=1}^{\infty} |\langle v, e_{\phi(j)}\rangle|^2 = \lim_{n\to\infty} \sum_{j=1}^{n} |\langle v, e_{\phi(j)}\rangle|^2 \leq \|v\|^2.$$

Thus we have

$$\sum_{i\in I} |\langle v, e_i\rangle|^2 \leq \|v\|^2$$

for every index set I. Since this is a sum of positive terms, the series

$$\sum_{i\in I} |\langle v, e_i\rangle|^2$$

converges for arbitrary index sets I. ■

Bessel's inequality makes the following definition reasonable.

4.4.7 Definition (Orthonormal expansion) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space, and let $\{e_i\}_{i\in I}$ be an orthonormal set. The ***orthonormal expansion*** of $v \in \mathsf{V}$ with respect to $\{e_i\}_{i\in I}$ is the series

$$\sum_{i\in I} \langle v, e_i\rangle e_i,$$

disregarding convergence. •

Let us give some examples of orthonormal sets.

4.4.8 Examples (Orthonormal sets)

1. In $\mathbb{F}^n$ with the standard inner product, one can check that the standard basis,

$$\boldsymbol{e}_1 = (1, 0, \dots, 0),\ \boldsymbol{e}_2 = (0, 1, \dots, 0),\ \dots,\ \boldsymbol{e}_n = (0, 0, \dots, 1),$$

is orthonormal. That is to say, the set $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is orthonormal. Moreover, the set $\{\lambda_1\boldsymbol{e}_1, \dots, \lambda_n\boldsymbol{e}_n\}$ is orthogonal for any collection of constants $\lambda_1, \dots, \lambda_n \in \mathbb{F}\setminus\{0\}$. Orthonormality of this set occurs precisely when $\lambda_j = 1$, $j \in \{1, \dots, n\}$.

It is easy to see that $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is a maximal orthonormal set. Indeed, let us consider an orthonormal set $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n, \boldsymbol{e}_{n+1}, \dots, \boldsymbol{e}_k\}$ containing $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$. We claim that $k = n$. Suppose otherwise. Since $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is a basis for $\mathbb{F}^n$ it follows that for each $a \in \{n+1, \dots, k\}$,

$$\boldsymbol{e}_a = c_{a1}\boldsymbol{e}_1 + \dots + c_{an}\boldsymbol{e}_n$$

for some constants $c_{a1}, \dots, c_{an}$. Since $\langle \boldsymbol{e}_a, \boldsymbol{e}_j\rangle = 0$ it follows that $c_{aj} = 0$ for $a \in \{n+1, \dots, k\}$ and $j \in \{1, \dots, n\}$. Thus $\boldsymbol{e}_{n+1} = \dots = \boldsymbol{e}_k = \boldsymbol{0}$, contradicting the orthonormality of $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_k\}$. Thus $k = n$.

Moreover, since $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ is a basis for $\mathbb{F}^n$ it follows that $\operatorname{span}_{\mathbb{F}}(\boldsymbol{e}_1, \dots, \boldsymbol{e}_n) = \mathbb{F}^n$, and so the orthonormal set is total and basic.

2. Next let us consider the inner product space $(\mathbb{F}_0^\infty, \langle\cdot,\cdot\rangle_2)$. We note that the standard basis $\{\boldsymbol{e}_j\}_{j\in\mathbb{Z}_{>0}}$, which we recall is defined by

$$\boldsymbol{e}_j(k) = \begin{cases} 1, & j = k, \\ 0, & j \neq k, \end{cases}$$

is orthonormal; this is straightforward to verify. Moreover, the set $(\lambda_j\boldsymbol{e}_j)_{j\in\mathbb{Z}_{>0}}$ is orthogonal for every collection of constants $\lambda_j \in \mathbb{F}\setminus\{0\}$, $j \in \mathbb{Z}_{>0}$, and is orthonormal if and only if $\lambda_j = 1$, $j \in \mathbb{Z}_{>0}$.

We leave it to the reader to show in Exercise 4.4.2 to show that $\{\boldsymbol{e}_j\}_{j\in\mathbb{Z}_{>0}}$ is a maximal orthonormal family.

Moreover, since $\{\boldsymbol{e}_j\}_{j\in\mathbb{Z}_{>0}}$ is a basis for $\mathbb{F}_0^\infty$ it follows that $\operatorname{span}_{\mathbb{F}}(\{\boldsymbol{e}_j\}_{j\in\mathbb{Z}_{>0}}) = \mathbb{F}_0^\infty$, and so the orthonormal set is total and basic.

3. The preceding examples might make one believe that the notions of maximal, total, and basic orthonormal sets are equivalent for general inner product spaces. They are not. Let us give an example to illustrate this. We consider the Hilbert space $(\ell^2(\mathbb{F}), \langle\cdot,\cdot\rangle_2)$ with $\{\boldsymbol{e}_j\}_{j\in\mathbb{Z}_{>0}}$ the orthonormal set from the preceding example, i.e., the standard (Hamel) basis for $\mathbb{F}_0^\infty \subseteq \ell^2(\mathbb{F})$. We then take the subspace

$$\mathsf{U} = \operatorname{span}_{\mathbb{F}}\Big(\sum_{j=1}^\infty \frac{\boldsymbol{e}_j}{j}, \boldsymbol{e}_2, \boldsymbol{e}_3, \dots\Big)$$

and consider $(\mathsf{U}, \langle\cdot,\cdot\rangle_2)$ as an inner product space. We claim that $\mathscr{B} = \{\boldsymbol{e}_2, \boldsymbol{e}_3, \dots\}$ is a maximal orthonormal set in U that is neither total nor basic.

To show that it is maximal, suppose that $\boldsymbol{u} \in \mathsf{U}$ is orthogonal to $\mathscr{B}$. Since $\boldsymbol{u} \in \mathsf{U}$ we can write

$$\boldsymbol{u} = c_1\Big(\sum_{j=1}^\infty \frac{\boldsymbol{e}_j}{j}\Big) + c_2\boldsymbol{e}_2 + \dots + c_k\boldsymbol{e}_k$$

for some $k \in \mathbb{Z}_{>0}$ and for $c_1, \dots, c_k \in \mathbb{F}$. Since

$$\Big\langle \boldsymbol{u}, \sum_{j=1}^\infty \frac{\boldsymbol{e}_j}{j}\Big\rangle = 0, \ \langle \boldsymbol{u}, \boldsymbol{e}_j\rangle = 0, \qquad j \in \{2, 3, \dots\},$$

it follows that $c_j = 0$, $j \in \{1, \dots, k\}$, and so $\boldsymbol{u} = \boldsymbol{0}_{\mathbb{F}_0^\infty}$. Thus there can be no orthonormal subset of U containing $\mathscr{B}$.

That $\mathscr{B}$ is not basic is plain since $\sum_{j=1}^\infty \frac{\boldsymbol{e}_j}{j}$ is in U but is not a sum of the form $\sum_{j=2}^\infty c_j\boldsymbol{e}_j$ (this follows from Proposition 4.4.5).

That $\mathscr{B}$ is not total follows since the subspace $\operatorname{span}_{\mathbb{F}}(\boldsymbol{e}_2, \boldsymbol{e}_3, \dots)$ is a closed subspace containing $\mathscr{B}$ but is a strict subspace of V. •

The preceding examples illustrate that the notions of maximal, total, and basic need not be equivalent for an orthonormal set. Let us explore the relationships between these concepts in a general setting.

4.4.9 Theorem (Relationship between maximal, total, and basic orthonormal sets) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space, and let $\mathscr{B} = \{\mathrm{e}_{\mathrm{i}}\}_{\mathrm{i}\in\mathrm{I}}$ be an orthonormal set. The following four statements are equivalent:*

(i) *$\mathscr{B}$ is basic;*

(ii) *$\mathscr{B}$ is total;*

(iii) *for every $\mathrm{v} \in \mathsf{V}$ the equality*

$$\|\mathrm{v}\|^2 = \sum_{\mathrm{i}\in\mathrm{I}} |\langle \mathrm{v}, \mathrm{e}_{\mathrm{i}}\rangle|^2$$

*holds, where convergence of the sum on the right is interpreted as in Definition I-2.4.31 (**Parseval's equality**);*

(iv) *for all $\mathrm{u}, \mathrm{v} \in \mathsf{V}$ we have*

$$\langle \mathrm{u}, \mathrm{v}\rangle = \sum_{\mathrm{i}\in\mathrm{I}} \langle \mathrm{u}, \mathrm{e}_{\mathrm{i}}\rangle \overline{\langle \mathrm{v}, \mathrm{e}_{\mathrm{i}}\rangle},$$

where convergence of the sum on the right is interpreted as in Definition I-2.4.31.

Also, the following two statements are equivalent:

(v) *$\mathscr{B}^{\perp} = \{0_{\mathsf{V}}\}$;*

(vi) *$\mathscr{B}$ is maximal.*

Finally, if V is a Hilbert space, the first four equivalent statements are equivalent to the last two equivalent statements.

Proof (i) $\Longrightarrow$ (ii) Let $\mathscr{B} = \{e_i\}_{i\in I}$ be basic and let $v \in \mathsf{V}$. We can then write

$$v = \sum_{i\in I} c_i e_i$$

for some coefficients $c_i \in \mathbb{F}$, $i \in I$. If I is finite this immediately implies that $v \in \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\mathscr{B}))$. If I is not finite, by Proposition 3.4.18 and Theorem 3.4.20 there exists an injection $\phi\colon \mathbb{Z}_{>0} \to I$ such that $c_i = 0$ for $i \notin \mathrm{image}(\phi)$ and such that

$$v = \sum_{j=1}^{\infty} c_j e_{\phi(j)}.$$

If we define

$$v_k = \sum_{j=1}^{k} c_j e_j$$

then the sequence $(v_k)_{k\in\mathbb{Z}_{>0}}$ converges to v. Thus $v \in \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\mathscr{B}))$ and so $\mathscr{B}$ is total.

(ii) $\Longrightarrow$ (iii) Let $v \in \mathsf{V}$. Since $\mathscr{B}$ is total there exists a sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathrm{span}_{\mathbb{F}}(\mathscr{B})$ such that $v = \lim_{j\to\infty} v_j$. For each $j \in \mathbb{Z}_{>0}$ write

$$v_j = c_{j1} e_{i_{j1}} + \cdots + c_{jk_j} e_{i_{jk_j}}$$

for $k_j \in \mathbb{Z}_{>0}$, coefficients $c_{j1}, \dots, c_{jk_j} \in \mathbb{F}$, and distinct $i_{j1}, \dots, i_{jk_j} \in I$. By Exercise 4.4.4 it follows that $c_{jl} = \langle v_j, e_{i_{jl}}\rangle$ for each $j \in \mathbb{Z}_{>0}$, $l \in \{1, \dots, k_j\}$. Note that the set

$\cup_{j\in\mathbb{Z}_{>0}}\{i_{j1},\dots,i_{jk_j}\}$ is countable by Proposition I-1.7.16. This means that there exists a countable set $K\subseteq I$ such that

$$v_j = \sum_{k\in K}\langle v_j, e_k\rangle e_k$$

for each $j\in\mathbb{Z}_{>0}$, with the sum being finite. We claim that $\langle v, e_i\rangle = 0$ for $i\notin K$. Indeed, for $i\in I$,

$$\langle v, e_i\rangle = \lim_{j\to\infty}\langle v_j, e_i\rangle = \lim_{j\to\infty}\sum_{k\in K}\langle v_j, e_k\rangle\langle e_k, e_i\rangle = 0,$$

using continuity of the inner product and Theorem 3.5.2.

We now have

$$\begin{aligned}\|v_j\|^2 &= \Big\langle \sum_{k\in K}\langle v_j, e_k\rangle e_k, \sum_{k'\in K}\langle v_j, e_{k'}\rangle e_{k'}\Big\rangle\\ &= \sum_{k\in K}\sum_{k'\in K}\langle v_j, e_k\rangle\overline{\langle v_j, e_{k'}\rangle}\langle e_k, e_{k'}\rangle\\ &= \sum_{k\in K}|\langle v_j, e_k\rangle|^2,\end{aligned}$$

using the fact that the inner product commutes with finite sums. Now, using continuity of the norm and inner product, along with Theorem 3.5.2, gives

$$\|v\|^2 = \lim_{j\to\infty}\|v_j\|^2 = \lim_{j\to\infty}\sum_{k\in K}|\langle v_j, e_k\rangle|^2 = \sum_{k\in K}|\langle v, e_k\rangle|^2 = \sum_{i\in I}|\langle v, e_i\rangle|^2,$$

as desired.

(iii) $\Longrightarrow$ (iv) For $u, v\in\mathsf{V}$ we have

$$\begin{aligned}&\|u+v\|^2 = \sum_{i\in I}|\langle u+v, e_i\rangle|^2\\ \Longrightarrow\quad &\|u\|^2+\|v\|^2+\langle u, v\rangle+\overline{\langle u, v\rangle}\\ &\quad= \sum_{i\in I}|\langle u, e_i\rangle|^2+\sum_{i\in I}|\langle v, e_i\rangle|^2+\sum_{i\in I}(\langle u, e_i\rangle\overline{\langle v, e_i\rangle}+\overline{\langle u, e_i\rangle\overline{\langle v, e_i\rangle}})\\ \Longrightarrow\quad &\mathrm{Re}(\langle u, v\rangle) = \sum_{i\in I}\mathrm{Re}(\langle u, e_i\rangle\overline{\langle v, e_i\rangle}).\end{aligned}$$

If $\mathbb{F}=\mathbb{R}$ this establishes the result. If $\mathbb{F}=\mathbb{C}$, a similar computation using the equality

$$\|u+\mathrm{i}v\|^2 = \sum_{i\in I}|\langle u+\mathrm{i}v, e_i\rangle|^2$$

gives

$$\mathrm{Im}(\langle u, v\rangle) = \sum_{i\in I}\mathrm{Im}(\langle u, e_i\rangle\overline{\langle v, e_i\rangle}).$$

(iv) $\Longrightarrow$ (i) Since part (iv) obviously implies part (iii), we shall prove that (iii) implies (i). Thus we have

$$\|v\|^2 = \sum_{i\in I}|\langle v, e_i\rangle|^2$$

for every $v \in \mathsf{V}$. By Proposition I-2.4.33, for $v \in \mathsf{V}$, it follows that there exists a bijection $\phi\colon \mathbb{Z}_{>0} \to I$ such that $\langle v, e_i \rangle = 0$ for $i \notin \text{image}(\phi)$ and such that

$$\|v\|^2 = \sum_{j=1}^{\infty} |\langle v, e_{\phi(j)} \rangle|^2.$$

For $k \in \mathbb{Z}_{>0}$ let us define

$$v_k = \sum_{j=1}^{k} \langle v, e_{\phi(j)} \rangle e_{\phi(j)}.$$

Note that

$$\begin{aligned}
\langle v - v_k, v_k \rangle &= \Big\langle v - \sum_{j=1}^{k} \langle v, e_{\phi(j)} \rangle e_{\phi(j)}, \sum_{l=1}^{k} \langle v, e_{\phi(l)} \rangle e_{\phi(l)} \Big\rangle \\
&= \Big\langle v, \sum_{l=1}^{k} \langle v, e_{\phi(l)} \rangle e_{\phi(l)} \Big\rangle - \Big\langle \sum_{j=1}^{k} \langle v, e_{\phi(j)} \rangle e_{\phi(j)}, \sum_{l=1}^{k} \langle v, e_{\phi(l)} \rangle e_{\phi(l)} \Big\rangle \\
&= \sum_{l=1}^{k} |\langle v, e_{\phi(l)} \rangle|^2 - \sum_{j=1}^{k} |\langle v, e_{\phi(j)} \rangle|^2 = 0
\end{aligned}$$

for every $k \in \mathbb{Z}_{>0}$. By the Pythagorean equality,

$$\|v\|^2 = \|v - v_k + v_k\|^2 = \|v - v_k\|^2 + \|v_k\|^2 \quad \Longrightarrow \quad \|v - v_k\|^2 = \|v\|^2 - \|v_k\|^2.$$

By assumption,

$$\lim_{k\to\infty} \|v_k\|^2 = \|v\|^2$$

and so

$$\lim_{k\to\infty} \|v - v_k\| = 0,$$

implying that

$$v = \sum_{i\in I} \langle v, e_i \rangle e_i,$$

and so in particular implying that $\mathscr{B}$ is basic.

(v) $\Longrightarrow$ (vi) Suppose that $\mathscr{B}$ is not maximal. Then there exists an orthonormal set $\mathscr{B}'$ such that $\mathscr{B} \subset \mathscr{B}'$. Let $v \in \mathscr{B}' \setminus \mathscr{B}$. Then, clearly, $v \in \mathscr{B}^\perp$ and $v \neq 0_{\mathsf{V}}$. Thus $\mathscr{B}^\perp \neq \{0_{\mathsf{V}}\}$.

(vi) $\Longrightarrow$ (v) Suppose that $\mathscr{B}^\perp \neq \{0_{\mathsf{V}}\}$ and let $v \in \mathscr{B}^\perp$ have unit length. Then the set $\mathscr{B} \cup \{v\}$ is an orthonormal set that strictly contains $\mathscr{B}$. Thus $\mathscr{B}$ is not maximal.

(ii) $\Longrightarrow$ (v) By Proposition 4.1.13(iv) we have $\mathscr{B}^\perp = \text{cl}(\text{span}_{\mathbb{F}}(\mathscr{B}))^\perp$. From this fact, if $\mathscr{B}$ is total it immediately follows that $\mathscr{B}^\perp = \{0_{\mathsf{V}}\}$.

(vi) $\Longrightarrow$ (i) (assuming V is a Hilbert space) Let $v \in \mathsf{V}$. Bessel's inequality gives

$$\sum_{i\in I} |\langle v, e_i \rangle|^2 \leq \|v\|^2,$$

and this implies that the series on the right converges and so is Cauchy. Let $\epsilon \in \mathbb{R}_{>0}$ and let $J \subseteq I$ be a finite set for which

$$\sum_{j \in J'} |\langle v, e_j \rangle|^2 < \epsilon$$

for every finite subset $J' \subseteq I$ such that $J \cap J' = \varnothing$ (see Definition 3.4.16). A direct computation using properties of inner products then gives

$$\Big\| \sum_{j \in J'} \langle v, e_j \rangle e_j \Big\|^2 = \sum_{j \in J'} |\langle v, e_j \rangle|^2 < \epsilon,$$

which shows that the series

$$\sum_{i \in I} \langle v, e_i \rangle e_i$$

is Cauchy. By Theorem 3.4.17 this series converges, implying that $\mathscr{B}$ is basic. ■

The following result records the fact that completeness is essential if all six of the statements in the preceding theorem are to be equivalent.

4.4.10 Theorem (Maximal orthonormal sets are not generally Hilbert bases for non-Hilbert spaces) *For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and for a $\mathbb{F}$-inner product space $(\mathsf{V}, \langle \cdot, \cdot \rangle)$, the following statements are equivalent:*

(i) V is a Hilbert space;

(ii) every maximal orthonormal set is a Hilbert basis.

Proof The implication of part (ii) from part (i) follows from Theorem 4.4.9, so we prove the converse implication. Let U be a proper closed subspace of V. By Theorem 4.1.23, to show that V is a Hilbert space it suffices to show that $\mathsf{U}^\perp \neq \{0_{\mathsf{V}}\}$. So suppose otherwise. Now let $\mathscr{B} = \{e_i\}_{i \in I}$ be a maximal orthonormal set in U and let $\mathscr{B}' = \mathscr{B} \cup \{f_j\}_{j \in J}$ be a maximal orthonormal set in V that extends that $\mathscr{B}$ (that such a set exists may be proved just as one proves Theorem I-4.5.26). Let $j_0 \in J$. Since $f_{j_0} \neq 0_{\mathsf{V}}$ it follows that $f_{j_0} \notin \mathsf{U}^\perp$. Thus there exists $u \in \mathsf{U}$ such that $\langle u, f_{j_0} \rangle \neq 0$. By hypothesis, $\mathscr{B}'$ is a basic orthonormal set and so we may write

$$u = \sum_{i \in I} a_i e_i + \sum_{j \in J} b_j f_j$$

for some coefficients $a_i \in \mathbb{F}$, $i \in I$, $b_j \in \mathbb{F}$, $j \in J$. Then

$$\sum_{j \in J} b_j f_j = u - \sum_{i \in I} a_i e_i \in \mathsf{U}.$$

We also have

$$\sum_{j \in J} b_j f_j \in \mathscr{B}^\perp.$$

Since $\mathscr{B}$ is a maximal orthonormal set in U it follows that

$$\sum_{j\in J} b_j f_j = 0_{\mathsf{V}}$$

and so

$$\langle u, f_{j_0}\rangle = \Big\langle \sum_{i\in I} a_i e_i, f_{j_0}\Big\rangle = \sum_{i\in I} a_i\langle e_i, f_{j_0}\rangle = 0,$$

where we have used Proposition 4.2.1. This is a contradiction. Thus it must be the case that $\mathsf{U}^{\perp} \neq \{0_{\mathsf{V}}\}$. ■

Now that we have a clear understanding of the relationships between basic, total, and maximal orthonormal sets, let us introduce some useful terminology.

4.4.11 Definition (Hilbert basis) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. A ***Hilbert basis*** for an $\mathbb{F}$-inner product space $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ is a basic (or, equivalently, total) orthonormal set in V. •

As we shall see, the notion of a Hilbert basis and a basis (sometimes also called a Hamel basis, cf. Remark I-4.5.21) can be different in a potentially confusing way. In particular, we refer to to clarify some aspects of the relationship between the two notions of basis. what

We have already seen in Example 4.4.8–3 that not every inner product space possesses a Hilbert basis. This, however, is where the value of the notion of a maximal orthonormal set arises.

4.4.12 Theorem (Every inner product spaces possesses a maximal orthonormal set) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ is a $\mathbb{F}$-inner product space, then there exists a maximal orthonormal set in V.*

Proof The proof goes very much like that for existence of a (Hamel) basis. Let $\mathscr{O}$ be the collection of orthonormal subsets of V. This set is nonempty since, if V is not the trivial vector space, $\{v\} \in \mathscr{O}$ for any vector v of unit length. Place a partial order $\preceq$ on $\mathscr{O}$ by asking that $S_1 \preceq S_2$ if $S_1 \subseteq S_2$. Let $\mathscr{S} \subseteq \mathscr{O}$ be a totally ordered subset. Note that $\cup_{S\in\mathscr{S}} S$ is an element of $\mathscr{O}$. Indeed, let $\{v_1, \dots, v_k\} \subseteq \cup_{S\in\mathscr{S}} S$. Then $v_j \in S_j$ for some $S_j \in \mathscr{S}$. Let $j_0 \in \{1, \dots, k\}$ be chosen such that S_{j_0} is the largest of the sets $S_1, \dots, S_k$ according to the partial order $\preceq$, this being possible since $\mathscr{S}$ is totally ordered. Then $\{v_1, \dots, v_k\} \subseteq S_{j_0}$ and so $\{v_1, \dots, v_k\}$ is orthonormal since S_{j_0} is orthonormal. It is also evident that $\cup_{S\in\mathscr{S}} S$ is an upper bound for $\mathscr{S}$. Thus every totally ordered subset of $\mathscr{O}$ possesses an upper bound, and so by Zorn's Lemma possesses a maximal element. Let $\mathscr{B}$ be such a maximal element. By construction $\mathscr{B}$ is orthonormal. We claim that it is also a maximal orthonormal set. Indeed, let $\mathscr{B}'$ be an orthonormal set such that $\mathscr{B} \subseteq \mathscr{B}'$. This immediately contradicts the fact that $\mathscr{B}$ is a maximal element of $\mathscr{O}$, and so we can conclude that $\mathscr{B}$ is a maximal orthonormal set. ■

For Hilbert spaces this leads to the following important result.

4.4.13 Corollary (Hilbert spaces possess a Hilbert basis) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and if* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *is an* $\mathbb{F}$*-Hilbert space, then there exists a Hilbert basis for* V*.*

Proof By Theorem 4.4.12 V possesses a maximal orthonormal set. By Theorem 4.4.9 every maximal orthonormal set is a Hilbert basis. ■

Note that it is not necessary for an inner product space to be a Hilbert space in order that it possess a Hilbert basis, cf. Example 4.4.8–2.

Now we consider a few important special cases of inner product spaces with orthonormal bases. While many of the result we give in the next two sections are actually special cases of the results above, we give independent proofs that are not dependent on the notion of a sum with an arbitrary index set. The relieves some of the complication present in the general setup.

4.4.2 Finite orthonormal sets and finite Hilbert bases

In this section we essentially generalise Example 4.4.8–1 to arbitrary finite-dimensional inner product spaces. The starting point is the following result. Note that we independently prove the existence of a Hilbert basis in this case, although this actually follows from Theorem 4.4.12.

4.4.14 Theorem (Characterisation of existence of finite Hilbert bases) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and if* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *is an inner product space of dimension* $\mathrm{n} \in \mathbb{Z}_{\geq 0}$*, then there exists a Hilbert basis for* V*. Moreover, every Hilbert basis for* V *is a basis and so has cardinality* n*.*

Proof If $\mathsf{V} = \{0_{\mathsf{V}}\}$ then there is nothing to prove, so let us suppose that $n \in \mathbb{Z}_{>0}$. By Theorem I-4.5.22 V possesses a basis and by Theorem I-4.5.25 the cardinality of any two bases are the same. Let $\{v_1, \dots, v_n\}$ be a basis for V and by Gram–Schmidt orthonormalisation construct an orthonormal set $\{u_1, \dots, u_n\}$. This set is linearly independent by Exercise 4.4.1 and so forms a basis for an n-dimensional subspace of V. By Proposition I-4.5.19 this subspace must be V. That is to say, $\{u_1, \dots, u_n\}$ is a basis for V. We claim that this implies that $\{u_1, \dots, u_n\}$ is a *Hilbert* basis. Since finite-dimensional inner product spaces are Hilbert spaces, it suffices to show that $\mathscr{B}$ is maximal. To prove maximality, suppose that $\{u_1, \dots, u_n, u_{n+1}, \dots, u_k\}$ is an orthonormal set containing $\{u_1, \dots, u_n\}$. By Exercise 4.4.1 it follows that $\{u_1, \dots, u_k\}$ is linearly independent. By Lemma I-1 from the proof of Theorem I-4.5.25 it follows that $k = n$, so proving maximality. This gives the existence of a Hilbert basis.

For the last assertion of the theorem, suppose that we have a Hilbert basis $\{u_1, \dots, u_m\}$ for V. Since $\{u_1, \dots, u_m\}$ is linearly independent by Exercise 4.4.1 it follows that $m \leq n$ by Lemma I-1 from the proof of Theorem I-4.5.25. To see that $m = n$ suppose otherwise so that $n > m$. Then $\operatorname{span}_{\mathbb{F}}(u_1, \dots, u_m)$ is a subspace of V of dimension $m < n$. By Theorem I-4.5.26 there exists $u_{m+1}, \dots, u_n \in \mathsf{V}$ such that $\{u_1, \dots, u_n\}$ is a basis for V. Applying the Gram–Schmidt orthonormalisation procedure gives a set $\{u'_1, \dots, u'_m, u'_{m+1}, \dots, u'_n\}$ where, by Exercise 4.4.3, $u'_j = u_j$ for $j \in \{1, \dots, m\}$. This contradicts the maximality of $\{u_1, \dots, u_m\}$ and so shows that we must have $m = n$. Thus every Hilbert basis is a linearly independent set of vectors having the same cardinality as the dimension of V, i.e., a basis. ■

A companion to the preceding result is the following more or less obvious fact.

4.4.15 Proposition (Necessary conditions for a finite Hilbert basis) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and if $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ is an $\mathbb{F}$-inner product space having a finite Hilbert basis, then V is finite-dimensional.*

Proof Let $\{e_1, \dots, e_n\}$ be a finite Hilbert basis for V. We claim that $\dim(\mathsf{V}) = n$. Suppose otherwise. Then, by Theorem I-4.5.26, there exists a basis $\mathscr{B}$ for V such that $\{e_1, \dots, e_n\} \subset \mathscr{B}$. Let $v \in \mathscr{B} \setminus \{e_1, \dots, e_n\}$. By applying Gram–Schmidt orthonormalisation procedure to $\{e_1, \dots, e_n, v\}$ we arrive at an orthonormal set $\{e_1, \dots, e_n, e_{n+1}\}$; by virtue of Exercise 4.4.3 the first n vectors remain unchanged. This, however, contradicts the maximality of $\{e_1, \dots, e_n\}$, and so we must have $\dim(\mathsf{V}) = n$. ■

Having established the existence of a Hilbert basis for a finite-dimensional inner product space, let us examine the set of all such bases. To motivate how one does this, recall from Section I-5.4.5 that there is a 1–1 correspondence between bases and invertible matrices. That is to say, if one chooses a basis $\mathscr{B} = \{e_1, \dots, e_n\}$ for V, then any other basis $\mathscr{B}' = \{e'_1, \dots, e'_n\}$ is uniquely determined by the invertible change of basis matrix $\boldsymbol{P}_{\mathscr{B}}^{\mathscr{B}'} \in \mathrm{Mat}_{n\times n}(\mathbb{F})$ which is defined by its satisfying the equality

$$e_{j_0} = \sum_{j=1}^{n} \boldsymbol{P}_{\mathscr{B}}^{\mathscr{B}'}(j, j_0) e'_j$$

for each $j_0 \in \{1, \dots, n\}$. We wish to understand the character of the change of basis matrix in the case where $\mathscr{B}$ and $\mathscr{B}'$ are both Hilbert bases.

The following result tells the story. In the statement, $\langle\cdot,\cdot\rangle_2$ denotes the standard inner product on $\mathbb{F}^n$ and $\|\cdot\|_2$ denotes the corresponding norm. Also, for a matrix $\boldsymbol{A}$ we denote by $\bar{\boldsymbol{A}}$ the matrix obtained by applying $\bar{\cdot}$ to the entries of $\boldsymbol{A}$.

4.4.16 Theorem (Change of basis matrices for finite Hilbert bases) *For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, for an n-dimensional $\mathbb{F}$-inner product space $(\mathsf{V}, \langle\cdot,\cdot\rangle)$, for a Hilbert basis $\mathscr{B} = \{e_1, \dots, e_n\}$ for V, and for $\mathbf{U} \in \mathrm{Mat}_{n\times n}(\mathbb{F})$ the following statements are equivalent:*

(i) *there exists a Hilbert basis $\mathscr{B}' = \{e'_1, \dots, e'_n\}$ for V such that $\mathbf{U} = \mathbf{P}_{\mathscr{B}}^{\mathscr{B}'}$;*

(ii) *$\|\mathbf{U}\mathbf{x}\|_2 = \|\mathbf{x}\|_2$ for all $\mathbf{x} \in \mathbb{F}^n$;*

(iii) *$\langle \mathbf{U}\mathbf{x}, \mathbf{U}\mathbf{y}\rangle_2 = \langle \mathbf{x}, \mathbf{y}\rangle_2$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{F}^n$;*

(iv) *$\mathbf{U}\bar{\mathbf{U}}^T = \bar{\mathbf{U}}^T\mathbf{U} = \mathbf{I}_n$;*

(v) *$\mathbf{U}$ is invertible and $\mathbf{U}^{-1} = \bar{\mathbf{U}}^T$.*

Proof (i) $\Longrightarrow$ (ii) By hypothesis we have

$$e_{j_0} = \sum_{j=1}^{n} \boldsymbol{U}(j, j_0) e'_j, \qquad j_0 \in \{1, \dots, n\},$$

so that, for every $j_1, j_2 \in \{1, \dots, n\}$,

$$\langle e_{j_1}, e_{j_2}\rangle = \Big\langle \sum_{k=1}^{n} \boldsymbol{U}(k, j_1) e'_k, \sum_{l=1}^{n} \boldsymbol{U}(l, j_2) e'_l \Big\rangle = \sum_{k=1}^{n} \boldsymbol{U}(k, j_1)\bar{\boldsymbol{U}}(k, j_2). \tag{4.9}$$

That is,

$$\sum_{k=1}^{n} \boldsymbol{U}(k, j_1)\bar{\boldsymbol{U}}(k, j_2) = \begin{cases} 1, & j_1 = j_2, \\ 0, & j_1 \neq j_2. \end{cases} \tag{4.10}$$

Now, for $\boldsymbol{x} \in \mathbb{F}^n$, a direct computation gives

$$\|\boldsymbol{U}\boldsymbol{x}\|_2^2 = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \boldsymbol{U}(i,j)\bar{\boldsymbol{U}}(i,k)\boldsymbol{x}(j)\boldsymbol{x}(k)$$

which gives $\|\boldsymbol{U}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2$ after using (4.10). This part of the result now follows by taking square roots.

(ii) $\implies$ (iii) We are assuming that $\|\boldsymbol{U}\boldsymbol{x}\|_2 = \|\boldsymbol{x}\|_2$ which implies that

$$\|\boldsymbol{U}\boldsymbol{x}\|_2^2 = \|\boldsymbol{x}\|_2^2 \quad\implies\quad \langle \boldsymbol{U}\boldsymbol{x}, \boldsymbol{U}\boldsymbol{x}\rangle_2 = \langle \boldsymbol{x}, \boldsymbol{x}\rangle_2,$$

this holding for all $\boldsymbol{x} \in \mathbb{F}^n$. Thus, for every $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{F}^n$,

$$\begin{aligned} & \langle \boldsymbol{U}(\boldsymbol{x}+\boldsymbol{y}), \boldsymbol{U}(\boldsymbol{x}+\boldsymbol{y})\rangle_2 = \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}\rangle_2 \\ \implies\quad & \langle \boldsymbol{U}\boldsymbol{x}, \boldsymbol{U}\boldsymbol{x}\rangle_2 + \langle \boldsymbol{U}\boldsymbol{y}, \boldsymbol{U}\boldsymbol{y}\rangle_2 + 2\,\mathrm{Re}(\langle \boldsymbol{U}\boldsymbol{x}, \boldsymbol{U}\boldsymbol{y}\rangle_2) = \langle \boldsymbol{x}, \boldsymbol{x}\rangle_2 + \langle \boldsymbol{y}, \boldsymbol{y}\rangle_2 + 2\,\mathrm{Re}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle_2) \\ \implies\quad & \mathrm{Re}(\langle \boldsymbol{U}\boldsymbol{x}, \boldsymbol{U}\boldsymbol{y}\rangle_2) = \mathrm{Re}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle_2). \end{aligned}$$

If $\mathbb{F} = \mathbb{R}$ then this gives this part of the result. If $\mathbb{F} = \mathbb{C}$, a computation entirely similar to the preceding one shows that

$$\langle \boldsymbol{U}(\boldsymbol{x}+\mathrm{i}\boldsymbol{y}), \boldsymbol{U}(\boldsymbol{x}+\mathrm{i}\boldsymbol{y})\rangle_2 = \langle \boldsymbol{x}+\mathrm{i}\boldsymbol{y}, \boldsymbol{x}+\mathrm{i}\boldsymbol{y}\rangle_2 \implies \mathrm{Im}(\langle \boldsymbol{U}\boldsymbol{x}, \boldsymbol{U}\boldsymbol{y}\rangle_2) = \mathrm{Im}(\langle \boldsymbol{x}, \boldsymbol{y}\rangle_2),$$

which gives this part of the result.

(iii) $\implies$ (iv) Letting $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ be the standard basis for $\mathbb{F}^n$ we have

$$\langle \boldsymbol{U}\boldsymbol{e}_j, \boldsymbol{U}\boldsymbol{e}_k\rangle_2 = \langle \boldsymbol{e}_j, \boldsymbol{e}_k\rangle_2, \qquad j,k \in \{1,\dots,n\}.$$

We have

$$\langle \boldsymbol{e}_j, \boldsymbol{e}_k\rangle_2 = \boldsymbol{I}_n(j,k) = \begin{cases} 1, & j = k, \\ 0, & j \neq k \end{cases}$$

and a direct calculation shows that

$$\langle \boldsymbol{U}\boldsymbol{e}_j, \boldsymbol{U}\boldsymbol{e}_k\rangle_2 = \sum_{i=1}^{n} \boldsymbol{U}(i,j)\bar{\boldsymbol{U}}(i,k) = (\boldsymbol{U}^T\bar{\boldsymbol{U}})(j,k).$$

Thus $\boldsymbol{U}^T\bar{\boldsymbol{U}} = \boldsymbol{I}_n$ which, upon conjugation, gives $\bar{\boldsymbol{U}}^T\boldsymbol{U} = \boldsymbol{I}_n$. From Theorem I-5.1.42 this means that $\boldsymbol{U}$ is invertible with inverse $\bar{\boldsymbol{U}}^T$. This means that we also have $\boldsymbol{U}\bar{\boldsymbol{U}}^T = \boldsymbol{I}_n$.

(iv) $\implies$ (v) This was proved in the preceding part of the proof.

(v) $\implies$ (i) By hypothesis we have

$$\bar{\boldsymbol{U}}^T\boldsymbol{U} = \boldsymbol{I}_n \quad\implies\quad \boldsymbol{U}^{-1}\bar{\boldsymbol{U}}^{-T} = \boldsymbol{I}_n.$$

By Theorem I-5.1.42 this implies that $\boldsymbol{U}^{-1}$ is invertible with inverse $\bar{\boldsymbol{U}}^{-T}$. Thus

$$\bar{\boldsymbol{U}}^{-T}\boldsymbol{U}^{-1} = \boldsymbol{I}_n \quad\implies\quad \boldsymbol{U}^{-T}\bar{\boldsymbol{U}}^{-1} = \boldsymbol{I}_n.$$

Let us define a basis $\{e'_1, \dots, e'_n\}$ for V by asking that

$$e'_{j_0} = \sum_{j=1}^{n} \boldsymbol{U}^{-1}(j, j_0) e_j. \tag{4.11}$$

The computation (4.9), but using $\boldsymbol{U}^{-1}$ in place of $\boldsymbol{U}$, gives

$$\langle e'_{j_1}, e'_{j_2} \rangle = \sum_{k=1}^{n} \boldsymbol{U}^{-1}(k, j_1)\bar{\boldsymbol{U}}^{-1}(k, j_2) = (\boldsymbol{U}^{-T}\bar{\boldsymbol{U}}^{-1})(j_1, j_2) = \boldsymbol{I}_n(j_1, j_2).$$

Thus

$$\langle e'_{j_1}, e'_{j_2} \rangle = \begin{cases} 1, & j_1 = j_2, \\ 0, & j_1 \neq j_2, \end{cases}$$

showing that $\{e'_1, \dots, e'_n\}$ is a Hilbert basis. Since (4.11) implies that

$$e_{j_0} = \sum_{j=1}^{n} \boldsymbol{U}(j, j_0) e'_j,$$

this part of the result follows. ■

In the case where $\mathbb{F} = \mathbb{R}$ the previous result, along with Theorem II-1.3.18, shows that the change of basis matrices between Hilbert bases are precisely the orthogonal matrices. The set of $n \times n$ orthogonal matrices were denoted by $\mathsf{O}(n)$. In the case where $\mathbb{F} = \mathbb{C}$ the matrices of the preceding result are called ***unitary*** matrices and the set of $n \times n$ unitary matrices are denoted by $\mathsf{U}(n)$.

One of the interesting features of Hilbert bases is that it is easy to determine the components of a vector relative to the basis. The following result records this.

4.4.17 Proposition (Components relative to a finite orthonormal set) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $(\mathsf{V}, \langle\cdot,\cdot\rangle_2)$ *be a (not necessarily finite-dimensional)* $\mathbb{F}$*-inner product space, and let* $\{\mathrm{e}_1, \dots, \mathrm{e}_n\}$ *be a finite orthonormal set. If* $\mathrm{v} \in \mathrm{span}_{\mathbb{F}}(\mathrm{e}_1, \dots, \mathrm{e}_n)$ *then*

$$\mathrm{v} = \langle \mathrm{v}, \mathrm{e}_1 \rangle \mathrm{e}_1 + \dots + \langle \mathrm{v}, \mathrm{e}_n \rangle \mathrm{e}_n.$$

Proof This is Exercise 4.4.4. ■

The preceding result has the following obvious corollary.

4.4.18 Corollary (Components relative to a finite Hilbert basis) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $(\mathsf{V}, \langle\cdot,\cdot\rangle_2)$ *be a finite-dimensional* $\mathbb{F}$*-inner product space, and let* $\{\mathrm{e}_1, \dots, \mathrm{e}_n\}$ *be a finite Hilbert basis for* V*. For* $\mathrm{v} \in \mathsf{V}$ *the components of* v *are* $\langle \mathrm{v}, \mathrm{e}_j \rangle$, $\mathrm{j} \in \{1, \dots, \mathrm{n}\}$.

We shall now give some properties of Hilbert bases for finite-dimensional inner product spaces that may, at first glance, seem obvious and/or silly. However, they arise in the infinite-dimensional setting in a rather less obvious and hopefully less silly way. Therefore, it is worth recording them in the present setup.

The first result is the finite-dimensional version of Bessel's inequality.

4.4.19 Proposition (Bessel's inequality for finite orthonormal sets) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*,* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *is a (not necessarily finite-dimensional)* $\mathbb{F}$*-inner product space, and if* $\{e_1, \dots, e_n\}$ *is a finite orthonormal set, then, for any* $v \in \mathsf{V}$*,*

$$\sum_{j=1}^{n} |\langle v, e_j\rangle|^2 \le \|v\|^2.$$

Proof This is Exercise 4.4.5. ■

Our final result gives several conditions equivalent to that of being a Hilbert basis. These are more or less "obvious" in finite-dimensions, but are a little less so in infinite-dimensions.

4.4.20 Theorem (Characterisations of finite Hilbert bases) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *be a finite-dimensional* $\mathbb{F}$*-inner product space, and let* $\mathscr{B} = \{e_1, \dots, e_n\}$ *be an orthonormal set. The following statements are equivalent:*

(i) $\mathscr{B}$ *is basic;*

(ii) $\mathscr{B}$ *is total;*

(iii) *for all* $v \in \mathsf{V}$ *we have*

$$\|v\|^2 = \sum_{j=1}^{n} |\langle v, e_j\rangle|^2$$

*(**Parseval's equality**);*

(iv) *for all* $u, v \in \mathsf{V}$ *we have*

$$\langle u, v\rangle = \sum_{j=1}^{n} \langle u, e_j\rangle \overline{\langle v, e_j\rangle};$$

(v) $\mathscr{B}^\perp = \{0_{\mathsf{V}}\}$*;*

(vi) $\mathscr{B}$ *is a maximal.*

Proof We leave this to the reader as Exercise 4.4.6. ■

4.4.3 Enumerable orthonormal sets and enumerable Hilbert bases

In the finite-dimensional case we see that Hilbert bases are always bases in the usual sense. Thus a Hilbert basis for a finite-dimensional inner product space is simply an instance of something we are already familiar with. This is no longer true in infinite-dimensions. Complications can arise in multiple ways. From Theorem 4.4.12 we know that every inner product space possesses a maximal orthonormal subset. For Hilbert spaces, these maximal orthonormal sets are necessarily Hilbert bases by Corollary 4.4.13. However, in infinite-dimensions it is not necessarily the case that a Hilbert basis is a basis. It *can* be the case that a Hilbert basis is a basis (see Example 4.4.8–2), but it is also true that enumerable Hilbert bases for Hilbert spaces are *never* bases. Also, for non-Hilbert spaces it can happen that they do not possess a Hilbert basis (see Example 4.4.8–3).

What we do in this section is consider the special case of inner product spaces that admit an enumerable Hilbert basis. Thus we consider the case where we have an enumerable orthonormal set $(e_j)_{j\in\mathbb{Z}_{>0}}$ for an inner product space and we assume that for any $v \in \mathsf{V}$ we can write

$$v = \sum_{j=1}^{\infty} c_j e_j. \tag{4.12}$$

Note that this sum is infinite, not finite as for a Hamel basis. The definition of convergence we use for this sum is made exactly as with the discussion of series in Banach spaces in Definition 3.4.1. That is to say, the existence of the infinite sum in (4.12) means that, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that

$$\left\| v - \sum_{j=1}^{k} c_j e_j \right\| < \epsilon$$

for every $k \geq N$. Note that it is not obvious that this coincides with the notion of convergence used in our general discussion in Section 4.4.1. Indeed, convergence for series using general index sets as used in Section 4.4.1 is equivalent to unconditional convergence for series using the index set $\mathbb{Z}_{>0}$. This sort of convergence *implies* convergence in the usual sense, but is not equivalent to it. This notwithstanding, we shall see that the usual definition of convergence for series is the appropriate one to use in the setting of enumerable Hilbert bases.

First we establish the appropriate condition under which an inner product space admits a enumerable Hilbert basis. In Theorem 4.4.14 we saw that the appropriate condition for the existence of a finite Hilbert basis was that the inner product space be, not surprisingly, finite-dimensional. For enumerable Hilbert bases, the condition turns out to be that the inner product space be separable (see Definition 3.6.12).

4.4.21 Theorem (Characterisation of existence of enumerable Hilbert bases) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and if* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *is a separable, infinite-dimensional* $\mathbb{F}$*-inner product space, then the following statements hold:*

(i) if V *is a Hilbert space then it possesses an enumerable Hilbert basis;*

(ii) every Hilbert basis for V *is enumerable.*

Proof By Corollary 4.4.13 we know that if V is a Hilbert space then it possesses a Hilbert basis and by Proposition 4.4.15 we know that every Hilbert basis is infinite. It remains to show that every Hilbert basis is enumerable. Suppose otherwise and so there exists an uncountable Hilbert basis $\mathscr{B} = \{e_i\}_{i\in I}$. If $i_1, i_2 \in I$ then

$$\|e_{i_1} - e_{i_2}\| = (\langle e_{i_1} - e_{i_2}, e_{i_1} - e_{i_2}\rangle)^{1/2} = (\|e_{i_1}\|^2 + \|e_{i_2}\|^2)^{1/2} = \sqrt{2}. \tag{4.13}$$

since e_{i_1} and e_{i_2} are orthogonal. For each $i \in I$ define $U_i = \mathsf{B}(\frac{1}{4}, e_i)$ and note that $U_{i_1} \cap U_{i_2} = \emptyset$ by (4.13). Now let $S \subseteq \mathsf{V}$ be enumerable. Then there exists an uncountable set $J \subseteq I$ such that $S \cap (\cup_{j\in J} U_j) = \emptyset$. Thus $S \subseteq \mathsf{V} \setminus (\cup_{j\in J} U_j)$ and so $\mathrm{cl}(S) \subseteq \mathsf{V} \setminus (\cup_{j\in J} U_j)$. Thus $\mathrm{cl}(S) \neq \mathsf{V}$ and so V is not separable. ■

The companion result to this is that enumerable Hilbert bases exist *only* for separable inner product spaces.

4.4.22 Theorem (Necessary conditions for a enumerable Hilbert basis) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and if* $(\mathsf{V}, \langle \cdot, \cdot \rangle)$ *is an* $\mathbb{F}$*-inner product space having an enumerable Hilbert basis, then* V *is separable and infinite-dimensional.*

Proof That V is infinite-dimensional follows from Proposition 4.4.15. To show that V is separable we let $\mathscr{B} = \{e_j\}_{j \in \mathbb{Z}_{>0}}$ be an enumerable Hilbert basis and let $\mathsf{V}_0 = \mathrm{span}_{\mathbb{F}}(\mathscr{B})$. By Theorem 4.4.9 we know that $\mathscr{B}$ is total and so $\mathrm{cl}(\mathsf{V}_0) = \mathsf{V}$. Now define

$$\mathbb{F}_{\mathbb{Q}} = \begin{cases} \mathbb{Q}, & \mathbb{F} = \mathbb{R}, \\ q_r + \mathrm{i} q_i, & \mathbb{F} = \mathbb{C} \end{cases}$$

and consider the set

$$S_{\mathscr{B}} = \{q_1 e_{j_1} + \cdots + q_k e_{j_k} \mid k \in \mathbb{Z}_{>0},\ q_1, \dots, q_k \in \mathbb{F}_{\mathbb{Q}}\}$$

of finite linear combinations of elements from $\mathscr{B}$ with coefficients in $\mathbb{F}_{\mathbb{Q}}$. Using Proposition I-1.7.16 we may conclude that $S_{\mathscr{B}}$ is enumerable. We claim that $S_{\mathscr{B}}$ is dense in V. From Exercise 3.6.2 it suffices to show that $S_{\mathscr{B}}$ is dense in V_0. Let $v \in \mathsf{V}_0$ so that we may write

$$v = c_1 e_{j_1} + \cdots + c_k e_{j_k}$$

for some $j_1, \dots, j_k \in \mathbb{Z}_{>0}$ and $c_1, \dots, c_k \in \mathbb{F}$. Let $\epsilon \in \mathbb{R}_{>0}$ and choose $q_a \in \mathbb{F}_{\mathbb{Q}}$ such that $|c_a - q_a| < \frac{\epsilon}{k}$, $a \in \{1, \dots, k\}$. Then

$$\|v - q_1 e_{j_1} - \cdots - q_k e_{j_k}\| \leq |c_1 - q_1| \|e_{j_1}\| + \cdots + |c_k - q_k| \|e_{j_k}\| < \epsilon$$

by the triangle inequality. Thus $v \in \mathrm{cl}(S_{\mathscr{B}})$ by Proposition 3.6.8. We have thus shown that the enumerable set $S_{\mathscr{B}}$ is dense in V, as desired. ■

First let us determine the form of the coefficients in the summation (4.12) if it does indeed converge. The reader should compare this result to Proposition 4.4.17.

4.4.23 Proposition (Components relative to an enumerable orthonormal set) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *let* $(\mathsf{V}, \langle \cdot, \cdot \rangle)$ *be an* $\mathbb{F}$*-inner product space, and let* $\{\mathrm{e}_j\}_{j \in \mathbb{Z}_{>0}}$ *be an orthonormal set in* V*. If the sum*

$$\sum_{j=1}^{\infty} \mathrm{c}_j \mathrm{e}_j$$

converges to $\mathrm{v} \in \mathsf{V}$, *then for each* $\mathrm{j} \in \mathbb{Z}_{>0}$, $\mathrm{c}_j = \langle \mathrm{v}, \mathrm{e}_j \rangle$.

Proof By Proposition 4.2.1 and Theorem 3.5.2 we have

$$\langle v, e_k \rangle = \Big\langle \sum_{j=1}^{\infty} c_j e_j, e_k \Big\rangle = \Big\langle \lim_{n \to \infty} \sum_{j=1}^{n} c_j e_j, e_k \Big\rangle = \lim_{n \to \infty} \sum_{j=1}^{n} c_j \langle e_j, e_k \rangle = c_k$$

for every $k \in \mathbb{Z}_{>0}$. ■

The reader should be sure to appreciate that, while the formula for the coefficients is exactly as given in the finite-dimensional case in Proposition 4.4.17, one must be a little more careful in arriving at this formula as there are issues with swapping limits with the inner product that must be accounted for.

The following result holds even for orthonormal sets that are not basic and should be compared to Proposition 4.4.19.

4.4.24 Theorem (Bessel's inequality for enumerable orthonormal sets) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* $(\mathsf{V}, \langle \cdot, \cdot \rangle)$ *be a* $\mathbb{F}$*-inner product space, and let* $\mathscr{B} = \{\mathrm{e}_j\}_{j \in \mathbb{Z}_{>0}}$ *be an enumerable orthonormal set. Then, for any* $\mathrm{v} \in \mathsf{V}$*, the sum*

$$\sum_{j=1}^{\infty} |\langle \mathrm{v}, \mathrm{e}_j \rangle|^2 \tag{4.14}$$

converges and satisfies

$$\sum_{j=1}^{\infty} |\langle \mathrm{v}, \mathrm{e}_j \rangle|^2 \leq \|\mathrm{v}\|^2.$$

Proof Let v_k denote the kth partial sum:

$$v_k = \sum_{j=1}^{k} \langle v, e_j \rangle e_j.$$

We claim that for $j \in \{1, \dots, k\}$, e_j is orthogonal to $v - v_k$. Indeed,

$$\langle v - v_k, e_j \rangle = \langle v, e_j \rangle - \langle v_k, e_j \rangle.$$

We also have, by a direct computation, $\langle v_k, e_j \rangle$ as the jth term in the sum, i.e., $\langle v_k, e_j \rangle = \langle v, e_j \rangle$. Thus $\langle v - v_k, e_j \rangle = 0$ as claimed. From this, since v_k is a linear combination of $\{e_1, \dots, e_k\}$, it follows that $v - v_k$ and v_k are orthogonal. By the Pythagorean identity (Exercise 4.1.12) we then have

$$\|v\|^2 = \|v - v_k + v_k\|^2 = \|v - v_k\|^2 + \|v_k\|^2,$$

giving

$$\|v_k\|^2 \leq \|v\|^2. \tag{4.15}$$

Since the vectors $\{e_1, \dots, e_k\}$ are orthonormal we compute

$$\|v_k\|^2 = \Big\langle \sum_{j=1}^{k} \langle v, e_j \rangle e_j, \sum_{l=1}^{k} \langle v, e_l \rangle e_l \Big\rangle = \sum_{j=1}^{k} \sum_{l=1}^{k} \langle v, e_j \rangle \overline{\langle v, e_l \rangle} \langle e_j, e_l \rangle = \sum_{j=1}^{k} |\langle v, e_j \rangle|^2. \tag{4.16}$$

Thus, combining (4.15) and (4.16), we have shown that the inequality

$$\sum_{j=1}^{k} |\langle v, e_j \rangle|^2 \leq \|v\|^2$$

holds for any $k \in \mathbb{Z}_{>0}$. Thus the sum (4.14) is a sum of positive terms with each partial sum being bounded above by $\|v\|^2$. It follows that the sequence of partial sums must converge to a number being at most $\|v\|^2$. ■

We also have the following result which should be compared to Theorem 4.4.20.

4.4.25 Theorem (Characterisations of enumerable Hilbert bases) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a separable $\mathbb{F}$-inner product space, and let $\mathscr{B} = \{e_j\}_{j\in\mathbb{Z}_{>0}}$ be an orthonormal set. The following four statements are equivalent:*

(i) $\mathscr{B}$ is basic;

(ii) $\mathscr{B}$ is total;

(iii) for every $v \in \mathsf{V}$ the equality

$$\|v\|^2 = \sum_{j=1}^{\infty} |\langle v, e_j\rangle|^2$$

*holds (**Parseval's equality**);*

(iv) for all $u, v \in \mathsf{V}$ we have

$$\langle u, v\rangle = \sum_{j=1}^{\infty} \langle u, e_j\rangle \overline{\langle v, e_j\rangle};$$

Also, the following two statements are equivalent:

(v) $\mathscr{B}^\perp = \{0_{\mathsf{V}}\}$;

(vi) $\mathscr{B}$ is maximal.

Finally, if V is a Hilbert space, the first four equivalent statements are equivalent to the last two equivalent statements.

Proof (i) $\implies$ (ii) Let $\mathscr{B} = \{e_j\}_{j\in\mathbb{Z}_{>0}}$ be basic and let $v \in \mathsf{V}$. We can then write

$$v = \sum_{j\in\mathbb{Z}_{>0}} c_j e_j$$

for some coefficients $c_j \in \mathbb{F}$, $j \in \mathbb{Z}_{>0}$. If we define

$$v_k = \sum_{j=1}^{k} c_j e_j$$

then the sequence $(v_k)_{k\in\mathbb{Z}_{>0}}$ converges to v. Thus $v \in \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\mathscr{B}))$ and so $\mathscr{B}$ is total.

(ii) $\implies$ (iii) Let $v \in \mathsf{V}$. Since $\mathscr{B}$ is total there exists a sequence $(v_k)_{k\in\mathbb{Z}_{>0}}$ in $\mathrm{span}_{\mathbb{F}}(\mathscr{B})$ such that $v = \lim_{k\to\infty} v_k$. For each $k \in \mathbb{Z}_{>0}$ write

$$v_k = c_{k1} e_{j_{k1}} + \cdots + c_{km_k} e_{j_{km_k}}$$

for $m_k \in \mathbb{Z}_{>0}$, coefficients $c_{k1}, \dots, c_{km_k} \in \mathbb{F}$, and distinct $j_{k1}, \dots, j_{km_k} \in I$. By Proposition 4.4.23 it follows that $c_{kl} = \langle v_k, e_{j_{kl}}\rangle$ for each $k \in \mathbb{Z}_{>0}$, $l \in \{1, \dots, m_k\}$. This means that we can write

$$v_k = \sum_{j=1}^{\infty} \langle v_k, e_j\rangle e_j$$

for each $k \in \mathbb{Z}_{>0}$, with the sum being finite.

We may also directly compute (cf. the proof of Theorem 4.4.24)

$$\|v_k\|^2 = \sum_{j=1}^{\infty} |\langle v_k, e_j\rangle|^2,$$

using the fact that the inner product commutes with finite sums. Now, using continuity of the norm and inner product, along with Theorem 3.5.2, gives

$$\|v\|^2 = \lim_{k\to\infty} \|v_k\|^2 = \lim_{k\to\infty} \sum_{j=1}^{\infty} |\langle v_k, e_j\rangle|^2 = \sum_{j=1}^{\infty} |\langle v, e_j\rangle|^2,$$

as desired.

(iii) $\implies$ (iv) For $u, v \in \mathsf{V}$ we have

$$\begin{aligned}
& \|u+v\|^2 = \sum_{j=1}^{\infty} |\langle u+v, e_j\rangle|^2 \\
\implies\quad & \|u\|^2 + \|v\|^2 + \langle u, v\rangle + \overline{\langle u, v\rangle} \\
& = \sum_{j=1}^{\infty} |\langle u, e_j\rangle|^2 + \sum_{j=1}^{\infty} |\langle v, e_j\rangle|^2 + \sum_{j=1}^{\infty} (\langle u, e_j\rangle\overline{\langle v, e_j\rangle} + \overline{\langle u, e_j\rangle\overline{\langle v, e_j\rangle}}) \\
\implies\quad & \mathrm{Re}(\langle u, v\rangle) = \sum_{j=1}^{\infty} \mathrm{Re}(\langle u, e_j\rangle\overline{\langle v, e_j\rangle}).
\end{aligned}$$

If $\mathbb{F} = \mathbb{R}$ this establishes the result. If $\mathbb{F} = \mathbb{C}$, a similar computation using the equality

$$\|u + \mathrm{i}v\|^2 = \sum_{j=1}^{\infty} |\langle u + \mathrm{i}v, e_j\rangle|^2$$

gives

$$\mathrm{Im}(\langle u, v\rangle) = \sum_{j=1}^{\infty} \mathrm{Im}(\langle u, e_j\rangle\overline{\langle v, e_j\rangle}).$$

(iv) $\implies$ (i) Since part (iv) obviously implies part (iii), we shall prove that (iii) implies (i). Thus we have

$$\|v\|^2 = \sum_{j=1}^{\infty} |\langle v, e_j\rangle|^2$$

for every $v \in \mathsf{V}$. For $k \in \mathbb{Z}_{>0}$ let us define

$$v_k = \sum_{j=1}^{k} \langle v, e_j\rangle e_j.$$

Note that

$$\begin{aligned}\langle v - v_k, v_k\rangle &= \Big\langle v - \sum_{j=1}^{k}\langle v, e_j\rangle e_j, \sum_{l=1}^{k}\langle v, e_l\rangle e_l\Big\rangle \\ &= \Big\langle v, \sum_{l=1}^{k}\langle v, e_l\rangle e_l\Big\rangle - \Big\langle \sum_{j=1}^{k}\langle v, e_j\rangle e_j, \sum_{l=1}^{k}\langle v, e_l\rangle e_l\Big\rangle \\ &= \sum_{l=1}^{k}|\langle v, e_l\rangle|^2 - \sum_{j=1}^{k}|\langle v, e_j\rangle|^2 = 0\end{aligned}$$

for every $k \in \mathbb{Z}_{>0}$. By the Pythagorean equality,

$$\|v\|^2 = \|v - v_k + v_k\|^2 = \|v - v_k\|^2 + \|v_k\|^2 \quad \Longrightarrow \quad \|v - v_k\|^2 = \|v\|^2 - \|v_k\|^2.$$

By assumption,

$$\lim_{k\to\infty}\|v_k\|^2 = \|v\|^2$$

and so

$$\lim_{k\to\infty}\|v - v_k\| = 0,$$

implying that

$$v = \sum_{j=1}^{\infty}\langle v, e_j\rangle e_j,$$

and so in particular implying that $\mathscr{B}$ is basic.

(v) $\Longrightarrow$ (vi) Suppose that $\mathscr{B}$ is not maximal. Then there exists an orthonormal set $\mathscr{B}'$ such that $\mathscr{B} \subset \mathscr{B}'$. Let $v \in \mathscr{B}' \setminus \mathscr{B}$. Then, clearly, $v \in \mathscr{B}^\perp$ and $v \neq 0_{\mathsf{V}}$. Thus $\mathscr{B}^\perp \neq \{0_{\mathsf{V}}\}$.

(vi) $\Longrightarrow$ (v) Suppose that $\mathscr{B}^\perp \neq \{0_{\mathsf{V}}\}$ and let $v \in \mathscr{B}^\perp$ have unit length. Then the set $\mathscr{B} \cup \{v\}$ is an orthonormal set that strictly contains $\mathscr{B}$. Thus $\mathscr{B}$ is not maximal.

(ii) $\Longrightarrow$ (v) By Proposition 4.1.13(iv) we have $\mathscr{B}^\perp = \mathrm{cl}(\mathrm{span}_{\mathbb{F}}(\mathscr{B}))^\perp$. From this fact, if $\mathscr{B}$ is total it immediately follows that $\mathscr{B}^\perp = \{0_{\mathsf{V}}\}$.

(vi) $\Longrightarrow$ (i) (assuming V is a Hilbert space) Let $v \in \mathsf{V}$. Bessel's inequality gives

$$\sum_{j=1}^{\infty}|\langle v, e_j\rangle|^2 \le \|v\|^2,$$

and this implies that the series on the right converges and so is Cauchy. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that

$$\sum_{j=k+1}^{l}|\langle v, e_j\rangle|^2 < \epsilon$$

for every $k, l \ge N$ with $l > k$. A direct computation using properties of inner products then gives

$$\Big\|\sum_{j=k+1}^{l}\langle v, e_j\rangle e_j\Big\|^2 = \sum_{j=k+1}^{l}|\langle v, e_j\rangle|^2 < \epsilon,$$

which shows that the series

$$\sum_{j=1}^{\infty} \langle v, e_j \rangle e_j$$

is Cauchy. By Theorem 3.4.17 this series converges, implying that $\mathscr{B}$ is basic. ■

4.4.4 Generalised Fourier series

In our general framework, the notion of a Fourier series is easily discussed. We shall discuss Fourier series (although we will think of this as being a means of getting at the inverse of the so-called CDFT) in Chapter IV-5. In this case, as we shall see, other issues not present in our general inner product space constructions, become relevant. Thus we focus our discussion in this section on the generalities. This will allow us to separate out these general considerations from the more specific ones in Chapter IV-5.

We begin with a definition that at this point is simply the giving of a name to something we already have been talking about.

4.4.26 Definition (Generalised Fourier series) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle \cdot, \cdot \rangle)$ be a $\mathbb{F}$-inner product space. If $\{e_i\}_{i \in I}$ is a Hilbert basis for V and if $v \in \mathsf{V}$, the ***generalised Fourier series*** for v is the series

$$v = \sum_{i \in I} \langle v, e_i \rangle e_i,$$

which converges to v. •

Let us consider some general and, therefore, more or less elementary examples.

4.4.27 Examples (Generalised Fourier series)

1. Let $(\mathsf{V}, \langle \cdot, \cdot \rangle)$ be a finite-dimensional inner product space with Hilbert basis $\{e_1, \dots, e_n\}$. The generalised Fourier series for $v \in \mathsf{V}$ is then simply the representation of v in the (Hamel) basis $\{e_1, \dots, e_n\}$, just as prescribed by Corollary 4.4.18:

$$v = \langle v, e_1 \rangle e_1 + \cdots + \langle v, e_n \rangle e_n.$$

2. Next consider the inner product space $(\mathbb{F}_0^\infty, \langle \cdot, \cdot \rangle_1)$ with its standard basis $\{\boldsymbol{e}_j\}_{j \in \mathbb{Z}_{>0}}$; this is a Hilbert basis as we saw in Example 4.4.8–2. In this case the Hilbert basis is also a basis in the usual sense. Thus the generalised Fourier series for $\boldsymbol{v} \in \mathbb{F}_0^\infty$,

$$\boldsymbol{v} = \sum_{j=1}^{\infty} \boldsymbol{v}(j) \boldsymbol{e}_j,$$

is simply the representation of $\boldsymbol{v}$ with respect to a basis in the usual sense.

3. Finally, let us consider the completion $(\ell^2(\mathbb{F}), \langle\cdot,\cdot\rangle_2)$ of $(\mathbb{F}_0^\infty, \langle\cdot,\cdot\rangle_2)$. In this case the generalised Fourier series for $v \in \ell^2(\mathbb{F})$ has the form

$$v = \sum_{j=1}^{\infty} v(j)e_j.$$

Note that this is *not* the representation of v in a basis in the usual sense because the sum is possibly finite. Indeed, it is quite clear that $\{e_j\}_{j\in\mathbb{Z}_{>0}}$ is not a (Hamel;) basis. Moreover, we shall see in Theorem 4.4.36 that any (Hamel) basis for $\ell^2(\mathbb{F})$ has cardinality strictly greater than that of $\mathbb{Z}_{>0}$. •

The preceding two examples illustrate the difference between the purely algebraic notion of a Hamel basis and the analytical notion of a Hilbert basis. It is probably worth understanding the message these examples are trying to pass on.

Let us now give a useful geometric interpretation of the generalised Fourier series. We recall from Section 4.1.5 the notation $\mathrm{dist}(v, S)$ for the distance from $v \in \mathsf{V}$ to a subset $S \subseteq \mathsf{V}$.

4.4.28 Theorem (The best approximation property of generalised Fourier series) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space, and let $\{\mathrm{e}_i\}_{i\in\mathrm{I}}$ be a Hilbert basis for V. For $\mathrm{J} \subseteq \mathrm{I}$ let us abbreviate*

$$\mathsf{V}_\mathrm{J} = \mathrm{cl}(\mathrm{span}_\mathbb{F}(\mathrm{e_j} |\ \mathrm{j} \in \mathrm{J})),$$

and assume that V_J is complete. If $\mathrm{v} \in \mathsf{V}$ and if $\mathrm{J} \subseteq \mathrm{I}$, then

$$\mathrm{v_J} \triangleq \sum_{\mathrm{j}\in\mathrm{J}} \langle \mathrm{v}, \mathrm{e_j}\rangle \mathrm{e_j}$$

is the unique vector in V_J for which $\mathrm{dist}(\mathrm{v}, \mathsf{V}_\mathrm{J}) = \|\mathrm{v} - \mathrm{v_J}\|$.

Proof We first claim that the series v_J converges. Since we are assuming that V_J is complete, it suffices by Theorem 3.4.17 to show that the series v_J is Cauchy. Let $\epsilon \in \mathbb{R}_{>0}$. Since the series

$$\sum_{j\in J} |\langle v, c_j\rangle|^2$$

is convergent by Theorem 4.4.6 it is also Cauchy. Thus there exists a finite subset $J \subseteq I$ such that

$$\sum_{j\in J'} |\langle v, e_j\rangle|^2 < \epsilon$$

for every finite subset $J' \subseteq I$ for which $J' \cap J = \varnothing$. Then, by Theorem 4.4.20,

$$\Big\|\sum_{j\in J'} \langle v, e_j\rangle e_j\Big\|^2 = \sum_{j\in J'} |c_j|^2 < \epsilon$$

for every finite subset $J' \subseteq I$ for which $J' \cap J = \varnothing$. This gives convergence of the series for v_J, as desired.

Now, by Theorem 4.1.26, it suffices to show that $v - v_J \in \mathsf{V}_J^{\perp}$. By Proposition 4.1.13(iv) it suffices to show that $\langle v - v_J, e_j\rangle = 0$ for every $j \in J$. But this holds since

$$\langle v - v_J, e_j\rangle = \Big\langle v - \sum_{j' \in J} \langle v, e_{j'}\rangle e_{j'}, e_j\Big\rangle = \langle v, e_{j'}\rangle - \langle v, e_{j'}\rangle = 0,$$

where we swap the sum and inner product by Proposition 4.2.1 and Theorem 3.5.2. ■

The preceding discussion has to do with representing a vector in an inner product space by a generalised Fourier series. The next result tells us that any "reasonable" collection of coefficients are those of a generalised Fourier series.

4.4.29 Theorem (Riesz–Fischer[5] Theorem) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-Hilbert space, and let $(e_i)_{i\in I}$ be an orthonormal family. If $(c_i)_{i\in I}$ is a family of numbers such that the series*

$$\sum_{i\in I} |c_i|^2 \tag{4.17}$$

converges in the sense of Definition I-2.4.31, then the series

$$\sum_{i\in I} c_i e_i \tag{4.18}$$

converges in the sense of Definition 3.4.16. Moreover, if the series converges to $v \in \mathsf{V}$ then $c_i = \langle v, e_i\rangle$, $i \in I$.

Proof We claim that the sum (4.18) is Cauchy. Let $\epsilon \in \mathbb{R}_{>0}$. Since the series (4.17) is convergent and so Cauchy, there exists a finite set $J \subseteq I$ such that

$$\sum_{j\in J'} |c_j|^2 < \epsilon$$

for every finite subset $J' \subseteq I$ for which $J \cap J' = \emptyset$. By Theorem 4.4.20 we then have

$$\Big\|\sum_{j\in J'} c_j e_j\Big\|^2 = \sum_{j\in J'} |c_j|^2 < \epsilon$$

for every finite subset $J' \subseteq I$ for which $J \cap J' = \emptyset$. Thus the series (4.18) is Cauchy, and so convergent by Theorem 3.4.17. The last assertion is simply Proposition 4.4.5. ■

4.4.5 Classification of Hilbert spaces

In this section we use the idea of Hilbert bases to characterise all Hilbert spaces. As we shall see, the classification is actually quite simple, just as with the classification of all vector spaces induced by the size of their bases.

First let us assert that the dimension of an inner product space, when it exists, is well defined.

[5] Frigyes Riesz (1880–1956) was born in what is now Hungary and was one of the founders of functional analysis. His younger brother Marcel was also a mathematician of some note. Ernst Sigismund Fischer (1875–1954) was an Austrian mathematician whose contributions to mathematics were in the areas of algebra and analysis.

4.4.30 Theorem (Invariance of cardinality of maximal orthonormal sets) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *if* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *is an* $\mathbb{F}$*-inner product space, and if* $\{e_i\}_{i\in I}$ *and* $\{f_j\}_{j\in J}$ *are Hilbert bases for* V*, then* $\mathrm{card}(I) = \mathrm{card}(J)$.

Proof If V possesses a finite maximal orthonormal set, then this set is a Hilbert basis and so also a Hamel basis by Theorem 4.4.14. Moreover, from the same result, every Hilbert basis for V is a Hamel basis. By Theorem I-4.5.25 every Hamel basis for V has the same cardinality, and so the result follows when V has a finite maximal orthonormal set.

Next suppose that V has two infinite maximal orthonormal sets $\{e_i\}_{i\in I}$ and $\{f_j\}_{j\in J}$. For $j \in J$ denote

$$I_j = \{i \in I \mid \langle f_j, e_i\rangle \neq 0\}.$$

Since, by Theorem 4.4.6, we have

$$\sum_{i\in I} \langle f_j, e_i\rangle \leq \|f_j\|^2 = 1,$$

it follows from Proposition I-2.4.33 that I_j is enumerable for each $j \in J$. We claim that $I = \cup_{j\in J} I_j$. It is clear that $\cup_{j\in J} I_j \subseteq I$. Suppose that the converse inclusion does not hold and let $i \in I \setminus (\cup_{j\in J} I_j)$. This means, by definition of the sets I_j, $j \in J$, that $\langle f_j, e_i\rangle = 0$ for every $j \in J$. By Theorem 4.4.9 this means that $f_j = 0_{\mathsf{V}}$; from this we conclude that $I \subseteq \cup_{j\in J} I_j$. Now we have

$$\mathrm{card}(I) = \mathrm{card}(\cup_{j\in J} I_j) \leq \mathrm{card}(\mathbb{Z}_{>0})\,\mathrm{card}(J) \leq \mathrm{card}(J)\,\mathrm{card}(J) = \mathrm{card}(J),$$

using Theorem I-1.7.17 and its Corollary I-1.7.18. By swapping the rôles of I and J we similarly prove that $\mathrm{card}(J) \leq \mathrm{card}(I)$, and so the theorem follows from Theorem I-1.7.12. ■

The result has the following obvious (by Theorem 4.4.9) corollary.

4.4.31 Corollary (Invariance of cardinality of Hilbert bases) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *if* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *is an* $\mathbb{F}$*-inner product space, and if* $\{e_i\}_{i\in I}$ *and* $\{f_j\}_{j\in J}$ *are Hilbert bases for* V*, then* $\mathrm{card}(I) = \mathrm{card}(J)$.

The preceding theorem and corollary make sense of the following definition.

4.4.32 Definition (Hilbert dimension) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be a $\mathbb{F}$-inner product space. The ***Hilbert dimension*** of V is the cardinality of any maximal orthonormal set in V. We denote by $\mathrm{hdim}_{\mathbb{F}}(\mathsf{V})$ the Hilbert dimension of V. •

For vector spaces we saw in Proposition I-4.5.30 that the dimension was an isomorphism invariant, indeed the only isomorphism. That is to say, two vector spaces are isomorphic if and only if they have the same dimension. We would like to establish a similar assertion for inner product spaces, but replacing "dimension" with "Hilbert dimension" and replacing "isomorphism" with "isomorphism of inner product spaces." But such a result is not actually true, as we shall see. The desired result *is* true, however, if we restrict ourselves to the most interesting case of Hilbert spaces.

4.4.33 Theorem (Hilbert dimension characterises Hilbert spaces) *If $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and if $(\mathsf{V}_1, \langle\cdot,\cdot\rangle_1)$ and $(\mathsf{V}_2, \langle\cdot,\cdot\rangle_2)$ are $\mathbb{F}$-Hilbert spaces, then the following statements are equivalent:*

(i) V_1 *and* V_2 *are isomorphic as inner product spaces;*

(ii) $\mathrm{hdim}_{\mathbb{F}}(\mathsf{V}_1) = \mathrm{hdim}_{\mathbb{F}}(\mathsf{V}_2)$.

Proof (i) $\Longrightarrow$ (ii) Let $\mathsf{L}\colon \mathsf{V}_1 \to \mathsf{V}_2$ be an inner product space isomorphism and let $\mathscr{B}_1$ be a Hilbert basis for V_1. Define

$$\mathscr{B}_2 = \{\mathsf{L}(u) \mid u \in \mathscr{B}_1\};$$

we claim that $\mathscr{B}_2$ is a Hilbert basis for V_2. First let us prove that $\mathscr{B}_2$ is orthonormal. If $\mathsf{L}(u_1), \mathsf{L}(u_2) \in \mathscr{B}_2$ we have

$$\langle \mathsf{L}(u_1), \mathsf{L}(u_2)\rangle_2 = \langle u_1, u_2\rangle_2,$$

using the fact that L is an isomorphism of inner product spaces. Thus $\mathsf{L}(u_1)$ and $\mathsf{L}(u_2)$ are orthogonal if and only if they are distinct. Similarly one computes $\|\mathsf{L}(u)\| = 1$ for $u \in \mathscr{B}_1$. Thus $\mathscr{B}_2$ is indeed orthonormal. Now suppose that $v_0 \in \mathscr{B}_2^{\perp}$ and let $u_0 = \mathsf{L}^{-1}(v)$. Then, for every $u \in \mathscr{B}_1$,

$$\langle v_0, \mathsf{L}(u)\rangle_2 = \langle u_0, u\rangle_1 = 0,$$

implying that $u_0 = 0_{\mathsf{V}_1}$ by Theorem 4.4.9, since L is an isomorphism of inner product spaces, and since $\mathscr{B}_1$ is maximal. We conclude that $\mathscr{B}_2$ is maximal and so a Hilbert basis By Theorem 4.4.9.

(ii) $\Longrightarrow$ (i) Let $\mathscr{B}_1$ and $\mathscr{B}_2$ be Hilbert bases for V_1 and V_2, respectively. By assumption there exists a bijection $\phi\colon \mathscr{B}_1 \to \mathscr{B}_2$. Note that by Theorem 4.4.9 every vector in V_1 can be written as

$$\sum_{u\in\mathscr{B}_1} c_u u$$

for coefficients $c_u \in \mathbb{F}$, $u \in \mathscr{B}_1$, such that

$$\sum_{u\in\mathscr{B}_1} |c_u|^2 < \infty.$$

Using this fact, let us define $\mathsf{L}\colon \mathsf{V}_1 \to \mathsf{V}_2$ by

$$\mathsf{L}\Big(\sum_{u\in\mathscr{B}_1} c_u u\Big) = \sum_{u\in\mathscr{B}_1} c_u \phi(u).$$

We must show that L is well-defined and is an isomorphism of inner product spaces. To show that L is well-defined, we must show that it defines an element of V_2. This, however, follows from the Riesz-Fischer Theorem. Linearity of L follows from the fact that $\mathsf{L}(u) = \phi(u)$ for every $u \in \mathscr{B}_1$ (why?) and from the calculations

$$\mathsf{L}\Big(\sum_{u\in\mathscr{B}_1} (a_u u + b_u u)\Big) = \sum_{u\in\mathscr{B}} a_u\phi(u) + \sum_{u\in\mathscr{B}} b_u\phi(u) = \mathsf{L}\Big(\sum_{u\in\mathscr{B}_1} a_u u\Big) + \mathsf{L}\Big(\sum_{u\in\mathscr{B}_1} b_u u\Big),$$

for $a_u, b_u \in \mathbb{F}$, $u \in \mathscr{B}_1$, and

$$\mathsf{L}\Big(\sum_{u\in\mathscr{B}_1} \alpha(c_u u)\Big) = \alpha \sum_{u\in\mathscr{B}_1} c_u\phi(u) = \alpha\mathsf{L}\Big(\sum_{u\in\mathscr{B}_1} c_u u\Big),$$

for $\alpha \in \mathbb{F}$ and $c_u \in \mathbb{F}$, $u \in \mathscr{B}_1$. (Of course, in the above computations we require that $\sum_{u\in\mathscr{B}_1}|a_u|^2$, $\sum_{u\in\mathscr{B}_1}|b_u|^2$, and $\sum_{u\in\mathscr{B}_1}|c_u|^2$ be finite.) The swapping of sums with addition and multiplication is justified by Proposition 4.2.1 and Theorem 3.5.2. Finally, we must show that L preserves the inner product. Using Theorem 4.4.9 we compute

$$\begin{aligned}\Big\langle \mathsf{L}\Big(\sum_{u\in\mathscr{B}_1} a_u u\Big), \mathsf{L}\Big(\sum_{u'\in\mathscr{B}_1} b_{u'}u'\Big)\Big\rangle_2 &= \Big\langle \sum_{u\in\mathscr{B}_1} a_u\phi(u), \sum_{u'\in\mathscr{B}_1} b_{u'}\phi(u')\Big\rangle_2\\ &= \sum_{u\in\mathscr{B}_1} a_u\overline{b_u}\\ &= \Big\langle \sum_{u\in\mathscr{B}_1} a_u u, \sum_{u'\in\mathscr{B}_1} b_{u'}u'\Big\rangle_1,\end{aligned}$$

as desired. ■

Now that we have decided that the Hilbert dimension of a Hilbert space is its only property invariant under isomorphism of inner product spaces, let us provide for the set of Hilbert spaces with a prescribed Hilbert dimension a simple representative. It is perhaps useful to remind ourselves how this is done for vector spaces. If V is a vector space over a field F with dimension $\operatorname{card}(I)$, then we showed in Theorem I-4.5.45 that V is isomorphic to the direct sum $\bigoplus_{i\in I}\mathsf{F}$. Thus the direct sum $\bigoplus_{i\in I}\mathsf{F}$ serves as a simple representative of *all* vector spaces with dimension equal to V. The situation is rather similar for Hilbert spaces.

The following theorem describes the simple representative we are after.

4.4.34 Theorem (A "canonical" Hilbert space of a prescribed Hilbert dimension) *For* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and for a set* I*, define*

$$\ell^2(\mathrm{I};\mathbb{F}) = \Big\{\phi\colon \mathrm{I} \to \mathbb{F} \;\Big|\; \sum_{\mathrm{i}\in\mathrm{I}}|\phi(\mathrm{i})|^2 < \infty\Big\}$$

and define an inner product on $\ell^2(\mathrm{I};\mathbb{F})$ *by*

$$\langle\phi,\psi\rangle_2 = \sum_{\mathrm{i}\in\mathrm{I}} \phi(\mathrm{i})\overline{\psi(\mathrm{i})}.$$

Then $(\ell^2(\mathrm{I};\mathbb{F}), \langle\cdot,\cdot\rangle_2)$ *is a Hilbert space with Hilbert dimension* $\operatorname{card}(\mathrm{I})$*.*

Proof Note that $\ell^2(I;\mathbb{F}) = \ell^2(\bigoplus_{i\in I}\mathbb{F})$ in the context of Definition 3.8.26. It then follows from Theorem 3.8.27 that $\ell^2(I;\mathbb{F})$ is a Banach space with respect to the norm $\|\cdot\|_2$ defined by

$$\|\phi\|_2 = \sum_{i\in I}|\phi(i)|^2.$$

In order to show that it is a Hilbert space we should show that the norm is derived from the given inner product $\langle\cdot,\cdot\rangle_2$. First of all, for $\phi,\psi\in\ell^2(I;\mathbb{F})$, by Proposition I-2.4.33 there exists an injection $\kappa\colon\mathbb{Z}_{>0}\to I$ such that $\phi(i)=\psi(i)=0$ for $i\notin\operatorname{image}(\kappa)$ and such that

$$\sum_{i\in I}|\phi(i)|^2=\sum_{j=1}^{\infty}|\phi(\kappa(j))|^2,\quad \sum_{i\in I}|\psi(i)|^2=\sum_{j=1}^{\infty}|\psi(\kappa(j))|^2.$$

Then, for $n\in\mathbb{Z}_{>0}$,

$$\Big|\sum_{j=1}^{n}\phi(\kappa(j))\overline{\psi(\kappa(j))}\Big|\leq\Big(\sum_{j=1}^{n}|\phi(\kappa(j))|^2\Big)^{1/2}\Big(\sum_{j=1}^{n}|\psi(\kappa(j))|^2\Big)^{1/2},$$

using the Cauchy–Bunyakovsky–Schwarz inequality. Letting $n\to\infty$ we get

$$\Big|\sum_{i\in I}\phi(i)\overline{\psi(i)}\Big|^2\leq\|\phi\|_2\|\psi\|_2<\infty.$$

Thus the sum defining the inner product converges. Completing the proof is now a matter of verifying the inner product axioms for $\langle\cdot,\cdot\rangle_2$, justifying the swapping of infinite sums and inner products using Proposition 4.2.1 and Theorem 3.5.2. ■

From the preceding result and from Theorem 4.4.33 (and its proof) we deduce the following interesting conclusion.

4.4.35 Corollary (Characterisation of Hilbert spaces up to isomorphism of inner product spaces) *If $\mathbb{F}\in\{\mathbb{R},\mathbb{C}\}$, if $(\mathsf{V},\langle\cdot,\cdot\rangle)$ is a Hilbert space, and if $\{e_i\}_{i\in I}$ is a Hilbert basis for V, then the map $\mathsf{L}\colon\mathsf{V}\to\ell^2(I;\mathbb{F})$ defined by*

$$\mathsf{L}\Big(\sum_{i\in I}c_ie_i\Big)=\sum_{i\in I}c_i\mathbf{e}_i$$

is an isomorphism of inner product spaces.

It is worth digesting and understanding clearly the difference between the preceding corollary and its counterpart Theorem I-4.5.45 for vector spaces. For a given set I the "canonical" $\mathbb{F}$-vector space of Hamel dimension $\operatorname{card}(I)$ is $\mathbb{F}_0^I$ and the "canonical" $\mathbb{F}$-Hilbert space of Hilbert dimension $\operatorname{card}(I)$ is $\ell^2(I;\mathbb{F})$. Both are subspaces of $\mathbb{F}^I$ (see Notation I-4.5.44 for this notation). Moreover, $\mathbb{F}_0^I$ is a subspace of $\ell^2(I;\mathbb{F})$, and is a strict subspace unless $\operatorname{card}(I)$ is finite. Indeed, $\mathbb{F}_0^I$ and $\ell^2(I;\mathbb{F})$ are rather different objects when $\operatorname{card}(I)$ is not finite. For example, to make sense of the vector space $\ell^2(I;\mathbb{F})$ requires some analysis that is not required to make sense of $\mathbb{F}_0^I$. Note, for example, that we have not defined $\ell^2(I;\mathsf{F})$ for a general field F as a general field does not possess the absolute value structure of $\mathbb{R}$ or $\mathbb{C}$ that is needed to make things go. Thus $\ell^2(I;\mathbb{F})$ is, in some sense, a "deeper" object than $\mathbb{F}_0^I$. However, there is a strong connection between $\mathbb{F}_0^I$ and $\ell^2(I;\mathbb{F})$ in that the latter is the completion of the former if one uses the inner product

$$\langle\phi,\psi\rangle_2=\sum_{i\in I}\phi(i)\overline{\psi(i)}\quad\text{(sum finite)}$$

on $\mathbb{F}_0^I$.

The preceding discussion leads one to the following natural question: "What is the relationship between the Hamel dimension and the Hilbert dimension of an inner product space?" For inner product spaces the answer can be, "They are equal." For example, $\mathbb{F}_0^I$ with the inner product $\langle\cdot,\cdot\rangle_2$ defined above has the same Hilbert and Hamel dimension. The question is deeper for Hilbert spaces. Indeed, from Theorem 3.6.26 and Theorem 4.4.21 we have the following result.

4.4.36 Theorem (Dimension of separable Hilbert space) *If* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and if* $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ *is a separable infinite-dimensional* $\mathbb{F}$*-Hilbert space, then* $\dim_{\mathbb{F}}(\mathsf{V}) = \mathrm{card}(\mathbb{R})$*.*

In particular, this shows that $\mathbb{F}_0^\infty$ and $\ell^2(\mathbb{F})$ have different Hamel dimension, and so are not isomorphic. The story for Hilbert spaces of general dimension is more complicated, and we refer to the notes in Section 4.4.6.

4.4.6 Notes

The Riesz–Fischer Theorem was published independently by Fischer [1907] and Riesz [1907a] and Riesz [1907b].

Evans and Tapia [1970] study the relationship between the Hamel and Schauder dimensions of a Banach space. Applying their result to Hilbert spaces, their conclusions are that there is a condition on the cardinal numbers that characterise those infinite-dimensional Hilbert spaces whose Hamel and Hilbert dimensions agree. They point out that $\aleph_0 = \mathrm{card}(\mathbb{Z}_{>0})$ does not satisfy this condition (and so separable Hilbert spaces necessarily have different Hamel and Hilbert dimension) while $\aleph_1 = \mathrm{card}(\mathbb{R})$ does satisfy this condition (and so the Hilbert space $\ell^2(\mathbb{R};\mathbb{F})$ has equal Hamel and Hilbert dimension). The proof of Evans and Tapia assumes the so-called Generalised Continuum Hypothesis which asserts that $2^{\aleph_o} = \aleph_{o+1}$ for every ordinal o.[6]

Exercises

4.4.1 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space. Show that an orthogonal set is linearly independent.

4.4.2 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Show that the standard basis $\{e_j\}_{j\in\mathbb{Z}_{>0}}$ for $\mathbb{F}_0^\infty$ is a maximal orthonormal family.

4.4.3 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, let $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ be an $\mathbb{F}$-inner product space, and let J be either the set $\{1, \dots, n\}$ for some $n \in \mathbb{Z}_{>0}$ or the set $\mathbb{Z}_{>0}$. Show that if $(u_j)_{j\in J}$ is orthonormal, then applying the Gram–Schmidt orthonormalisation procedure to this family gives the same family back again.

[6] The cardinals $\aleph_o$, defined for ordinals o, are defined using transfinite recursion as follows. Take $\aleph_0$ to be the cardinality of $\mathbb{Z}_{>0}$. Assuming that $\aleph_o$ has been defined, one defines $\aleph_{o+1}$ to be the successor (see Definition I-1.4.1) of $\aleph_o$.

4.4.4 Prove Proposition 4.4.17. Point out the parts of your argument that are not valid in the infinite-dimensional case.

4.4.5 Prove Proposition 4.4.19. Point out the parts of your argument that are not generally valid for enumerable orthonormal sets $(e_j)_{j\in\mathbb{Z}_{>0}}$.

4.4.6 Prove Theorem 4.4.20. Point out the parts of your argument that are not generally valid for enumerable orthonormal sets $(e_j)_{j\in\mathbb{Z}_{>0}}$.

In the following exercise you will see just how fine is the notion of a maximal orthonormal set. Taking away any vector, or attempting to add a vector, ruins the maximality.

4.4.7 Let $\mathscr{B} = \{e_j\}_{j\in\mathbb{Z}_{>0}}$ be a maximal orthonormal set in an inner product space $(\mathsf{V}, \langle\cdot,\cdot\rangle)$.

(a) Show that for any $k \in \mathbb{Z}_{>0}$ the set $\mathscr{B} \setminus \{e_k\}$ is not maximal.

(b) Show that there is no vector $e_0 \in \mathsf{V}$ with the property that $\{e_0\} \cup \mathscr{B}$ is a maximal orthonormal set.

4.4.8 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \langle\cdot,\cdot\rangle_2)$ be an $\mathbb{F}$-Hilbert space. Let $\{e_i\}_{i\in I}$ be a Hilbert basis. Show that if $\alpha \in \mathsf{V}^*$ then the vector $v_\alpha \in \mathsf{V}$ associated with α by Corollary 4.2.5 satisfies $\langle v_\alpha, e_i\rangle = \overline{\alpha(e_i)}$ for each $i \in I$.

Chapter 5
Convexity

In Section II-1.9 we studied convexity in the context of n-dimensional Euclidean space $\mathbb{R}^n$. In this chapter we generalise this discussion to arbitrary vector spaces, and also provide more a more advanced discussion of topics in convexity. The results in this chapter will be essential in our constructions of locally convex topologies in Chapter 6.

Do I need to read this chapter? The material in this chapter is essential reading for the important material in Chapter 6. •

Section 5.1

Convex bodies

In this section we study in some detail a particular class of convex sets, those with nonempty interior. Of course, since every convex set has a nonempty interior in the affine hull of the convex set, i.e., the smallest affine subspace containing the convex set. Thus, convex sets with a nonempty interior are very general. However, thinking about them specifically reveals many interesting features of these bodies.

Do I need to read this section? The results in this section can be bypassed on a first reading and/or until needed. Some of the constructions are interesting, however, so if you like interesting things, then read this section. •

5.1.1 Definitions

In this section we give a few results concerning convex sets with nonempty interior. These sets have a name.

move balanced?

5.1.1 Definition (Convex body, balanced convex body) A ***convex body*** is a convex set $C \subseteq \mathbb{R}^n$ for which $\operatorname{int}(C) \neq \emptyset$. A convex body C is ***balanced*** if $\boldsymbol{x} \in C$ implies that $-\boldsymbol{x} \in C$. •

Associated with a convex body is the following notion.

5.1.2 Definition (Gauge of a convex body) If $C \subseteq \mathbb{R}^n$ is a convex body, the ***gauge*** of C is the map $\mu_C \colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ given by

$$\mu_C(\boldsymbol{x}) = \inf\{\lambda \in \mathbb{R}_{>0} \mid \boldsymbol{x} \in \lambda C\}. \quad \bullet$$

The gauge of a compact convex set is important.

5.1.3 Theorem (The gauge of a balanced compact convex body is a norm) *If $\mathrm{C} \subseteq \mathbb{R}^{\mathrm{n}}$ is a balanced compact convex body, then the gauge μ_{C} of C is a norm on $\mathbb{R}^{\mathrm{n}}$. That is to say,*

(i) $\mu_C(\alpha\mathbf{x}) = |\alpha|\mu_C(\mathbf{x})$ *for* $\alpha \in \mathbb{R}$ *and* $\mathbf{x} \in \mathbb{R}^n$,

(ii) $\mu_C(\mathbf{x}) \geq 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$,

(iii) $\mu_C(\mathbf{x}) = 0$ *only if* $\mathbf{x} = \mathbf{0}$,

(iv) $\mu_C(\mathbf{x}_1 + \mathbf{x}_2) \leq \mu_C(\mathbf{x}_1) + \mu_C(\mathbf{x}_2)$.

Moreover, C is the unit ball for the norm μ_{C}, i.e.,

$$C = \{\mathbf{x} \in \mathbb{R}^n \mid \mu_C(\mathbf{x}) \leq 1\}.$$

Proof (i) If $\alpha = 0$ and $\boldsymbol{x} \in C$, then

$$\mu_C(\alpha\boldsymbol{x}) = \mu_C(\mathbf{0}) = \inf\{\lambda \in \mathbb{R}_{>0} \mid \mathbf{0} \in \lambda C\} = \inf \mathbb{R}_{>0} = 0,$$

since $\mathbf{0} \in C$. Now let $\alpha \in \mathbb{R}_{>0}$. Then

$$\begin{aligned}\mu_C(\alpha\boldsymbol{x}) &= \inf\{\lambda \in \mathbb{R}_{>0} \mid \alpha\boldsymbol{x} \in \lambda C\}\\ &= \inf\{\lambda \in \mathbb{R}_{>0} \mid \boldsymbol{x} \in \tfrac{\lambda}{\alpha}C\}\\ &= \inf\{\alpha\lambda \in \mathbb{R}_{>0} \mid \boldsymbol{x} \in \lambda C\}\\ &= \alpha\inf\{\lambda \in \mathbb{R}_{>0} \mid \boldsymbol{x} \in \lambda C\} = \alpha\mu_C(\boldsymbol{x}),\end{aligned}$$

using Proposition I-2.2.28. Finally, if $\alpha \in \mathbb{R}_{<0}$, then

$$\begin{aligned}\mu_C(\alpha\boldsymbol{x}) &= \inf\{\lambda \in \mathbb{R}_{>0} \mid \alpha\boldsymbol{x} \in \lambda C\}\\ &= \inf\{\lambda \in \mathbb{R}_{>0} \mid \alpha\boldsymbol{x} \in (-\lambda)C\}\\ &= \inf\{\lambda \in \mathbb{R}_{>0} \mid (-\alpha)\boldsymbol{x}) \in \lambda C\}\\ &= -\alpha\inf\{\lambda \in \mathbb{R}_{>0} \mid \boldsymbol{x} \in \lambda C\} = -\alpha\mu_C(\boldsymbol{x}),\end{aligned}$$

using the fact that C is balanced and using the fact that this part of the result has been proved for $\alpha \in \mathbb{R}_{>0}$.

(ii) This is obvious.

(iii) Suppose that $\mu_C(\boldsymbol{x}) = 0$ and let $\epsilon \in \mathbb{R}_{>0}$. Let $\delta \in \mathbb{R}_{>0}$ be such that $\delta C \subseteq \overline{\mathsf{B}}^n(\epsilon, \mathbf{0})$. By definition of μ_C, $\boldsymbol{x} \in \overline{\mathsf{B}}^n(\epsilon, \mathbf{0})$. As this holds for every $\epsilon \in \mathbb{R}_{>0}$ we much have $\boldsymbol{x} = \mathbf{0}$.

(iv) Let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$. Note that if $\boldsymbol{x}_1 \in \lambda_1 C$ and if $\boldsymbol{x}_2 \in \lambda_2 C$ for $\lambda_1, \lambda_2 \in \mathbb{R}_{>0}$, then we write $\boldsymbol{x}_1 = \lambda_1\boldsymbol{y}_1$ and $\boldsymbol{x}_2 = \lambda_2\boldsymbol{y}_2$ for $\boldsymbol{y}_1, \boldsymbol{y}_2 \in C$. Then

$$\begin{aligned}\frac{\lambda_1}{\lambda_1+\lambda_2}\boldsymbol{y}_1 + \frac{\lambda_2}{\lambda_1+\lambda_2}\boldsymbol{y}_1 \in C \quad &\Longrightarrow \quad \lambda_1\boldsymbol{y}_1 + \lambda_2\boldsymbol{y}_2 \in (\lambda_1+\lambda_2)C\\ &\Longrightarrow \quad \boldsymbol{x}_1 + \boldsymbol{x}_2 \in (\lambda_1+\lambda_2)C.\end{aligned}$$

Thus

$$\{\lambda_1, \lambda_2 \in \mathbb{R}_{>0} \mid \boldsymbol{x}_1 \in \lambda_1 C,\ \boldsymbol{x}_2 \in \lambda_2 C\} \subseteq \{\lambda_1, \lambda_2 \in \mathbb{R}_{>0} \mid \boldsymbol{x}_1 + \boldsymbol{x}_2 \in (\lambda_1+\lambda_2)C\}.$$

Thus we have

$$\begin{aligned}\mu_C(\boldsymbol{x}_1+\boldsymbol{x}_2) &= \inf\{\lambda \in \mathbb{R} \mid \boldsymbol{x}_1+\boldsymbol{x}_2 \in \lambda C\}\\ &= \inf\{\lambda_1+\lambda_2 \mid \boldsymbol{x}_1+\boldsymbol{x}_2 \in (\lambda_1+\lambda_2)C\}\\ &\leq \inf\{\lambda_1+\lambda_2 \mid \boldsymbol{x}_1 \in \lambda_1 C,\ \boldsymbol{x}_2 \in \lambda_2 C\}\\ &= \inf\{\lambda_1 \mid \boldsymbol{x}_1 \in \lambda_1 C\} + \inf\{\lambda_2 \mid \boldsymbol{x}_2 \in \lambda_2 C\} = \mu_C(\boldsymbol{x}_1) + \mu_C(\boldsymbol{x}_2),\end{aligned}$$

using Proposition I-2.2.28.

For the final assertion of the theorem, first let $\boldsymbol{x} \in \overline{\mathsf{B}}(1, \mathbf{0})$ be in the unit ball for the norm $\|\cdot\|$. Then $\mu_C(\boldsymbol{x}) \in [0, 1]$. Let $(\epsilon_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence converging to 0 and note that $\boldsymbol{x} \in (1+\epsilon_j)C$ for every $j \in \mathbb{Z}_{>0}$. Thus, for each $j \in \mathbb{Z}_{>0}$, there exists $\boldsymbol{x}_j \in C$ such that $\boldsymbol{x} = (1+\epsilon_j)\boldsymbol{x}_j$. Thus $\lim_{j\to\infty}\boldsymbol{x}_j = \boldsymbol{x}$, giving $\boldsymbol{x} \in C$ since C is closed. Next, suppose that $\boldsymbol{x} \in C$. Then, since C is symmetric, $\lambda\boldsymbol{x} \in C$ for $\lambda \in [0, 1]$. Therefore,

$$\mu_C(\boldsymbol{x}) = \inf\{\lambda \in \mathbb{R}_{>0} \mid \boldsymbol{x} \in \lambda C\} \leq 1,$$

giving $\boldsymbol{x} \in \overline{\mathsf{B}}(1, \mathbf{0})$, as desired. ■

An important class of convex sets are the ellipsoids.

5.1.4 Definition (Ellipsoid) An ***ellipsoid*** is a subset $E \subseteq \mathbb{R}^n$ that is given by $f(\overline{\mathsf{B}}^n(1, \mathbf{0}))$ where $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is an invertible affine map. •

By Exercise II-1.9.7 it follows that an ellipsoid is convex. Let us show that this definition of ellipsoid agrees with the perhaps more familiar one. We refer the reader ahead to for a discussion of linear maps defined on finite-dimensional inner product spaces.

what

5.1.5 Proposition (Characterisation of ellipsoids) *For a subset* $\mathrm{E} \subseteq \mathbb{R}^{\mathrm{n}}$*, the following statements are equivalent:*

(i) $\mathrm{E} \subseteq \mathbb{R}^{\mathrm{n}}$ *is an ellipsoid;*

(ii) there exists a symmetric and positive-definite map $\mathbf{A} \in \mathrm{L}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{n}})$ *and* $\mathbf{b} \in \mathbb{R}^{\mathrm{n}}$ *such that*

$$\mathrm{E} = \{\mathbf{A}\mathbf{x} + \mathbf{b} \mid \mathbf{x} \in \overline{\mathsf{B}}^{\mathrm{n}}(1, \mathbf{0})\};$$

(iii) there exists $\mathbf{x}_0 \in \mathbb{R}^{\mathrm{n}}$*,* $\lambda_1, \dots, \lambda_{\mathrm{n}} \in \mathbb{R}_{>0}$*, and an orthonormal basis* $\{\mathbf{f}_1, \dots, \mathbf{f}_{\mathrm{n}}\}$ *for* $\mathbb{R}^{\mathrm{n}}$ *such that*

$$\mathrm{E} = \{\mathrm{y}_1\mathbf{f}_1 + \dots + \mathrm{y}_{\mathrm{n}}\mathbf{f}_{\mathrm{n}} \in \mathbb{R}^{\mathrm{n}} \mid \lambda_1(\mathrm{y}_1 - \mathrm{y}_{01})^2 + \dots + \lambda_{\mathrm{n}}(\mathrm{y}_{\mathrm{n}} - \mathrm{y}_{0\mathrm{n}})^2 \leq 1\},$$

where

$$\mathbf{x}_0 = \mathrm{y}_1\mathbf{f}_1 + \dots + \mathrm{y}_{\mathrm{n}}\mathbf{f}_{\mathrm{n}};$$

(iv) there exists a symmetric and positive-definite map $\mathbf{B} \in \mathrm{L}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{n}})$ *and* $\mathbf{x}_0 \in \mathbb{R}^{\mathrm{n}}$ *such that*

$$\mathrm{E} = \{\mathbf{x} \in \mathbb{R}^{\mathrm{n}} \mid \langle \mathbf{B}(\mathbf{x} - \mathbf{x}_0), \mathbf{x} - \mathbf{x}_0 \rangle_{\mathbb{R}^{\mathrm{n}}} \leq 1\}.$$

what? *Moreover, if* $\mathbf{A}$ *and* $\mathbf{B}$ *are as in parts (ii) and (iv), respectively, then* $\mathbf{A} = \mathbf{B}^{-1/2}$ *(refer to for the notion of the square root of a symmetric positive-definite linear map).*

Proof (i) $\Longrightarrow$ (ii) By definition, $E = f'(x)$ where $f'(x) = A'x + b$ with A invertible. By the Polar Decomposition Theorem, , $A' = A \circ U$ where $U \in \mathsf{O}(n)$ and where A is symmetric and positive-definite. Then

what?

$$\begin{aligned} E &= \{A'x + b \mid x \in \overline{\mathsf{B}}^n(1, \mathbf{0})\} \\ &= \{AUx + b \mid x \in \overline{\mathsf{B}}^n(1, \mathbf{0})\} \\ &= \{Ax + b \mid x \in \overline{\mathsf{B}}^n(1, \mathbf{0})\}, \end{aligned}$$

noting that $U(\overline{\mathsf{B}}^n(1, \mathbf{0})) = \overline{\mathsf{B}}^n(1, \mathbf{0})$.

(ii) $\Longrightarrow$ (iii) By assumption, $E = f(\overline{\mathsf{B}}^n(1, \mathbf{0})$ where the affine map $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is given by $f(x) = Ax + b$ where $A \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ is symmetric and positive-definite and $b \in \mathbb{R}^n$. Thus, by , there exist eigenvalues $\mu'_1, \dots, \mu'_n \in \mathbb{R}_{>0}$ for A and a basis $\{f_1, \dots, f_n\}$ of orthonormal eigenvectors for A. Let $x_0 = A^{-1}b$ and write

what?

$$x_0 = y_{01}f_1 + \dots + y_{0n}f_n.$$

Also, if $x \in \mathbb{R}^n$, let us write

$$x = y_1f_1 + \dots + y_nf_n,$$

and note that
$$A(y_1 f_1 + \cdots + y_n f_n) = \mu'_1 y_1 f_1 + \cdots + \mu'_n y_n f_n.$$
Thus
$$A^{-1}(y_1 f_1 + \cdots + y_n f_n) = \mu_1 y_1 f_1 + \cdots + \mu_n y_n f_n,$$
where $\mu_j = \frac{1}{\mu'_j}$, $j \in \{1, \dots, n\}$. Thus, if $E = f(\overline{\mathsf{B}}^n(1, \mathbf{0}))$, we have

$$\begin{aligned}
E &= \{f(z) \mid z \in \overline{\mathsf{B}}^n(1,\mathbf{0})\} = \{Az + b \mid z \in \overline{\mathsf{B}}^n(1,\mathbf{0})\}\\
&= \{x \in \mathbb{R}^n \mid A^{-1}(x - b) \in \overline{\mathsf{B}}^n(1,\mathbf{0})\}\\
&= \{y_1 f_1 + \cdots + y_n f_n \mid \mu_1(y_1 - y_{01}) f_1 + \cdots + \mu_n(y_n - y_{0n}) f_n \in \overline{\mathsf{B}}^n(1,\mathbf{0})\}\\
&= \{y_1 f_1 + \cdots + y_n f_n \mid \lambda_1(y_1 - y_{01})^2 + \cdots + \lambda_n(y_n - y_{0n})^2 \le 1\}
\end{aligned}$$

where $\lambda_j = \mu_j^2$, $j \in \{1, \dots, n\}$, since the basis $\{f_1, \dots, f_n\}$ is orthonormal.

(iii) $\Longrightarrow$ (iv) With $\lambda_1, \dots, \lambda_n \in \mathbb{R}_{>0}$ and $\{f_1, \dots, f_n\}$ as in part (iii), define $B \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ by asking that $Bf_j = \lambda_j f_j$, $j \in \{1, \dots, n\}$, noting that this defines a linear map by Theorem I-4.5.24. Also let $x_0 \in \mathbb{R}^n$ be as in part (iii). If we write

$$x_0 = y_1 f_1 + \cdots + y_n f_n,$$

then the relation
$$\lambda_1(y_1 - y_{01})^2 + \cdots + \lambda_n(y_n - y_{0n})^2 \le 1$$
exactly corresponds to the relation $\langle B(x - x_0), x - x_0\rangle_{\mathbb{R}^n} \le 1$, if we write

$$x = y_1 f_1 + \cdots + y_n f_n.$$

(iv) $\Longrightarrow$ (i) We have

$$\begin{aligned}
\{x \in \mathbb{R}^n \mid \langle B(x - x_0), x - x_0\rangle_{\mathbb{R}^n} \le 1\} &= \{x \in \mathbb{R}^n \mid \langle B^{1/2}(x - x_0), B^{1/2}(x - x_0)\rangle_{\mathbb{R}^n} \le 1\}\\
&= \{x \in \mathbb{R}^n \mid B^{1/2}(x - x_0) \in \overline{\mathsf{B}}^n(1,\mathbf{0})\}\\
&= \{B^{-1/2} y + x_0 \mid y \in \overline{\mathsf{B}}^n(1,\mathbf{0})\},
\end{aligned}$$

showing that E is an ellipsoid if it satisfies condition (iv). This also proves the final assertion of the proposition. ■

5.1.2 Maximal and minimal ellipsoids

Given a convex body, one can obviously fit an ellipsoid inside the body; since a convex body has a nonempty interior, there exists a ball around every point contained in the convex body. One can then imagine growing an ellipsoid within the convex body. A question that arises is, “How large can an ellipsoid within a convex body be made?” The following result gives the somewhat unsurprising result that there is an ellipsoid of maximal volume, and the more surprising result that this maximal volume ellipsoid is unique.

5.1.6 Theorem (The maximal volume ellipsoid contained in a convex body) *Let* $\mathrm{C} \subseteq \mathbb{R}^{\mathrm{n}}$ *be a compact convex set with nonempty interior. Then the following statements hold:*

(i) there exists a unique ellipsoid $\mathrm{E}_{\mathrm{C}} \subseteq \mathrm{C}$ *such that, among all ellipsoids contained in* C*,* E_{C} *has the maximum volume;*

(ii) there exists $\mathrm{m} \in \mathbb{Z}_{>0}$*,* $\mathbf{u}_1, \dots, \mathbf{u}_{\mathrm{m}} \in \mathrm{bd}(\mathrm{C}) \cap \mathrm{bd}(\mathrm{E}_{\mathrm{C}})$*, and* $\lambda_1, \dots, \lambda_{\mathrm{m}} \in \mathbb{R}_{>0}$ *such that*

(a) $\displaystyle\sum_{\mathrm{j}=1}^{\mathrm{m}} \lambda_{\mathrm{j}} \mathbf{A}^{-1}(\mathbf{u}_{\mathrm{j}} - \mathbf{x}_0) = \mathbf{0}$ *and*

(b) $\displaystyle\sum_{\mathrm{j}=1}^{\mathrm{m}} \lambda_{\mathrm{j}} \langle \mathbf{A}^{-1}(\mathbf{u}_{\mathrm{j}} - \mathbf{x}_0), \mathbf{x}\rangle_{\mathbb{R}^{\mathrm{n}}}^2 = \|\mathbf{x}\|_{\mathbb{R}^{\mathrm{n}}}^2$ *for all* $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$*,*

where $\mathbf{A} \in \mathrm{L}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{n}})$ *and* $\mathbf{x}_0 \in \mathbb{R}^{\mathrm{n}}$ *are such that*

$$\mathrm{E}_{\mathrm{C}} = \{\mathbf{A}\mathbf{y} + \mathbf{x}_0 \mid \mathbf{y} \in \overline{\mathsf{B}}^{\mathrm{n}}(1, \mathbf{0})\};$$

(iii) if $\mathbf{x}_{\mathrm{C}}$ *is the centre of* E_{C} *(see Definition II-1.9.46), then*

$$\mathrm{C} \subseteq \mathbf{x}_{\mathrm{C}} + \mathrm{n}(\mathrm{E}_{\mathrm{C}} - \mathbf{x}_{\mathrm{C}});$$

(iv) if C *is balanced then* $\mathrm{C} \subseteq \sqrt{\mathrm{n}}\mathrm{C}$*.*

Proof (i) First let us prove the existence of E_C. Consider the map $\mathrm{vol}\colon \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n) \times \mathbb{R}^n \to \mathbb{R}$ defined by asking that $\mathrm{vol}(\boldsymbol{A}, \boldsymbol{b})$ is the volume of the ellipsoid

$$E_{\boldsymbol{A},\boldsymbol{b}} = \{\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \mid \boldsymbol{x} \in \overline{\mathsf{B}}^n(1, \mathbf{0}_{\mathbb{R}^n})\}.$$

Note that vol is a continuous map. We claim that the set

$$M_C = \{(\boldsymbol{A}, \boldsymbol{b}) \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n) \times \mathbb{R}^n \mid E_{\boldsymbol{A},\boldsymbol{b}} \subseteq C\}$$

is compact. That it is bounded follows from the fact that if $S \subseteq \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n) \times \mathbb{R}^n$ is an unbounded set, then given $R \in \mathbb{R}_{>0}$ there exists $\boldsymbol{x} \in \mathbb{R}^n$ such that $\|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|_{\mathbb{R}^n} > R$ (we leave the simple verification of this fact to the reader). Thus we need only show that the complement of M_C is open. Let $(\boldsymbol{A}, \boldsymbol{b}) \in (\mathrm{L}(\mathbb{R}^n; \mathbb{R}^n) \times \mathbb{R}^n) \setminus M_C$ and let $\boldsymbol{x} \in \overline{\mathsf{B}}^n(1, \mathbf{0}_{\mathbb{R}^n})$ be such that $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \notin C$. Since $\mathbb{R}^n \setminus C$ is open and since the map

$$(\boldsymbol{A}, \boldsymbol{b}) \mapsto \boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}$$

is continuous, there exists a neighbourhood $U \subseteq \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n) \times \mathbb{R}^n$ of $(\boldsymbol{A}, \boldsymbol{b})$ such that, if $(\boldsymbol{A}', \boldsymbol{b}') \in U$, then $\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b} \notin C$. This shows that M_C is closed.

Now, note that $\mathrm{vol}|M_C$ is a continuous map defined on a compact subset of a finite-dimensional normed vector space. Thus it has a maximum by Theorem II-1.3.32, and this shows that a maximal volume ellipsoid exists.

To prove the uniqueness part of the theorem is a little harder. First we prove a few lemmata. For these lemmata we recall the binomial coefficient

$$\binom{n}{m} = \frac{n!}{m!(n-m)!}.$$

With this notation we have the following lemma.

1 Lemma *If* $\mathrm{a}_1, \dots, \mathrm{a}_\mathrm{n} \in \mathbb{R}$ *then*

$$\prod_{\mathrm{k}=1}^{\mathrm{n}}(1+\mathrm{a}_\mathrm{k}) = 1 + \sum_{\mathrm{k}=1}^{\mathrm{n}} \sum_{\substack{\mathrm{j}_1,\dots,\mathrm{j}_\mathrm{k}\in\{1,\dots,\mathrm{n}\}\\ \mathrm{j}_1<\dots<\mathrm{j}_\mathrm{k}}} \mathrm{a}_{\mathrm{j}_1}\cdots \mathrm{a}_{\mathrm{j}_\mathrm{k}}$$

and, for each $\mathrm{k} \in \{1, \dots, \mathrm{n}\}$.

$$\prod_{\substack{\mathrm{j}_1,\dots,\mathrm{j}_\mathrm{k}\in\{1,\dots,\mathrm{n}\}\\ \mathrm{j}_1<\dots<\mathrm{j}_\mathrm{k}}} \mathrm{a}_{\mathrm{j}_1}\cdots \mathrm{a}_{\mathrm{j}_\mathrm{k}} = (\mathrm{a}_1\cdots\mathrm{a}_\mathrm{n})^{\binom{\mathrm{n}-1}{\mathrm{k}-1}}.$$

Proof Let us prove the first assertion first. We prove this by induction on n. For $n = 1$ the result is clearly true. So suppose the result is true for $n = m$ and let $a_1, \dots, a_{m+1} \in \mathbb{R}$. Then, using the induction hypothesis,

$$\begin{aligned}
\prod_{k=1}^{m+1}(1+a_k) &= \Big(\prod_{k=1}^{m}(1+a_k)\Big)(1+a_{m+1})\\
&= \Big(1+\sum_{k=1}^{m}\sum_{\substack{j_1,\dots,j_k\in\{1,\dots,m\}\\ j_1<\dots<j_k}} a_{j_1}\cdots a_{j_k}\Big)(1+a_{m+1})\\
&= 1+\sum_{k=1}^{m}\sum_{\substack{j_1,\dots,j_k\in\{1,\dots,m\}\\ j_1<\dots<j_k}} a_{j_1}\cdots a_{j_k} + \sum_{k=1}^{m}\sum_{\substack{j_1,\dots,j_k\in\{1,\dots,m\}\\ j_1<\dots<j_k}} a_{j_1}\cdots a_{j_k}a_{m+1} + a_{m+1}\\
&= 1+\sum_{k=1}^{m+1}\sum_{\substack{j_1,\dots,j_k\in\{1,\dots,m\}\\ j_1<\dots<j_k}} a_{j_1}\cdots a_{j_k},
\end{aligned}$$

giving the first assertion.

We prove the second assertion by induction on k, fixing n. For $k = 1$ we have

$$\{j_1, \dots, j_k \in \{1, \dots, n\} \mid j_1 < \dots < j_k\} = \{1, \dots, n\}.$$

Then

$$\prod_{\substack{j_1,\dots,j_k\in\{1,\dots,n\}\\ j_1<\dots<j_k}} = a_1\cdots a_n.$$

Since $\binom{n-1}{0} = 1$ the assertion follows in this case. Now suppose that the second assertion holds for $k = m < n$ and compute finish

$$\prod_{\substack{j_1,\dots,j_m,j_{m+1}\in\{1,\dots,n\}\\ j_1<\dots<j_k<j_{k+1}}} a_{j_1}\cdots a_{j_k}a_{j_{k+1}} =$$

▼

2 Lemma *For* $\mathrm{k}, \mathrm{n} \in \mathbb{Z}_{>0}$ *with* $\mathrm{k} \leq \mathrm{n}$*,*

$$\mathrm{card}(\{\mathrm{j}_1, \dots, \mathrm{j}_\mathrm{k} \in \{1, \dots, \mathrm{n}\} \mid \mathrm{j}_1 < \dots < \mathrm{j}_\mathrm{k}\}) = \binom{\mathrm{n}}{\mathrm{k}}.$$

Proof We fix n and prove the lemma by induction on k. For $k = 1$ we have

$$\{j_1, \dots, j_k \in \{1, \dots, n\} \mid j_1 < \dots < j_k\} = \{1, \dots, n\}$$

and $\binom{n}{k} = n$ and so the result holds in this case. Next suppose that the lemma holds for $k \in \{1, \dots, m\}$. Let $m \in \{1, \dots, n\}$ and note that

$$\mathrm{card}(\{j_1, \dots, j_k \in \{1, \dots, n\} \mid j_1 < \dots < j_k < m\}) = \mathrm{card}(\{j_1, \dots, j_k \in \{1, \dots, m-1\} \mid j_1 < \dots < j_k\}) = \binom{m-1}{k}.$$

Note that this equality still holds when $k \geq m - 1$, as long as we adopt the convention that $\binom{m-1}{k} = 0$ in these cases. Therefore, using Pascal's Rule (see the proof of Proposition I-3.2.11),

$$\begin{aligned} \mathrm{card}(\{j_1, \dots, j_k, j_{k+1} \in & \{1, \dots, n\} \mid j_1 \leq \dots \leq j_k \leq j_{k+1}\} \\ &= \sum_{m=1}^{n} \binom{m-1}{k} = \sum_{m=1}^{n} \binom{m}{k+1} - \sum_{m=1}^{n} \binom{m-1}{k+1} \\ &= \sum_{m=1}^{n} \binom{m}{k+1} - \sum_{m=0}^{n-1} \binom{m}{k+1} = \binom{n}{k+1} - \binom{0}{k+1} = \binom{n}{k+1} \end{aligned}$$

as desired. ▼

3 Lemma *If* $\mathbf{A}, \mathbf{B} \in \mathrm{L}(\mathbb{R}^\mathrm{n}; \mathbb{R}^\mathrm{n})$ *are symmetric and positive-definite, then*

$$\det(\mathbf{A} + \mathbf{B})^{1/\mathrm{n}} \geq \det(\mathbf{A})^{1/\mathrm{n}} + \det(\mathbf{B})^{1/\mathrm{n}}.$$

Moreover, if equality holds and if $\det(\mathbf{A}) = \det(\mathbf{B})$*, then* $\mathbf{A} = \mathbf{B}$.

Proof First let us prove the result assuming that $\boldsymbol{A} = \boldsymbol{I}_n$. In this case, we let $\lambda_1, \dots, \lambda_n \in$ what? $\mathbb{R}_{>0}$ denote the eigenvalues of $\boldsymbol{B}$ (these are real and positive by). For $k \in \{1, \dots, n\}$ let us define

$$\sigma_k = \sum_{\substack{j_1, \dots, j_k \in \{1, \dots, n\} \\ j_1 < \dots < j_k}} \lambda_{j_1} \cdots \lambda_{j_k}.$$

Note that, by Lemma 2, σ_k has $\binom{n}{k}$ summands. Thus, by Corollary I-3.1.36,

$$\sum_{\substack{j_1, \dots, j_k \in \{1, \dots, n\} \\ j_1 < \dots < j_k}} \frac{\lambda_{j_1} \cdots \lambda_{j_k}}{\binom{n}{k}} \geq \prod_{\substack{j_1, \dots, j_k \in \{1, \dots, n\} \\ j_1 < \dots < j_k}} (\lambda_{j_1} \cdots \lambda_{j_k})^{1/\binom{n}{k}}.$$

We then compute, using the preceding relation and the second part of Lemma 1,

$$\begin{aligned}
\sigma_k &= \binom{n}{k} \sum_{\substack{j_1,\dots,j_k\in\{1,\dots,n\}\\ j_1<\dots<j_k}} \frac{\lambda_{j_1}\cdots\lambda_{j_k}}{\binom{n}{k}} \\
&\geq \binom{n}{k} \prod_{\substack{j_1,\dots,j_k\in\{1,\dots,n\}\\ j_1<\dots<j_k}} (\lambda_{j_1}\cdots\lambda_{j_k})^{1/\binom{n}{k}} \\
&= \binom{n}{k}(\lambda_1\cdots\lambda_n)^{\binom{n-1}{k-1}/\binom{n}{k}} = \binom{n}{k}(\lambda_1\cdots\lambda_n)^{k/n}.
\end{aligned}$$

By Corollary I-3.1.36 we have equality above if and only if

$$\lambda_{j_1}\cdots\lambda_{j_k} = \lambda_{l_1}\cdots\lambda_{l_k}$$

for every $j_1,\dots,j_k,l_1,\dots,l_k \in \{1,\dots,n\}$ satisfying

$$j_1 < \dots < j_k, \quad l_1 < \dots < l_k.$$

In particular, if this holds for $k = 1$ we see that $\lambda_1 = \dots = \lambda_n$. By Lemma 1,

$$\prod_{k=1}^{n}(1+\lambda_k) = 1 + \sigma_1 + \dots + \sigma_n$$

and so, using the Binomial Theorem, Exercise I-2.2.1,

$$(1 + (\lambda_1\cdots\lambda_n)^{1/n})^n = \sum_{k=0}^{n}\binom{n}{k}(\lambda_1\cdots)^{k/n} \leq 1 + \sigma_1 + \dots + \sigma_n = \prod_{k=1}^{n}(1+\lambda_k).$$

Noting that

$$\det(\boldsymbol{I}_n)^{1/n} = 1, \quad \det(\boldsymbol{B})^{1/n} = (\lambda_1\cdots\lambda_n)^{1/n}, \quad \det(\boldsymbol{I}_n+\boldsymbol{B})^{1/n} = \Big(\prod_{k=1}^{n}(1+\lambda_k)\Big)^{1/n},$$

we see that the first assertion of the lemma holds when $\boldsymbol{A} = \boldsymbol{I}_n$. Moreover, equality holds in the above inequality if and only if all eigenvalues of $\boldsymbol{B}$ are equal. If $\det(\boldsymbol{I}_n) = \det(\boldsymbol{B})$ then it follows that $\boldsymbol{B} = \boldsymbol{I}_n$, and this gives the second assertion of the lemma when $\boldsymbol{A} = \boldsymbol{I}_n$.

For the general case, write $\boldsymbol{A} = \boldsymbol{S}^2$ where $\boldsymbol{S}$ is positive-definition, using . Then what?

$$\begin{aligned}
& \boldsymbol{A}+\boldsymbol{B} = \boldsymbol{P}^2 + \boldsymbol{B} = \boldsymbol{P}(\boldsymbol{I}_n + \boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1})\boldsymbol{P} \\
\implies \quad & \det(\boldsymbol{A}+\boldsymbol{B})^{1/n} = \det(\boldsymbol{P})^{2/n}\det(\boldsymbol{I}_n+\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1})^{1/n}
\end{aligned}$$

and

$$\begin{aligned}
\det(\boldsymbol{A})^{1/n} + \det(\boldsymbol{B})^{1/n} &= \det(\boldsymbol{P}^2) + \det(\boldsymbol{P}\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1}\boldsymbol{P})^{1/n} \\
&= \det(\boldsymbol{P})^{2/n}(1 + \det(\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1})^{1/n}),
\end{aligned}$$

using Theorem I-5.4.35. Then the first assertion of the lemma is equivalent to

$$\det(\boldsymbol{I}_n + \boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1})^{1/n} \geq 1 + \det(\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1})^{1/n}.$$

Since $\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1}$ is symmetric and positive-definite (as can be directly verified), this follows from the first part of the proof where we assumed that $\boldsymbol{A} = \boldsymbol{I}_n$. From the first part of the proof, if equality holds and if $\det(\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1}) = \det I_n = 1$ then $\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1} = \boldsymbol{I}_n$ and so $\boldsymbol{A} = \boldsymbol{B}$. Since

$$\det(\boldsymbol{P}^{-1}\boldsymbol{B}\boldsymbol{P}^{-1}) = \det(\boldsymbol{B})\det(\boldsymbol{P})^{-2} = \det(\boldsymbol{B})\det(\boldsymbol{A})^{-1}$$

we get the second assertion of the lemma in the general case. ▼

Now we complete the proof. Let S_n^+ denote the set of symmetric positive-definite matrices. One can easily directly verify that S_n^+ is convex. Let

$$\mathscr{E}_C = \{(\boldsymbol{A}, \boldsymbol{b}) \in \mathsf{S}_n^+ \times \mathbb{R}^n \mid \boldsymbol{A}(\overline{\mathsf{B}}^n(1, \boldsymbol{0})) + \boldsymbol{b} \subseteq C\}.$$

One can check directly that $\mathscr{E}_C$ is itself convex. Suppose that $E_1 = \boldsymbol{A}_1(\overline{\mathsf{B}}^n(1, \boldsymbol{0})) + \boldsymbol{b}_1$ and $E_2 = \boldsymbol{A}_2(\overline{\mathsf{B}}^n(1, \boldsymbol{0})) + \boldsymbol{b}_2$ are ellipsoids contained in C having maximal volume. Since the volume of E_1 and E_2 are $\det(\boldsymbol{A}_1)$ and $\det(\boldsymbol{A}_2)$, respectively, it follows that $\det(\boldsymbol{A}_1) = \det(\boldsymbol{A}_2)$. Let $\boldsymbol{A}_3 = \frac{1}{2}(\boldsymbol{A}_1 + \boldsymbol{A}_2)$ and $\boldsymbol{b}_3 = \frac{1}{2}(\boldsymbol{b}_1 + \boldsymbol{b}_2)$, noting that $(\boldsymbol{A}_3, \boldsymbol{b}_3) \in \mathscr{E}_C$ since $\mathscr{E}_C$ is convex. By Lemma 3 it holds that

$$\det(\boldsymbol{A}_3)^{1/n} \geq \tfrac{1}{2}(\det(\boldsymbol{A}_1)^{1/n} + \det(\boldsymbol{A}_2)^{1/n}) = \det(\boldsymbol{A}_1)^{1/n}.$$

Since E_1 has maximal volume, the inequalities must be equalities in the above expression, and so, by the second assertion of Lemma 3, $\boldsymbol{A}_1 = \boldsymbol{A}_2 = \boldsymbol{A}_3$.

It remains to show that $\boldsymbol{b}_1 = \boldsymbol{b}_2$. Suppose that $\boldsymbol{b}_1 \neq \boldsymbol{b}_2$ and let $\boldsymbol{b}_3 = \frac{1}{2}(\boldsymbol{b}_1 + \boldsymbol{b}_2)$. Let $K = \operatorname{conv}(E_1 \cup E_2)$. For a suitable affine transformation $\boldsymbol{f}$ we have $\boldsymbol{f}(E_1) = \overline{\mathsf{B}}^n(1, r\boldsymbol{e}_n)$ and $\boldsymbol{f}(E_2) = \overline{\mathsf{B}}^n(1, -r\boldsymbol{e}_n)$ for some $r \in \mathbb{R}_{>0}$. Then define E_3 by asking that

$$\boldsymbol{f}(E_3) = \{(y_1, \dots, y_n) \in \mathbb{R}^n \mid y_1^2 + \dots + y_{n-1}^2 + \tfrac{1}{r^2}y_n^2 \leq 1\}.$$

Note that $\overline{\mathsf{B}}^n(1, \boldsymbol{0}) \subset \boldsymbol{f}(E_3)$ and so the volume of $\boldsymbol{f}(E_3)$ is greater than that of $\overline{\mathsf{B}}^n(1, \boldsymbol{0})$. Since the latter volume agrees with that of $\boldsymbol{f}(E_1)$ and $\boldsymbol{f}(E_2)$, it follows that the volume of E_3 is greater than that of E_1 and E_2, contradicting the maximality of the latter two volumes.

(ii) First let us prove this part of the theorem assuming that $E_C = \overline{\mathsf{B}}^n(1, \boldsymbol{0})$. For $\boldsymbol{u} \in \mathbb{R}^n$ let us denote by $\boldsymbol{L}_{\boldsymbol{u}} \in \mathsf{L}(\mathbb{R}^n; \mathbb{R}^n)$ the linear map

$$\boldsymbol{L}_{\boldsymbol{u}}(\boldsymbol{x}) = \langle \boldsymbol{u}, \boldsymbol{x} \rangle_{\mathbb{R}^n} \boldsymbol{u}.$$

Note that $\operatorname{rank}(\boldsymbol{L}_{\boldsymbol{u}}) = 1$. Let us denote

$$K_C = \operatorname{conv}(\{(\boldsymbol{L}_{\boldsymbol{u}}, \boldsymbol{u}) \in \mathsf{L}(\mathbb{R}^n; \mathbb{R}^n) \oplus \mathbb{R}^n \mid \boldsymbol{u} \in \operatorname{bd}(C) \cap \operatorname{bd}(\overline{\mathsf{B}}^n(1, \boldsymbol{0}))\}.$$

Note that $\operatorname{bd}(C) \cap \operatorname{bd}(\overline{\mathsf{B}}^n(1, \boldsymbol{0})) \neq \emptyset$ since $\overline{\mathsf{B}}^n(1, \boldsymbol{0})$ is the maximal ellipsoid for C. As useful observation is that if $(\boldsymbol{A}, \boldsymbol{x}) \in K_C$ then $\operatorname{tr}((\boldsymbol{A}) = 1$. Indeed, note that, if $(\boldsymbol{L}_{\boldsymbol{u}}, \boldsymbol{u}) \in K_C$ for $\boldsymbol{u} \in \operatorname{bd}(C) \cap \operatorname{bd}(\overline{\mathsf{B}}^n(1, \boldsymbol{0}))$, then a direct computation gives

$$\operatorname{tr}(\boldsymbol{L}_{\boldsymbol{u}}) = \|\boldsymbol{u}\|_{\mathbb{R}^n}^2 = 1.$$

Now, if $(\boldsymbol{A}, \boldsymbol{x}) \in K_C$ then

$$(\boldsymbol{A}, \boldsymbol{x}) = \sum_{j=1}^{k} \lambda_j (\boldsymbol{L}_{\boldsymbol{u}_j}, \boldsymbol{u}_j) \quad \Longrightarrow \quad \boldsymbol{A} = \sum_{j=1}^{k} \lambda_j \boldsymbol{L}_{\boldsymbol{u}_j}$$

where $\lambda_j \in [0, 1]$, $j \in \{1, \dots, k\}$, and $\sum_{j=1}^{k} \lambda_j = 1$. Therefore,

$$\operatorname{tr}(\boldsymbol{A}) = \sum_{j=1}^{k} \lambda_j \operatorname{tr}(\boldsymbol{L}_{\boldsymbol{u}_j}) = 1$$

as claimed.

We shall first show that $(\frac{1}{n}\boldsymbol{I}_n, \boldsymbol{0}) \in K_C$. Suppose otherwise. Then, by Corollary II-1.9.17, let $\boldsymbol{\alpha}\colon \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n) \oplus \mathbb{R}^n \to \mathbb{R}$ and $a \in \mathbb{R}$ be such that

$$\boldsymbol{\alpha}(\tfrac{1}{n}\boldsymbol{I}_n, \boldsymbol{0}) < a, \quad \boldsymbol{\alpha}(\boldsymbol{A}, \boldsymbol{x}) > a, \qquad (\boldsymbol{A}, \boldsymbol{x}) \in K_C.$$

Note that there exists $\boldsymbol{B}_{\boldsymbol{\alpha}} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ and $\boldsymbol{y}_{\boldsymbol{\alpha}} \in \mathbb{R}^n$ such that such that

$$\boldsymbol{\alpha}(\boldsymbol{A}, \boldsymbol{x}) = \operatorname{tr}(\boldsymbol{B}_{\boldsymbol{\alpha}}^T \boldsymbol{A}) + \langle \boldsymbol{y}_{\boldsymbol{\alpha}}, \boldsymbol{x} \rangle_{\mathbb{R}^n}$$

for any $(\boldsymbol{A}, \boldsymbol{x}) \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ (since

$$((\boldsymbol{A}, \boldsymbol{x}), (\boldsymbol{B}, \boldsymbol{y})) \mapsto \operatorname{tr}(\boldsymbol{A}^T \boldsymbol{B}) + \langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathbb{R}^n}$$

is an inner product, cf.). Since $\frac{1}{n}\boldsymbol{I}_n$ and all matrices $\boldsymbol{A}$ such that $(\boldsymbol{A}, \boldsymbol{x}) \in K_C$ for some $\boldsymbol{x} \in \mathbb{R}^n$ are symmetric, we can, without loss of generality, assume that $\boldsymbol{B}_{\boldsymbol{\alpha}}$ is symmetric. what?
Moreover, if we take

$$\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}} = \boldsymbol{B}_{\boldsymbol{\alpha}} - \operatorname{tr}(\boldsymbol{B}_{\boldsymbol{\alpha}}) \boldsymbol{I}_n$$

and define

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{A}, \boldsymbol{x}) = \operatorname{tr}(\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}} \boldsymbol{A}) + \langle \boldsymbol{y}_{\boldsymbol{\alpha}}, \boldsymbol{x} \rangle_{\mathbb{R}^n} = \boldsymbol{\alpha}(\boldsymbol{A}, \boldsymbol{x}) - \operatorname{tr}(\boldsymbol{B}_{\boldsymbol{\alpha}})$$

(recalling that $\operatorname{tr}(\boldsymbol{A}) = 1$), then we have

$$\begin{aligned} &\hat{\boldsymbol{\alpha}}(\tfrac{1}{n}\boldsymbol{I}_n, \boldsymbol{0}) = 0 = \boldsymbol{\alpha}(\tfrac{1}{n}\boldsymbol{I}_n, \boldsymbol{0}) - \operatorname{tr}(\boldsymbol{B}_{\boldsymbol{\alpha}}) < a - \operatorname{tr}(\boldsymbol{B}_{\boldsymbol{\alpha}}), \\ &\hat{\boldsymbol{\alpha}}(\boldsymbol{A}, \boldsymbol{x}) = \boldsymbol{\alpha}(\boldsymbol{A}, \boldsymbol{x}) - \operatorname{tr}(\boldsymbol{B}_{\boldsymbol{\alpha}}) > a - \operatorname{tr}(\boldsymbol{B}_{\boldsymbol{\alpha}}), \qquad (\boldsymbol{A}, \boldsymbol{x}) \in K_C. \end{aligned}$$

Therefore, for $(\boldsymbol{A}, \boldsymbol{x}) \in K_C$, we have

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{A}, \boldsymbol{x}) > a - \operatorname{tr}(\boldsymbol{B}_{\boldsymbol{\alpha}}) > 0.$$

Since

$$\operatorname{tr}(\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}} \boldsymbol{L}_{\boldsymbol{u}}) = \langle \hat{\boldsymbol{B}}_{\boldsymbol{\alpha}} \boldsymbol{u}, \boldsymbol{u} \rangle_{\mathbb{R}^n}$$

for every $\boldsymbol{u} \in \mathbb{R}^n$ (this is directly verifiable), it follows that

$$\langle \hat{\boldsymbol{B}}_{\boldsymbol{\alpha}} \boldsymbol{u}, \boldsymbol{u} \rangle_{\mathbb{R}^n} + \langle \boldsymbol{y}_{\boldsymbol{\alpha}}, \boldsymbol{u} \rangle_{\mathbb{R}^n} > 0 \tag{5.1}$$

for every $\boldsymbol{u} \in \operatorname{bd}(C) \cap \operatorname{bd}(\overline{\mathsf{B}}^n(1, \boldsymbol{0}))$.

Let $\delta \in \mathbb{R}_{>0}$ and let $\delta_0 \in \mathbb{R}_{>0}$ be sufficiently small that $\boldsymbol{I}_n + \delta\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}}$ is positive-definite for every $\delta \in [0, \delta_0]$. (This is possible since $\boldsymbol{I}_n + \delta\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}}$ is positive-definite for $\delta = 0$ and since the condition for positive-definiteness from Theorem I-5.6.19 are continuous in δ.) We shall always assume that $\delta \in (0, \delta_0]$. Then define

$$\boldsymbol{x}_\delta = -\frac{1}{2}(\boldsymbol{I}_n + \delta\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}})^{-1}\boldsymbol{y}_{\boldsymbol{\alpha}}$$

and

$$E_\delta = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle(\boldsymbol{I}_n + \delta\boldsymbol{B})(\boldsymbol{x} - \boldsymbol{x}_\delta), \boldsymbol{x} - \boldsymbol{x}_\delta\rangle_{\mathbb{R}^n} \leq 1\}.$$

Note that, given the definition of $\boldsymbol{x}_\delta$,

$$\langle(\boldsymbol{I}_n + \delta\boldsymbol{B})(\boldsymbol{x} - \boldsymbol{x}_\delta), \boldsymbol{x} - \boldsymbol{x}_\delta\rangle_{\mathbb{R}^n} = \|\boldsymbol{x}\|^2_{\mathbb{R}^n} + \delta\langle\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}}\boldsymbol{x}, \boldsymbol{x}\rangle_{\mathbb{R}^n} + \langle\boldsymbol{y}_{\boldsymbol{\alpha}}, \boldsymbol{x}\rangle_{\mathbb{R}^n}. \tag{5.2}$$

We claim that there exists $\delta \in \mathbb{R}_{>0}$ sufficiently small that $E_\delta \subset C$. Let $\boldsymbol{x} \in \mathrm{bd}(C) - \mathrm{bd}(\overline{\mathsf{B}}^n(1, \boldsymbol{0}))$. By continuity of the function

$$(\delta, \boldsymbol{y}) \mapsto \langle(\boldsymbol{I}_n + \delta\boldsymbol{B})(\boldsymbol{y} - \boldsymbol{x}_\delta), \boldsymbol{y} - \boldsymbol{x}_\delta\rangle_{\mathbb{R}^n}$$

and since the value of this function is greater than 1 at $\boldsymbol{x}$ when $\delta = 0$, it follows that there exists $\delta_{\boldsymbol{x}} \in \mathbb{R}_{>0}$ and a neighbourhood $U_{\boldsymbol{x}}$ of $\boldsymbol{x}$ such that

$$\langle(\boldsymbol{I}_n + \delta\boldsymbol{B})(\boldsymbol{y} - \boldsymbol{x}_\delta), \boldsymbol{y} - \boldsymbol{x}_\delta\rangle_{\mathbb{R}^n} > 1$$

for all $\delta \in (0, \delta_{\boldsymbol{x}}]$ and all $\boldsymbol{y} \in U_{\boldsymbol{x}}$. Now, if $\boldsymbol{x} \in \mathrm{bd}(C) \cap \mathrm{bd}(\overline{\mathsf{B}}^n(1, \boldsymbol{0}))$ we use (5.1) and (5.2) to get

$$\langle(\boldsymbol{I}_n + \delta\boldsymbol{B})(\boldsymbol{x} - \boldsymbol{x}_\delta), \boldsymbol{x} - \boldsymbol{x}_\delta\rangle_{\mathbb{R}^n} > 1$$

for $\delta \in (0, \delta_0]$. Thus there exists $\delta_{\boldsymbol{x}} \in \mathbb{R}_{>0}$ and a neighbourhood $U_{\boldsymbol{x}}$ of $\boldsymbol{x}$ such that

$$\langle(\boldsymbol{I}_n + \delta\boldsymbol{B})(\boldsymbol{y} - \boldsymbol{x}_\delta), \boldsymbol{y} - \boldsymbol{x}_\delta\rangle_{\mathbb{R}^n} > 1$$

for all $\delta \in (0, \delta_{\boldsymbol{x}}]$ and all $\boldsymbol{y} \in U_{\boldsymbol{x}}$. Now note that $(U_{\boldsymbol{x}})_{\boldsymbol{x}\in\mathrm{bd}(C)}$ covers $\mathrm{bd}(C)$ and so, by compactness of $\mathrm{bd}(C)$, there exists $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathrm{bd}(C)$ such that $\mathrm{bd}(C) \subseteq \cup_{j=1}^k U_{\boldsymbol{x}_j}$. If

$$\delta = \min\{\delta_{\boldsymbol{x}_1}, \dots, \delta_{\boldsymbol{x}_k}\}$$

then we have

$$\langle(\boldsymbol{I}_n + \delta\boldsymbol{B})(\boldsymbol{x} - \boldsymbol{x}_\delta), \boldsymbol{x} - \boldsymbol{x}_\delta\rangle_{\mathbb{R}^n} > 1$$

for all $\boldsymbol{x} \in \mathrm{bd}(C)$. Thus $E_\delta \subseteq C$ as claimed.

Next we claim that $\mathrm{vol}(E_\delta) \geq \mathrm{vol}(\overline{\mathsf{B}}^n(1, \boldsymbol{0}))$. Note that

$$E_\delta = \{(\boldsymbol{I}_n + \delta\hat{\boldsymbol{B}}_{\boldsymbol{\alpha}})^{-1/2}\boldsymbol{x} + \boldsymbol{x}_0 \mid \boldsymbol{x} \in \overline{\mathsf{B}}^n(1, \boldsymbol{0})\}$$

by Proposition 5.1.5. Let $\mu_1, \dots, \mu_n \in \mathbb{R}_{>0}$ be the eigenvalues for $\boldsymbol{I}_n + \delta\boldsymbol{B}$. By the change of variable theorem,

$$\mathrm{vol}(E_\delta) = \frac{\mathrm{vol}(\overline{\mathsf{B}}^n(1, \boldsymbol{0}))}{\det(\boldsymbol{I}_n + \delta\boldsymbol{B})^{1/2}} = \frac{\mathrm{vol}(\overline{\mathsf{B}}^n(1, \boldsymbol{0}))}{(\prod_{j=1}^n \mu_j)^{1/2}}.$$

Now note that $\sum_{j=1}^{n} \mu_j = \mathrm{tr}(\boldsymbol{I}_n + \delta\boldsymbol{B}) = n$. Then, using Corollary I-3.1.36,

$$\prod_{j=1}^{n} \mu_j \leq \Big(\frac{1}{n}\sum_{j=1}^{n} \mu_j\Big)^n = 1,$$

giving $\mathrm{vol}(E_\delta) \geq \mathrm{vol}(\overline{\mathsf{B}}^n(1,\boldsymbol{0}))$. Since $E_\delta \subset C$ there exists r slightly larger than 1 such that the ellipsoid $rE_\delta \subseteq C$ (do you spot the hidden use of compactness here?). Since

$$\mathrm{vol}(rE_\delta) > \mathrm{vol}(E_\delta) \geq \mathrm{vol}(\overline{\mathsf{B}}^n(1,\boldsymbol{0})),$$

we contradict the fact that $\overline{\mathsf{B}}^n(1,\boldsymbol{0})$ is the unique ellipsoid of maximal volume for C.

Thus our assumption that $(\frac{1}{n}\boldsymbol{I}_n, \boldsymbol{0}) \notin K_C$ is false and so, by Proposition II-1.9.4,

$$\frac{1}{n}\boldsymbol{I}_n = \sum_{j=1}^{m} c_j \boldsymbol{L}_{\boldsymbol{u}_j}, \quad \boldsymbol{0} = \sum_{j=1}^{n} c_j \boldsymbol{u}_j$$

for $c_j \in \mathbb{R}_{>0}$ and $\boldsymbol{u}_j \in \mathrm{bd}(C) \cap \mathrm{bd}(\overline{\mathsf{B}}^n(1,\boldsymbol{0}))$, $j \in \{1,\dots,m\}$. Let us define $\lambda_j = nc_j$, $j \in \{1,\dots,n\}$, and note that

$$\boldsymbol{I}_n = \sum_{j=1}^{m} \lambda_j \boldsymbol{L}_{\boldsymbol{u}_j}, \quad \boldsymbol{0} = \sum_{j=1}^{n} \lambda_j \boldsymbol{u}_j.$$

Thus

$$\sum_{j=1}^{m} \lambda_j \langle \boldsymbol{u}_j, \boldsymbol{x}\rangle_{\mathbb{R}^n} \boldsymbol{u}_j = \boldsymbol{x}$$

for every $\boldsymbol{x} \in \mathbb{R}^n$, and taking the inner product of this expression with $\boldsymbol{y} \in \mathbb{R}^n$ gives

$$\sum_{j=1}^{m} \lambda_j \langle \boldsymbol{u}_j, \boldsymbol{x}\rangle_{\mathbb{R}^n} \langle \boldsymbol{u}_j, \boldsymbol{y}\rangle_{\mathbb{R}^n} = \langle \boldsymbol{x}, \boldsymbol{y}\rangle_{\mathbb{R}^n}.$$

Letting $\boldsymbol{y} = \boldsymbol{x}$ gives this part of the theorem in the case that $E_C = \overline{\mathsf{B}}^n(1,\boldsymbol{0})$.

In the general case, write $E_C = \boldsymbol{f}(\overline{\mathsf{B}}^n(1,\boldsymbol{0}))$ for an invertible affine map $\boldsymbol{f}$. The above proof then shows that there are points $\boldsymbol{u}'_j \in \mathrm{bd}(\boldsymbol{f}^{-1}(C)) \cap \mathrm{bd}(\overline{\mathsf{B}}^n(1,\boldsymbol{0}))$ and $\lambda_j \in \mathbb{R}_{>0}$, $j \in \{1,\dots,m\}$, such that

$$\|\boldsymbol{x}\|^2_{\mathbb{R}^n} = \sum_{j=1}^{m} \lambda_j \langle \boldsymbol{u}'_j, \boldsymbol{x}\rangle_{\mathbb{R}^n}, \quad \boldsymbol{0} = \sum_{j=1}^{n} \lambda_j \boldsymbol{u}'_j$$

for every $\boldsymbol{x} \in \mathbb{R}^n$. Defining $\boldsymbol{u}_j = \boldsymbol{f}(\boldsymbol{u}'_j)$ we get the desired result.

(iii) We begin by assuming that $E_C = \overline{\mathsf{B}}^n(1,\boldsymbol{0})$. Using part (ii), let $\boldsymbol{u}_1,\dots,\boldsymbol{u}_m \in \mathrm{bd}(C) \cap \mathrm{bd}(\overline{\mathsf{B}}^n(1,\boldsymbol{0}))$ and $\lambda_1,\dots,\lambda_m \in \mathbb{R}_{>0}$ be such that

$$\sum_{j=1}^{m} \lambda_j \boldsymbol{L}_{\boldsymbol{u}_j} = \boldsymbol{I}_n.$$

Define

$$K = \{x \in \mathbb{R}^n \mid \langle u_j, x\rangle_{\mathbb{R}^n} \le 1,\ j \in \{1,\dots,m\}\}.$$

Note that K is convex (it is a convex polyhedron according to Definition II-1.9.49). We claim that $C \subseteq K$. To see this, note that for each $j \in \{1,\dots,m\}$ the hyperplane

$$\mathsf{P}_j = \{x \in \mathbb{R}^n \mid \langle u_j, x\rangle_{\mathbb{R}^n} = 1\}$$

is a support hyperplane for K passing through u_j. Moreover, it is also a support hyperplane for $\overline{\mathsf{B}}^n(1,\mathbf{0})$ passing through u_j; in fact, it is the unique support hyperplane for $\overline{\mathsf{B}}^n(1,\mathbf{0})$ passing through u_j by . Thus P_j is a support hyperplane for C passing through u_j. Since the hyperplanes P_j, $j \in \{1,\dots,m\}$, are the only supporting hyperplanes for K, we conclude that $C \subseteq K$ as claimed.

what?

Now let $x \in K$ and note that

$$-\|x\|_{\mathbb{R}^n} \le \langle u_j, x\rangle_{\mathbb{R}^n} \le 1,\ j \in \{1,\dots,m\},$$

the first inequality by Cauchy–Bunyakovsky–Schwartz and the second by definition of K. Also note that, as we saw in the proof of part (ii),

$$I_n = \sum_{j=1}^m \lambda_j L_{u_j}.$$

Taking traces and noting that $\operatorname{tr}(\mathsf{L}_{u_j}) = 1$, we have

$$\sum_{j=1}^m \lambda_j = n.$$

Therefore,

$$\begin{aligned}0 &\le \sum_{j=1}^m \lambda_j(1 - \langle u_j, x\rangle_{\mathbb{R}^n})(\|x\|_{\mathbb{R}^n} + \langle u_j, x\rangle_{\mathbb{R}^n})\\ &= \|x\|_{\mathbb{R}^n}\sum_{j=1}^n \lambda_j + (1 - \|x\|_{\mathbb{R}^n})\sum_{j=1}^m \lambda_j\langle u_j, x\rangle_{\mathbb{R}^n} - \sum_{j=1}^m \lambda_j \langle u_j, x\rangle_{\mathbb{R}^n}^2\\ &= n\|x\|_{\mathbb{R}^n} - \|x\|_{\mathbb{R}^n}^2.\end{aligned}$$

Thus $\|x\|_{\mathbb{R}^n} \le n$, giving this part of the result when $E_C = \overline{\mathsf{B}}^n(1,\mathbf{0})$ since $x \in K \supseteq C$.

In the general case, write

$$E_C = \{Ay + x_C \mid y \in \overline{\mathsf{B}}^n(1,\mathbf{0})\}.$$

Let $f(x) = Ax + x_C$. The first part of the proof above then gives $f^{-1}(C) \subseteq n\overline{\mathsf{B}}^n(1,\mathbf{0})$. Thus

$$C \subseteq \{Ay + x_C \mid y \in n\overline{\mathsf{B}}^n(1,\mathbf{0})\} = x_C + n(E_C - x_C),$$

as desired.

(iv) Suppose that C is balanced and let $\boldsymbol{u}'_1,\dots,\boldsymbol{u}'_m \in \mathrm{bd}(C)\cap\mathrm{bd}(\overline{\mathsf{B}}^n(1,\boldsymbol{0}))$ and $\lambda'_1,\dots,\lambda'_m \in \mathbb{R}_{>0}$ satisfy

$$\sum_{j=1}^{m} \lambda'_j \langle \boldsymbol{u}'_j, \boldsymbol{x}\rangle^2_{\mathbb{R}^n} = \|\boldsymbol{x}\|^2_{\mathbb{R}^n}.$$

Then define

$$\boldsymbol{u}_{2j-1} = \boldsymbol{u}'_j,\quad \boldsymbol{u}_{2j} = -\boldsymbol{u}'_j,\qquad j\in\{1,\dots,m\},$$

noting that $\boldsymbol{u}_k \in \mathrm{bd}(C)\cap\mathrm{bd}(\overline{\mathsf{B}}^n(1,\boldsymbol{0}))$, $k\in\{1,\dots,2m\}$, since C and $\overline{\mathsf{B}}^n(1,\boldsymbol{0})$ are balanced. Also define

$$\lambda_{2j-1} = \lambda_{2j} = \frac{1}{2}\lambda'_j,\qquad j\in\{1,\dots,m\}.$$

One then directly verifies that

$$\sum_{k=1}^{2m} \lambda_k \boldsymbol{u}_k = \boldsymbol{0},\quad \sum_{k=1}^{2m} \lambda_k \langle \boldsymbol{u}_k, \boldsymbol{x}\rangle^2_{\mathbb{R}^n} = \|\boldsymbol{x}\|^2_{\mathbb{R}^n},\qquad \boldsymbol{x}\in\mathbb{R}^n.$$

Now let

$$K = \{\boldsymbol{x}\in\mathbb{R}^n \mid \langle \boldsymbol{u}_k,\boldsymbol{x}\rangle_{\mathbb{R}^n} \le 1,\ k\in\{1,\dots,2m\}\}.$$

As in the previous part of the proof, $C\subseteq K$ and, if $\boldsymbol{x}\in K$ then

$$-\|\boldsymbol{x}\|_{\mathbb{R}^n} \le \langle \boldsymbol{u}_k,\boldsymbol{x}\rangle_{\mathbb{R}^n} \le 1,\qquad k\in\{1,\dots,2m\}.$$

Since if $\boldsymbol{u}\in\{\boldsymbol{u}_k \mid k\in\{1,\dots,2m\}\}$ it also holds that $-\boldsymbol{u}\in\{\boldsymbol{u}_k \mid k\in\{1,\dots,2m\}\}$, we also have

$$-1 \le \langle \boldsymbol{u}_k,\boldsymbol{x}\rangle_{\mathbb{R}^n} \le \|\boldsymbol{x}\|_{\mathbb{R}^n},\qquad k\in\{1,\dots,2m\}.$$

Thus $\langle \boldsymbol{u}_k,\boldsymbol{x}\rangle_{\mathbb{R}^n}\in[-1,1]$ for $k\in\{1,\dots,2m\}$ and $\boldsymbol{x}\in K$. As in the previous part of the proof, we have

$$\sum_{k=1}^{2m}\lambda_k = n.$$

Therefore, for $\boldsymbol{x}\in K$,

$$\|\boldsymbol{x}\|^2_{\mathbb{R}^n} = \sum_{k=1}^{2m}\lambda_k\langle \boldsymbol{u}_k,\boldsymbol{x}\rangle^2_{\mathbb{R}^n} \le n,$$

and so $\|\boldsymbol{x}\|_{\mathbb{R}^n} \le \sqrt{n}$. This gives this part of the theorem when $E_C = \overline{\mathsf{B}}^n(1,\boldsymbol{0})$.

In the general case, write $E_C = \boldsymbol{L}(\overline{\mathsf{B}}^n(1,\boldsymbol{0}))$ for an invertible linear (instead of linear since the centre of E_C is $\boldsymbol{0}$ if C is balanced) map $\boldsymbol{L}$. Then the preceding part of the proof gives $\boldsymbol{L}^{-1}(C)\subseteq\sqrt{n}\overline{\mathsf{B}}^n(1,\boldsymbol{0})$ which immediately gives $C\subseteq\sqrt{n}E_C$ by linearity of $\boldsymbol{L}$. ■

Let us give some examples that illustrate the preceding theorem, and also show that the estimates given in parts (iii) and (iv) are sharp.

5.1.7 Examples (Maximal ellipsoid of a convex body) Both examples will be convex polytopes, as per Definition II-1.9.49. It is useful, therefore, to consider the following general result. We let $S_{>0}(\mathbb{R}^n;\mathbb{R}^n)$ be the set of symmetric, positive-definite linear maps on $\mathbb{R}^n$. is this the right notation?

1 **Lemma** *Let* C *be a convex polytope defined by*

$$\mathsf{C} = \{\mathbf{x} \in \mathbb{R}^n \mid \langle \mathbf{a}_j, \mathbf{x}\rangle_{\mathbb{R}^n} \le b_j,\ j \in \{1, \dots, k\}\}$$

for $\mathbf{a}_j \in \mathbb{R}^n$ *and* $\mathbf{b} \in \mathbb{R}^k$, $j \in \{1, \dots, k\}$, *and let* E *be the ellipsoid*

$$\mathsf{E}_0 = \{\mathbf{B}_0\mathbf{x} + \mathbf{d}_0 \mid \mathbf{x} \in \overline{\mathsf{B}}^n(1, \mathbf{0})\}$$

for a symmetric, positive-definite $\mathbf{B_0} \in \mathsf{S}_{>0}(\mathbb{R}^n; \mathbb{R}^n)$ *and* $\mathbf{d}_0 \in \mathbb{R}^n$. *Define* $f\colon \mathsf{S}_{>0}(\mathbb{R}^n; \mathbb{R}^n) \oplus \mathbb{R}^n \to \mathbb{R}$ *and* $\mathbf{h}\colon \mathsf{S}_{>0}(\mathbb{R}^n; \mathbb{R}^n) \oplus \mathbb{R} \to \mathbb{R}^k$ *by*

$$f(\mathbf{B}, \mathbf{d}) = \log\det\mathbf{B}, \quad h_j(\mathbf{B}, \mathbf{x}) = \|\mathbf{B}\mathbf{a}_j\|_{\mathbb{R}^n} + \langle \mathbf{a}_j, \mathbf{d}\rangle_{\mathbb{R}^n}, \qquad j \in \{1, \dots, k\}.$$

Then E_0 *is the maximal volume ellipsoid for* C *if and only if* $(\mathbf{C}, \mathbf{x}_0)$ *is a local maximum of* $(f, \mathbf{h})$ *with inequality constraints (see Definition II-1.4.43).*

Proof For $(\boldsymbol{B}, \boldsymbol{d}) \in S_{>0}(\mathbb{R}^n; \mathbb{R}^n)$ define the ellipsoid

$$E_{\boldsymbol{B},\boldsymbol{d}} = \{\boldsymbol{B}\boldsymbol{x} + \boldsymbol{d} \mid \boldsymbol{x} \in \overline{\mathsf{B}}^n(1, \mathbf{0})\}.$$

Then

$$\begin{aligned}
& E_{\boldsymbol{B},\boldsymbol{d}} \subseteq C \\
\iff\quad & \langle \boldsymbol{a}_j, \boldsymbol{B}\boldsymbol{x}\rangle_{\mathbb{R}^n} + \langle \boldsymbol{a}_j, \boldsymbol{d}\rangle_{\mathbb{R}^n} \le b_j, \qquad j \in \{1, \dots, k\},\ \boldsymbol{x} \in \overline{\mathsf{B}}^n(1, \mathbf{0}) \\
\iff\quad & \sup\{\langle \boldsymbol{a}_j, \boldsymbol{B}\boldsymbol{x}\rangle_{\mathbb{R}^n} + \langle \boldsymbol{a}_j, \boldsymbol{d}\rangle_{\mathbb{R}^n} \mid \boldsymbol{x} \in \overline{\mathsf{B}}^n(1, \mathbf{0})\} \le b_j, \qquad j \in \{1, \dots, k\} \\
\iff\quad & \sup\{|\langle \boldsymbol{B}\boldsymbol{a}_j, \boldsymbol{x}\rangle_{\mathbb{R}^n}| \mid \boldsymbol{x} \in \overline{\mathsf{B}}^n(1, \mathbf{0})\} + \langle \boldsymbol{a}_j, \boldsymbol{d}\rangle_{\mathbb{R}^n} \le b_j, \qquad j \in \{1, \dots, k\} \\
\iff\quad & \|\boldsymbol{B}\boldsymbol{a}_j\|_{\mathbb{R}^n} + \langle \boldsymbol{a}_j, \boldsymbol{d}\rangle_{\mathbb{R}^n} \le b_j, \qquad j \in \{1, \dots, k\},
\end{aligned}$$

since $\boldsymbol{B}$ is symmetric, and using Proposition I-2.2.27 and the fact that the induced norm of the linear map $\boldsymbol{x} \mapsto \langle \boldsymbol{c}, \boldsymbol{x}\rangle_{\mathbb{R}^n}$ is $\|\boldsymbol{c}\|_{\mathbb{R}^n}$ according to Theorem II-1.1.14. Thus the set of ellipsoids contained in C is prescribed by the inequality constraints defined by $\boldsymbol{h}$. Since $\operatorname{vol}(E_{\boldsymbol{B},\boldsymbol{d}}) = \det\boldsymbol{B}\operatorname{vol}(\overline{\mathsf{B}}^n(1, \mathbf{0}))$ by the change of variables theorem, the maximum volume ellipsoid is the maximum of the function $(\boldsymbol{B}, \boldsymbol{d}) \mapsto \det\boldsymbol{B}$ subject to the inequality constraints defined by $\boldsymbol{h}$. Since log is monotonically increasing, the lemma follows, cf. Proposition I-2.2.27. ▼

Note that maximising $\boldsymbol{B} \mapsto \log\det\boldsymbol{B}$ is nothing more than maximising $\boldsymbol{B} \mapsto \det\boldsymbol{B}$. However, there are advantages to using log det rather than det. One of these is the simple form for the derivative of the function log det.

2 **Lemma** *For* $\mathbf{B}, \mathbf{V} \in \mathsf{L}(\mathbb{R}^n; \mathbb{R}^n)$ *with* $\mathbf{B}$ *invertible, the derivative of* $\log\det$ *at* $\mathbf{B}$ *in the direction of* $\mathbf{V}$ *is* $\mathbf{D}\log\det(\mathbf{B}) \cdot \mathbf{V} = \operatorname{tr}(\mathbf{B}^{-1}\mathbf{V})$.

Proof By Theorem I-5.3.10,

$$(\det\boldsymbol{B})\boldsymbol{I}_n = \boldsymbol{B}\operatorname{Adj}(\boldsymbol{B}).$$

In this formula, we shall think of $\boldsymbol{B} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n)$ as being a variable, and shall write the components of $\boldsymbol{B}$ as $\boldsymbol{B}_{jk}$, $j,k \in \{1,\dots,n\}$. Let $j,k \in \{1,\dots,n\}$ and note that the above formula gives

$$\det \boldsymbol{B} = \sum_{m=1}^{n} \boldsymbol{B}_{jm}\mathrm{Adj}(\boldsymbol{B})_{mj}.$$

Note that $\mathrm{Adj}(\boldsymbol{B})_{mj}$ is, possibly up to sign, the determinant of the $(n-1)\times(n-1)$ matrix obtained by deleting the jth row and mth column of $\boldsymbol{A}$. This implies, in particular, that $\mathrm{Adj}(\boldsymbol{B})_{mj}$ is independent of $\boldsymbol{B}_{jk}$ for every $m \in \{1,\dots,n\}$. Thus

$$\frac{\partial \det \boldsymbol{B}}{\partial \boldsymbol{B}_{jk}} = \sum_{m=1}^{n} \frac{\partial \boldsymbol{B}_{jm}}{\partial \boldsymbol{B}_{jk}}\mathrm{Adj}(\boldsymbol{B})_{mj} = \mathrm{Adj}(\boldsymbol{B})_{kj}.$$

Thus, by Proposition I-3.8.6(iii) and the Fundamental Theorem of Calculus,

$$\frac{\partial \log \det \boldsymbol{B}}{\partial \boldsymbol{B}_{jk}} = \frac{1}{\det \boldsymbol{B}}\mathrm{Adj}(\boldsymbol{B})_{kj} = (\boldsymbol{B}^{-T})_{jk}.$$

Thus

$$\boldsymbol{D}\log\det(\boldsymbol{B})\cdot\boldsymbol{V} = \sum_{j,k=1}^{n} \frac{\partial \log\det \boldsymbol{B}}{\partial \boldsymbol{B}_{jk}}\boldsymbol{V}_{jk} = \sum_{j,k=1}^{n} \boldsymbol{B}^{-1}_{kj}\boldsymbol{V}_{jk} = \mathrm{tr}(\boldsymbol{B}^{-1}\boldsymbol{V}),$$

as claimed. ▼

Let us now consider a few special cases of the preceding lemma.

1. Let us take $C = [-1,1]^n$ to be the unit cube in $\mathbb{R}^n$.

3 Lemma $E_C = \overline{\mathsf{B}}^n(1,\boldsymbol{0})$.

Proof Let $\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}$ be the standard basis for $\mathbb{R}^n$ and define $\boldsymbol{a}_1,\dots,\boldsymbol{a}_{2n} \in \mathbb{R}^n$ by

$$\boldsymbol{a}_{2j-1} = \boldsymbol{e}_j, \quad \boldsymbol{a}_{2j} = -\boldsymbol{e}_j, \qquad j \in \{1,\dots,n\}.$$

Note that

$$C = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{a}_k, \boldsymbol{x}\rangle_{\mathbb{R}^n} \le 1, k \in \{1,\dots,2n\}\}.$$

Note that, for $\boldsymbol{B} \in S_{>0}(\mathbb{R}^n;\mathbb{R}^n)$, we have $\boldsymbol{B}\boldsymbol{e}_j = \boldsymbol{c}(\boldsymbol{B},j)$ for each $j \in \{1,\dots,n\}$. Let E_C be the ellipsoid of maximal volume contained in C. Note that since C is balanced, E_C is also balanced (why?). Thus, according to Lemma 1 above, $E_C = \boldsymbol{B}_0(\overline{\mathsf{B}}^n(1,\boldsymbol{0}))$ where $\boldsymbol{B}_0$ minimises $\boldsymbol{B} \mapsto \log\det\boldsymbol{B}$ subject to the constraints $\|\boldsymbol{c}(\boldsymbol{B},j)\|_{\mathbb{R}^n} \le 1$, $j \in \{1,\dots,n\}$.

$$\boldsymbol{D}\log\det(\boldsymbol{I}_n)\cdot(\boldsymbol{B}-\boldsymbol{I}_n) = \mathrm{tr}(\boldsymbol{I}_n(\boldsymbol{B}-\boldsymbol{I}_n)) = \mathrm{tr}(\boldsymbol{B}-\boldsymbol{I}_n) \le 0$$

▼

5.1.3 Notes

Theorem 5.1.6 is often called the ***John Ellipsoid Theorem***, and was proved by John [1948]. John actually proves more, giving some measure of how much of the convex body is filled by the maximal ellipsoid.

Now note that, by Proposition II-1.1.11(iv),

$$\inf\{r \in \mathbb{R}_{>0} \mid C \subseteq r\overline{\mathsf{B}}^n(1, \mathbf{0})\} = \sqrt{n}.$$

Thus, in this case, $C \subseteq \sqrt{n}\overline{\mathsf{B}}^n(1, \mathbf{0})$, and $\sqrt{n}$ is the smallest factor for which this inclusion holds.

2.

Section 5.2

Dvoretzky's Theorem

In this section we present an important theorem on the nature of high-dimensional convex bodies. This theorem is valuable for understanding the so-called local geometry of Banach spaces. Indeed, we have already used Dvoretzky's Theorem in a crucial way in the proof of Theorem 3.7.5. The proof of Dvoretzky's Theorem is rather nonobvious, and has a probabilistic flavour. For this reason, our discussion of Dvoretzky's Theorem begins with a discussion of the so-called "concentration" of Gaussian measure. This subject itself is one of great independent interest.

Do I need to read this section? This section can be passed over unless one needs to use it. However, if one is interested in gaining some insights into the structure of infinite-dimensional Banach spaces, Dvoretzky's Theorem is indispensable. •

5.2.1 Concentration of Gaussian measure

Our proof of Dvoretzky's Theorem relies on a trick from measure theory known as concentration of measure. This is a rather general subject area, and we shall only touch upon a very specific facet of this, the concentration of Gaussian measure.

We work with $\mathbb{R}^n$ and on the σ-algebra $\mathscr{B}(\mathbb{R}^n)$ of Borel sets we define a measure, called the ***standard Gaussian measure***, γ^n by

$$\gamma^n(B) = \frac{1}{(\sqrt{2\pi})^n}\int_B \exp(-\tfrac{1}{2}\|\boldsymbol{x}\|^2_{\mathbb{R}^n})\,\mathrm{d}\boldsymbol{x},$$

where we use Riemann integral notation, but actually use the Lebesgue integral on $\mathbb{R}^n$. Let us state some useful facts about this measure.

5.2.1 Lemma (Properties of standard Gaussian measure) *The standard Gaussian measure has the following properties:*

(i) γ^n *is the* n*-fold product of the measures* γ^1*:*

$$\gamma^n = \gamma^1 \times \cdots \times \gamma^1;$$

(ii) $\gamma^n(\mathbb{R}^n) = 1$*;*

(iii) *if* $\mathbf{R} \in \mathsf{O}(n)$ *and if* $B \in \mathscr{B}(\mathbb{R}^n)$*, then* $\gamma^n(\mathbf{R}(B)) = \gamma^n(B)$.

Proof (i) Let $B_1, \dots, B_n \in \mathscr{B}(\mathbb{R})$ so that

$$B \triangleq B_1 \times \cdots \times B_n \in \mathscr{B}(\mathbb{R}^n)$$

by Proposition 2.5.7. We compute

$$\begin{aligned}\gamma^n(B) &= \frac{1}{(\sqrt{2\pi})^n}\int_B \exp(-\tfrac{1}{2}\|\boldsymbol{x}\|^2_{\mathbb{R}^n})\,\mathrm{d}\boldsymbol{x} = \frac{1}{(\sqrt{2\pi})^n}\int_{B_1\times\cdots\times B_n}\exp(-\tfrac{1}{2}x_1^2)\cdots\exp(-\tfrac{1}{2}x_n^2)\,\mathrm{d}\boldsymbol{x}\\ &= \Big(\frac{1}{\sqrt{2\pi}}\int_{B_1}\exp(-\tfrac{1}{2}x_1^2)\,\mathrm{d}x_1\Big)\cdots\Big(\frac{1}{\sqrt{2\pi}}\int_{B_n}\exp(-\tfrac{1}{2}x_n^2)\,\mathrm{d}x_n\Big) = \gamma^1(B_1)\cdots\gamma^1(B_n),\end{aligned}$$

giving this part of the result by Theorem 2.3.33.

(ii) By Lemma 1 from Example 2.3.32–4 and a change of variable we have $\gamma^1(\mathbb{R}) = 1$. This part of the lemma now follows from part (i).

(iii) We compute

$$\begin{aligned}\gamma^n(\boldsymbol{R}(B)) &= \int_{\boldsymbol{R}(B)}\exp(-\tfrac{1}{2}\|\boldsymbol{x}\|^2_{\mathbb{R}^n})\,\mathrm{d}\boldsymbol{x} = \int_B \exp(-\tfrac{1}{2}\|\boldsymbol{R}^{-1}\boldsymbol{y}\|^2_{\mathbb{R}^n})|\det\boldsymbol{R}|\,\mathrm{d}\boldsymbol{y}\\ &= \int_B \exp(-\tfrac{1}{2}\|\boldsymbol{y}\|^2_{\mathbb{R}^n})\,\mathrm{d}\boldsymbol{y} = \gamma^n(B),\end{aligned}$$

using the change of variables theorem and Theorem II-1.3.18. ■

Let us denote the integral of a $\mathscr{B}(\mathbb{R}^n)$-measurable function $f\colon \mathbb{R}^n \to \mathbb{R}$ by

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\,\mathrm{d}\gamma^n(\boldsymbol{x}),$$

noting by Proposition 2.7.65 that

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\exp(-\tfrac{1}{2}\|\boldsymbol{x}\|^2_{\mathbb{R}^n})\,\mathrm{d}\boldsymbol{x}.$$

The following estimate will be useful for us.

5.2.2 Lemma (Gaussian mean of supremum) *There exists* $C \in \mathbb{R}_{>0}$ *such that*

$$\int_{\mathbb{R}^n}\|\mathbf{x}\|_\infty\,\mathrm{d}\gamma^n(\mathbf{x}) \geq C\sqrt{\log(n)}$$

for every $n \in \mathbb{Z}_{>0}$.

Proof By the change of variable formula, $t \mapsto -\mathrm{e}^{frac12t^2}$ is a primitive of $t\mathrm{e}^{-\frac{1}{2}t^2}$. Let $\alpha \in \mathbb{R}_{>0}$ and compute, using this fact and integration by parts,

$$\int_\alpha^\infty \mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{d}t = \int_\alpha^\infty \frac{t\mathrm{e}^{-\frac{1}{2}t^2}}{t}\,\mathrm{d}t = \frac{\mathrm{e}^{-\frac{1}{2}\alpha^2}}{\alpha} - \int_\alpha^\infty \frac{\mathrm{e}^{-\frac{1}{2}t^2}}{t^2}\,\mathrm{d}t.$$

Now we compute

$$\int_\alpha^\infty \frac{\mathrm{e}^{-\frac{1}{2}t^2}}{t^2}\,\mathrm{d}t \leq \frac{1}{\alpha^3}\int_\alpha^\infty t\mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{d}t = \frac{\mathrm{e}^{-\frac{1}{2}\alpha^2}}{\alpha^3}.$$

Therefore,

$$\int_\alpha^\infty \mathrm{e}^{-\frac{1}{2}t^2}\,\mathrm{d}t \geq \Big(1 - \frac{1}{\alpha^2}\Big)\frac{\mathrm{e}^{-\frac{1}{2}\alpha^2}}{\alpha}.$$

make sure this is the right log

Now let $\alpha = \alpha_n \triangleq \sqrt{\log(n)}$ and note that, with this value of α,

$$\int_{\alpha_n}^{\infty} e^{-\frac{1}{2}t^2}\,dt \geq \frac{1 - \frac{1}{\log(n)}}{\sqrt{n\log(n)}}.$$

Note that

$$\lim_{n\to\infty} n\frac{1 - \frac{1}{\log(n)}}{\sqrt{n\log(n)}} = \infty.$$

Thus we let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\frac{1 - \frac{1}{\log(n)}}{\sqrt{n\log(n)}} \geq \frac{\sqrt{2\pi}}{2n}, \qquad n \geq N.$$

Thus

$$\frac{2}{\sqrt{2\pi}}\int_{\alpha_n}^{\infty} e^{-\frac{1}{2}t^2}\,dt \geq \frac{1}{n}, \qquad n \geq N.$$

Note that

$$\{x \in \mathbb{R}^n \mid \|x\|_\infty \leq \alpha_n\}$$

is the cube of radius α_n centred at $\mathbf{0}_{\mathbb{R}^n}$. Thus

$$\begin{aligned}\gamma^n(\{x \in \mathbb{R}^n \mid \|x\|_\infty \leq \alpha_n\}) &= \Big(\frac{1}{\sqrt{2\pi}}\int_{-\alpha_n}^{\alpha_n} e^{-\frac{1}{2}t^2}\,dt\Big)^n \\ &= \Big(\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{1}{2}t^2}\,dt - \frac{2}{2\sqrt{\pi}}\int_{\alpha_n}^{\infty} e^{-\frac{1}{2}t^2}\,dt\Big)^n \leq \Big(1 - \frac{1}{n}\Big)^n.\end{aligned}$$

Note that

$$\Big(1 - \frac{1}{n}\Big)^n = \Big(\frac{n-1}{n}\Big)^n = \Big(\frac{n}{n-1}\Big)^{-n} = \Big(\frac{n-1+1}{n-1}\Big)^{-n} = \Big(1 + \frac{1}{n-1}\Big)^{-n}.$$

Let $m = n - 1$ and note that

$$\Big(1 - \frac{1}{n}\Big)^n = \Big(1 + \frac{1}{m}\Big)^{-m-1} = \frac{1}{(1+\frac{1}{m})^m(1+\frac{1}{m})}.$$

Thus

$$\lim_{n\to\infty}\Big(1 - \frac{1}{n}\Big)^n = \lim_{m\to\infty}\frac{1}{(1+\frac{1}{m})^m(1+\frac{1}{m})} = \frac{1}{e},$$

by Proposition I-2.4.18, using Proposition I-2.3.23. Thus we assume that N is also sufficiently large that

$$\gamma^n(\{x \in \mathbb{R}^n \mid \|x\|_\infty \leq \alpha_n\}) \leq \frac{1}{2}, \qquad n \geq N.$$

Let

$$A_n = \{x \in \mathbb{R}^n \mid \|x\|_\infty \geq \alpha_n\}$$

so that $\gamma^n(A_n) \geq \frac{1}{2}$ when $n \geq N$. Then

$$\int_{\mathbb{R}^n} \|x\|_\infty \mathrm{d}\gamma^n(x) = \int_{A_n} \|x\| \, \mathrm{d}\gamma^n(x) + \int_{\mathbb{R}^n \setminus A_n} \|x\| \, \mathrm{d}\gamma^n(x) \geq \alpha_n \gamma^n(A_n) \geq \frac{1}{2}\sqrt{\log(n)},$$

provided that $n \geq N$. Letting

$$C = \min\Big\{\frac{1}{2}\Big\} \cup \left\{ \frac{\int_{\mathbb{R}^n} \|x\|_\infty \, \mathrm{d}\gamma^n(x)}{\sqrt{\log(n)}} \;\middle|\; n \in \{1, \dots, N\} \right\},$$

we have the estimate as given. ■

Now let us state the result we will use on the so-called concentration of Gaussian measure.

5.2.3 Theorem (Concentration of standard Gaussian measure) *If* $r \in \mathbb{R}_{>0}$ *and if* $f \colon \mathbb{R}^n \to \mathbb{R}$ *satisfies*

$$|f(\mathbf{x}_1) - f(\mathbf{x}_2)| \leq \|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbb{R}^n}$$

for every $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$, *then*

$$\gamma^n\Big(\Big\{\mathbf{x} \in \mathbb{R}^n \;\Big|\; \Big|f(\mathbf{x}) - \int_{\mathbb{R}^n} f(\mathbf{y}) \mathrm{d}\gamma^n(\mathbf{y})\Big| \geq r\Big\}\Big) \leq 2e^{-\frac{2}{\pi^2} r^2}.$$

Proof We first show that f is integrable so that the statement in the theorem makes sense.

1 Lemma $f \in \mathsf{L}^{(1)}((\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \gamma^n); \mathbb{R})$.

Proof Note that for every $x \in \mathbb{R}^n$ we have

$$\begin{aligned} & |f(x) - f(\mathbf{0})| \leq \|x\|_{\mathbb{R}^n} \\ \implies \quad & f(\mathbf{0}) - \|x\|_{\mathbb{R}^n} \leq f(x) \leq f(0) + \|x\|_{\mathbb{R}^n} \\ \implies \quad & |f(x)| \leq |f(\mathbf{0})| + \|x\|_{\mathbb{R}^n}. \end{aligned}$$

Therefore,

$$|f(x)\exp(-\tfrac{1}{2}\|x\|^2_{\mathbb{R}^n})| \leq |f(0)|\exp(-\tfrac{1}{2}\|x\|^2_{\mathbb{R}^n}) + \|x\|_{\mathbb{R}^n}\exp(-\tfrac{1}{2}\|x\|^2_{\mathbb{R}^n}).$$

Thus our claim will follow if we can show that

$$x \mapsto \|x\|_{\mathbb{R}^n}\exp(-\tfrac{1}{2}\|x\|^2_{\mathbb{R}^n})$$

is integrable with respect to Lebesgue measure. Note that one can use L'Hôpital's Rule to prove that

$$\lim_{r\to\infty} r^k \exp(-\tfrac{1}{2}r^2) = 0$$

for every $k \in \mathbb{Z}_{>0}$. Thus let $R \in \mathbb{R}_{>0}$ be sufficiently large that $r(1 + r^2)^n \exp(-\frac{1}{2}r^2) \leq 1$ for $r \geq R$. Then we compute

$$\begin{aligned} \int_{\mathbb{R}^n} \|x\|_{\mathbb{R}^n}\exp(-\tfrac{1}{2}\|x\|^2_{\mathbb{R}^n})\,\mathrm{d}x = & \int_{\overline{\mathsf{B}}^n(R,\mathbf{0})} \|x\|_{\mathbb{R}^n}\exp(-\tfrac{1}{2}\|x\|^2_{\mathbb{R}^n})\,\mathrm{d}x \\ & + \int_{\mathbb{R}^n\setminus\overline{\mathsf{B}}^n(R,\mathbf{0})} \|x\|_{\mathbb{R}^n}\exp(-\tfrac{1}{2}\|x\|^2_{\mathbb{R}^n})\,\mathrm{d}x. \end{aligned}$$

The first integral on the right is that of a bounded function on a compact set, and so is finite. As for the second integral on the right, we have

$$\begin{aligned}\int_{\mathbb{R}^n\setminus\overline{\mathsf{B}}^n(R,\mathbf{0})} \|\boldsymbol{x}\|_{\mathbb{R}^n} \exp(-\tfrac{1}{2}\|\boldsymbol{x}\|^2_{\mathbb{R}^n})\,\mathrm{d}\boldsymbol{x} &\le \int_{\mathbb{R}^n} \frac{1}{(1+\|\boldsymbol{x}\|^2)^n}\,\mathrm{d}\boldsymbol{x} \\ &\le \Big(\int_{\mathbb{R}} \frac{1}{1+|x_1|^2}\,\mathrm{d}x_1\Big)\cdots\Big(\int_{\mathbb{R}} \frac{1}{1+|x_n|^2}\,\mathrm{d}x_n\Big) = \pi^n,\end{aligned}$$

where we have used Fubini's Theorem and the fact that

$$\frac{\mathrm{d}}{\mathrm{d}x}\tan^{-1}(x) = \frac{1}{1+x^2},$$

cf. the proof of Theorem I-3.8.18. ▼

Note that f is Lipschitz with Lipschitz constant 1. By Rademacher's Theorem, Theorem II-1.10.60, it follows that f is almost everywhere differentiable and that, at points $\boldsymbol{x}$ where f is differentiable, $\|\boldsymbol{D}f(\boldsymbol{x})\|_{\mathbb{R}^n,\mathbb{R}} \le 1$. As can be see from Theorem II-1.1.14,

$$\|\boldsymbol{D}f(\boldsymbol{x})\|_{\mathbb{R}^n,\mathbb{R}} = \Big(\sum_{j=1}^n |\boldsymbol{D}_j f(\boldsymbol{x})|^2\Big)^{1/2}. \tag{5.3}$$

In the computations below, we shall only be integrating expressions involving the derivative of f, so we shall suppose that $\boldsymbol{D}f$ is everywhere defined by taking it to be $\mathbf{0}$ at points where f is not differentiable.

Now let us assume that

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\,\mathrm{d}\gamma^n(\boldsymbol{x}) = 0, \tag{5.4}$$

i.e., that $\mathrm{mean}(f) = 0$, and prove that

$$\gamma^n(\{\boldsymbol{x}\in\mathbb{R}^n \mid f(\boldsymbol{x}) \ge r\}) \le \mathrm{e}^{-\frac{2}{\pi^2}r^2}. \tag{5.5}$$

Note that, by the Chernoff inequality (Corollary 2.7.37) we have

$$\gamma^n(\{\boldsymbol{x}\in\mathbb{R}^n \mid f(\boldsymbol{x}) \ge r\}) \le \mathrm{e}^{-cr}\int_{\mathbb{R}^n} \exp(cf(\boldsymbol{x}))\,\mathrm{d}\gamma^n(\boldsymbol{x}) \tag{5.6}$$

for every $c \in \mathbb{R}_{>0}$. Let us estimate the integral in the preceding expression.

2 Lemma *If* f *is Lipschitz with Lipschitz constant* 1*, if* $\mathrm{mean(f)} = 0$*, and if* $\mathrm{c}\in\mathbb{R}_{>0}$*, then*

$$\int_{\mathbb{R}^n} \exp(\mathrm{cf}(\mathbf{x}))\,\mathrm{d}\gamma^n(\mathbf{x}) \le \exp\Big(\frac{\mathrm{c}^2\pi^2}{8}\Big).$$

Proof Note that the second derivative of $x \mapsto \exp(-cx)$ is $c^2\exp(-cx)$, which is positive. Therefore, by Proposition I-3.2.30(iii) we conclude that $x \mapsto \exp(-cx)$ is convex. Therefore, by Jensen's inequality, Theorem 2.7.31, we have

$$\int_{\mathbb{R}^n} \exp(-cf(\boldsymbol{y}))\,\mathrm{d}\gamma^n(\boldsymbol{y}) \ge \exp\Big(-c\int_{\mathbb{R}^n} f(\boldsymbol{y})\,\mathrm{d}\gamma^n(\boldsymbol{y})\Big) = 1,$$

the latter equality holding since we are assuming that $\int_{\mathbb{R}^n} f(\boldsymbol{y})\,\mathrm{d}\gamma^n = 0$. Therefore,

$$\begin{aligned} & \exp(cf(\boldsymbol{x})) \le \exp(cf(\boldsymbol{x})) \int_{\mathbb{R}^n} \exp(-cf(\boldsymbol{y}))\,\mathrm{d}\gamma^n(\boldsymbol{y}) \\ \Longrightarrow \quad & \int_{\mathbb{R}^n} \exp(cf(\boldsymbol{x}))\mathrm{d}\gamma^n \le \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \exp(c(f(\boldsymbol{x}) - f(\boldsymbol{y})))\,\mathrm{d}\gamma^n(\boldsymbol{x})\mathrm{d}\gamma^n(\boldsymbol{y}) \end{aligned} \tag{5.7}$$

by Proposition 2.7.19 and Fubini's Theorem.

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$ define a curve $\boldsymbol{\sigma}_{\boldsymbol{x},\boldsymbol{y}} \colon [0, \frac{\pi}{2}] \to \mathbb{R}^n \oplus \mathbb{R}^n$ by

$$\boldsymbol{\sigma}_{\boldsymbol{x},\boldsymbol{y}}(\theta) = (\cos\theta\boldsymbol{y}, \sin\theta\boldsymbol{x}).$$

Let $\boldsymbol{p} \colon \mathbb{R}^n \oplus \mathbb{R}^n \to \mathbb{R}^n$ be the map $\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x} + \boldsymbol{y}$. Note that

$$\boldsymbol{p}^* f \circ \boldsymbol{\sigma}_{\boldsymbol{x},\boldsymbol{y}}(\theta) = f(\cos\theta\boldsymbol{y} + \sin\theta\boldsymbol{x}).$$

Thus, for fixed $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ we have, by the Fundamental Theorem of Calculus,

$$\int_0^{\frac{\pi}{2}} \frac{\mathrm{d}}{\mathrm{d}\theta}\boldsymbol{p}^* f \circ \boldsymbol{\sigma}_{\boldsymbol{x},\boldsymbol{y}}(\theta)\,\mathrm{d}\theta = \boldsymbol{p}^* f \circ \boldsymbol{\sigma}_{\boldsymbol{x},\boldsymbol{y}}(\tfrac{\pi}{2}) - \boldsymbol{p}^* f \circ \boldsymbol{\sigma}_{\boldsymbol{x},\boldsymbol{y}}(0) = f(\boldsymbol{y}) - f(\boldsymbol{x}).$$

By Corollary 2.7.32 of Jensen's inequality,

$$\begin{aligned} \exp(c(f(\boldsymbol{y}) - f(\boldsymbol{x}))) &\le \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \exp\Big(\frac{c\pi}{2}\frac{\mathrm{d}}{\mathrm{d}\theta}p^* f \circ \boldsymbol{R}_\theta(\boldsymbol{x}, \boldsymbol{y})\Big)\,\mathrm{d}\theta \\ &= \frac{2}{\pi}\int_0^{\frac{\pi}{2}} \exp\Big(\frac{c\pi}{2}\boldsymbol{D}f(\cos\theta\boldsymbol{y} + \sin\theta\boldsymbol{x}) \cdot (-\sin\theta\boldsymbol{y} + \cos\theta\boldsymbol{x})\Big)\,\mathrm{d}\theta. \end{aligned}$$

Let us define a linear transformation $\boldsymbol{R}_\theta$ of $\mathbb{R}^n \oplus \mathbb{R}^n$ by

$$\boldsymbol{R}_\theta(\boldsymbol{x}, \boldsymbol{y}) = (\cos\theta\boldsymbol{y} + \sin\theta\boldsymbol{x}, -\sin\theta\boldsymbol{y} + \cos\theta\boldsymbol{x}).$$

One immediately computes

$$\|\boldsymbol{R}_\theta(\boldsymbol{x}, \boldsymbol{y})\|_{\mathbb{R}^{2n}} = \|(\boldsymbol{x}, \boldsymbol{y})\|_{\mathbb{R}^{2n}},$$

and so $\boldsymbol{R}_\theta \in \mathsf{O}(2n)$ by Theorem II-1.3.18. Thus, by the change of variable theorem and Fubini's Theorem, denoting $(\boldsymbol{\xi}, \boldsymbol{\eta}) = \boldsymbol{R}_\theta(\boldsymbol{x}, \boldsymbol{y})$,

$$\begin{aligned} &\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \exp(c(f(\boldsymbol{y}) - f(\boldsymbol{x})))\,\mathrm{d}\gamma^n(\boldsymbol{x})\mathrm{d}\gamma^n(\boldsymbol{y}) \\ &\qquad \le \frac{2}{\pi}\int_0^{\frac{\pi}{2}}\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \exp\Big(\frac{c\pi}{2}\boldsymbol{D}f(\boldsymbol{\xi}) \cdot \boldsymbol{\eta}\Big)\,\mathrm{d}\gamma^n(\boldsymbol{\xi})\mathrm{d}\gamma^n(\boldsymbol{\eta})\mathrm{d}\theta. \end{aligned} \tag{5.8}$$

Let us evaluate the integral

$$\int_{\mathbb{R}^n} \exp\Big(\frac{c\pi}{2}\boldsymbol{D}f(\boldsymbol{\xi}) \cdot \boldsymbol{\eta}\Big)\,\mathrm{d}\gamma^n(\boldsymbol{\eta}) = \frac{1}{(\sqrt{2\pi})^n}\int_{\mathbb{R}^n} \exp\Big(\frac{c\pi}{2}\boldsymbol{D}f(\boldsymbol{\xi}) \cdot \boldsymbol{\eta} - \frac{1}{2}\|\boldsymbol{\eta}\|^2_{\mathbb{R}^n}\Big)\,\mathrm{d}\boldsymbol{\eta}$$

with $\boldsymbol{\xi} \in \mathbb{R}^n$ fixed. Let us abbreviate $a_j = \frac{c\pi}{2}\boldsymbol{D}_j f(\boldsymbol{\xi})$. Then

$$\int_{\mathbb{R}^n} \exp(\langle \boldsymbol{a}, \boldsymbol{\eta}\rangle_{\mathbb{R}^n} - \tfrac{1}{2}\|\boldsymbol{\eta}\|^2_{\mathbb{R}^n})\,\mathrm{d}\boldsymbol{\eta} = \Big(\int_{\mathbb{R}} \exp(a_1\eta_1 - \tfrac{1}{2}\eta_1^2)\,\mathrm{d}\eta_1\Big)\cdots\Big(\int_{\mathbb{R}} \exp(a_n\eta_n - \tfrac{1}{2}\eta_n^2)\,\mathrm{d}\eta_n\Big)$$

by Fubini's Theorem. Noting that

$$-\tfrac{1}{2}\eta_j^2 + a_j\eta_j = -\tfrac{1}{2}(\eta_j - a_j)^2 + \tfrac{1}{2}a_j^2,$$

we have

$$\int_{\mathbb{R}} \exp(a_j\eta_j - \tfrac{1}{2}\eta_j^2)\,\mathrm{d}\eta_j = \exp(\tfrac{1}{2}a_j^2)\int_{\mathbb{R}} \exp(-\tfrac{1}{2}(\eta_j - a_j)^2)\,\mathrm{d}\eta_j = \sqrt{2\pi}\exp(\tfrac{1}{2}a_j^2)$$

for every $j \in \{1,\dots,n\}$, using Lemma 1 from Example 2.3.32–4. Note that we are assuming that $\|\boldsymbol{D}f(\boldsymbol{\xi})\|_{\mathbb{R}^n,\mathbb{R}} \le 1$ and so, using (5.3),

$$\int_{\mathbb{R}^n} \exp\Big(\frac{c\pi}{2}\boldsymbol{D}f(\boldsymbol{\xi})\cdot\boldsymbol{\eta}\Big)\,\mathrm{d}\gamma^n(\boldsymbol{\eta}) = \exp\Big(\frac{1}{2}\Big(\frac{c\pi}{2}\Big)^2\|\boldsymbol{D}f(\boldsymbol{\xi})\|^2_{\mathbb{R}^n,\mathbb{R}}\Big) \le \exp\Big(\frac{c^2\pi^2}{8}\Big). \tag{5.9}$$

Combining (5.7), (5.8) (and integrating the right-hand side with respect to θ and $\boldsymbol{\xi}$ in this equation), and (5.9) gives

$$\int_{\mathbb{R}^n} \exp(cf(\boldsymbol{x}))\mathrm{d}\gamma^n \le \exp\Big(\frac{c^2\pi^2}{8}\Big),$$

which is the lemma. ▼

Combining (5.6) with the lemma we have that

$$\gamma^n(\{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \ge r\}) \le \exp\Big(-rc + \frac{c^2\pi^2}{8}\Big)$$

for every $c \in \mathbb{R}_{>0}$. The expression on the right achieves its minimum when the argument of the exponential achieves its minimum, i.e., when $c = \frac{4r}{\pi^2}$. With this value of c we have

$$\gamma^n(\{\boldsymbol{x} \in \mathbb{R}^n \mid f(\boldsymbol{x}) \ge r\}) \le \exp\Big(-\frac{2}{\pi^2}r^2\Big)$$

This gives the estimate (5.5) in the case when (5.4) holds. The same analysis, applied to $-f$, gives

$$\gamma^n\Big(\Big\{\boldsymbol{x} \in \mathbb{R}^n \;\Big|\; f(\boldsymbol{x}) - \int_{\mathbb{R}^n} f(\boldsymbol{x})\mathrm{d}\gamma^n \le -r\Big\}\Big) \le \exp\Big(-\frac{2}{\pi^2}r^2\Big),$$

under the assumption that (5.4) holds. Thus the theorem holds when (5.4) holds.

In case (5.4) does not hold, then the function $\boldsymbol{x} \mapsto f(\boldsymbol{x}) - \mathrm{mean}(f)$ satisfies (5.4) and so

$$\gamma^n\Big(\Big\{\boldsymbol{x} \in \mathbb{R}^n \;\Big|\; \Big|(f(\boldsymbol{x}) - \mathrm{mean}(f)) - \int_{\mathbb{R}^n}(f(\boldsymbol{x}) - \mathrm{mean}(f))\mathrm{d}\gamma^n\Big| \ge r\Big\}\Big) \le 2\exp\Big(-\frac{2}{\pi^2}r^2\Big).$$

Since

$$\int_{\mathbb{R}^n} \operatorname{mean}(f)\,\mathrm{d}\gamma^n = \operatorname{mean}(f)$$

by Lemma 5.2.1(ii), we get the desired conclusion. ■

One should understand the theorem as indicating the extend to which the values of a Lipschitz function are concentrated around its mean.

5.2.2 Dvoretzky's Theorem

In this section we present an important theorem regarding the nature of convex bodies in sufficiently high-dimensional Euclidean spaces. As we shall see, this is related to the character of the norms on subspaces of sufficiently high-dimensional Banach spaces. In particular, it is related to the question of whether there are subspaces of Banach spaces for which the unit ball is "nearly round." Let us put this into some context with an example.

5.2.4 Example (Balls for the ∞-norm on $\mathbb{R}^n$) Note that the unit ball in $\mathbb{R}^n$ with respect to the ∞-norm

$$\|\boldsymbol{x}\|_\infty = \max\{|x_1|, \dots, |x_n|\}$$

is the unit cube $[-1, 1]^n$, and this cube is not very round. We shall show that as n gets large, there is a two-dimensional subspace of $\mathbb{R}^n$ for which the intersection of the unit cube with this subspace gets more and more round.

For $n \geq 2$ we define

$$\boldsymbol{v}_1 = (1, \cos(\tfrac{\pi}{n}), \cos(\tfrac{2\pi}{n}), \dots, \cos(\tfrac{(n-1)\pi}{n}))$$
$$\boldsymbol{v}_2 = (0, \sin(\tfrac{\pi}{n}), \sin(\tfrac{2\pi}{n}), \dots, \sin(\tfrac{(n-1)\pi}{n}))$$

and $\boldsymbol{u}_j = \frac{\boldsymbol{v}_j}{\|\boldsymbol{v}_j\|_{\mathbb{R}^n}}$, $j \in \{1, 2\}$. We claim that $\langle \boldsymbol{u}_1, \boldsymbol{u}_2\rangle_{\mathbb{R}^n} = 0$. To see this, note that $\mathrm{e}^{\frac{2\pi\mathrm{i}}{n}}$ is a primitive nth root of unity. Therefore, by Proposition II-3.2.6,

$$\sum_{j=0}^{n-1} \mathrm{e}^{2\pi\mathrm{i}\frac{j}{n}} = 0 \quad \implies \sum_{j=0}^{n-1} \sin(2\pi\tfrac{j}{n}) = 0,$$

the latter equality following by taking the imaginary part of the former. Thus

$$\langle \boldsymbol{v}_1, \boldsymbol{v}_2\rangle_{\mathbb{R}^n} = \sum_{j=0}^{n-1} \cos(\pi\tfrac{j}{n})\sin(\pi\tfrac{j}{n}) = \frac{1}{2}\sum_{j=0}^{n-1} \sin(2\pi\tfrac{j}{n}) = 0,$$

using the identity $2\cos\theta\sin\theta = \sin(2\theta)$ that follows by an easy application of
ref Euler's formula. Thus $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are also orthogonal, and so form an orthonormal basis for the two-dimensional subspace $\mathsf{P}_n = \operatorname{span}_{\mathbb{R}}(\boldsymbol{u}_1, \boldsymbol{u}_2)$. In Figure 5.1 we show the intersection of P_n with the unit cube in $\mathbb{R}^n$ for increasing n. As claimed, the unit ball in P_n becomes more round as n gets large. As we shall see, Dvoretzky's Theorem says that this phenomenon happens, in some sense, for general convex bodies. •

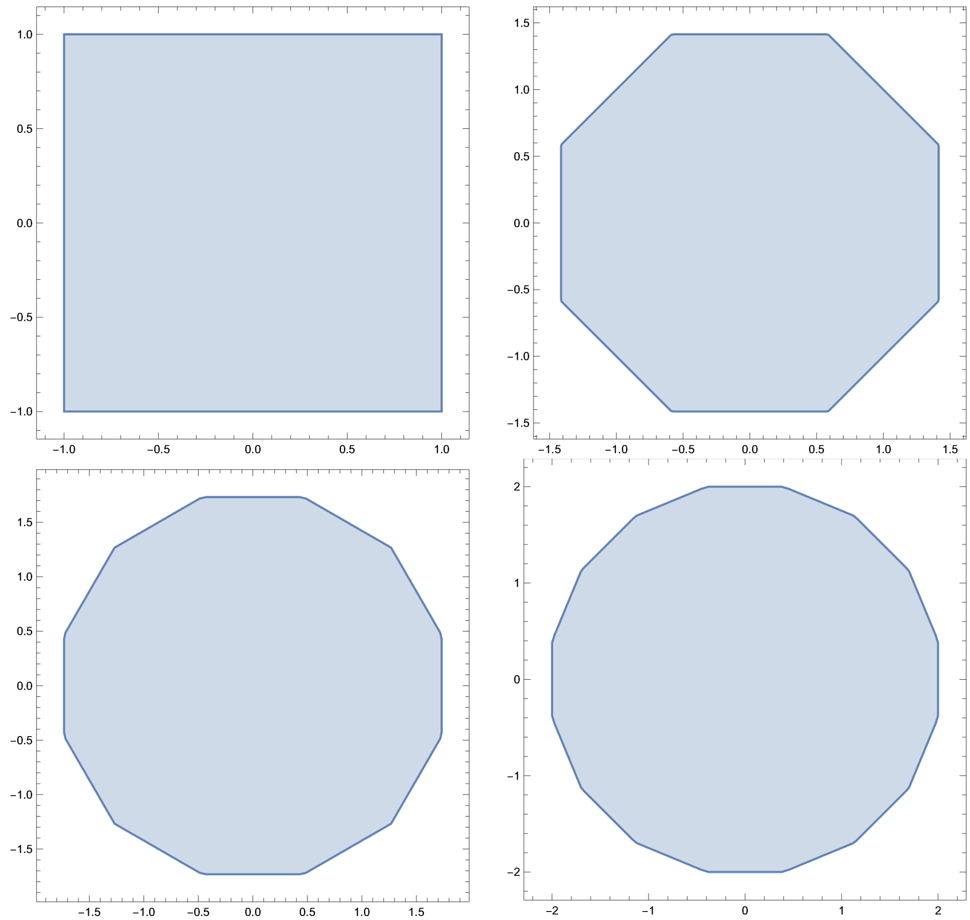


Figure 5.1 The intersection of the unit cube in $\mathbb{R}^n$ with the plane P_n for $n \in \{2, 4, 6, 8\}$

Let us be precise about this.

5.2.5 Theorem (Dvoretzky's Theorem) *If $\epsilon \in \mathbb{R}_{>0}$ then there exists $N_\epsilon \in \mathbb{Z}_{>0}$ and $\Delta_\epsilon \in \mathbb{R}_{>0}$ such that,*

(i) if $n > N_\epsilon$,

(ii) if $k \in \mathbb{Z}_{>0}$ satisfies $k \leq \lfloor \Delta_\epsilon \log(n) \rfloor$, and

(iii) if $(\mathsf{V}, \|\cdot\|)$ is an n-dimensional Banach space

then there are vectors $v_1, \dots, v_k \in \mathsf{V}$ for which

$$(1-\epsilon)\|(a_1, \dots, a_k)\|_{\mathbb{R}^k} \leq \Big\|\sum_{j=1}^{k} a_j v_j\Big\| \leq (1+\epsilon)\|(a_1, \dots, a_k)\|_{\mathbb{R}^k}$$

for every $(a_1, \dots, a_k) \in \mathbb{R}^k$.

Proof We start by proving a bunch of rather technical lemmata that seem rather irrelevant, but will nonetheless be important.

1 Lemma *Let* $(\mathsf{V}, \|\cdot\|)$ *be an* n*-dimensional Banach space and let* $\mathrm{m} = \lfloor \frac{\mathrm{n}}{2} \rfloor$ *be the largest integer less than or equal to* $\frac{\mathrm{n}}{2}$*. Then there exists* $\mathrm{w}_1, \dots, \mathrm{w}_\mathrm{m} \in \mathsf{V}$ *such that*

(i) $\|\mathrm{w}_\mathrm{j}\| \geq \frac{1}{2}$, $\mathrm{j} \in \{1, \dots, \mathrm{m}\}$*, and*

(ii) $\|\sum_{\mathrm{j}=1}^{\mathrm{m}} \mathrm{x}_\mathrm{j}\mathrm{w}_\mathrm{j}\| \leq \|(\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{m})\|_{\mathbb{R}^\mathrm{m}}$ *for every* $\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{m} \in \mathbb{R}$.

Proof Let $\alpha = 2 + \frac{m(m-1)}{n}$ and let $\mathsf{L}\colon \mathbb{R}^n \to \mathsf{V}$ be an isomorphism for which $\|\mathsf{L}\|_{\mathbb{R}^n,\mathsf{V}} \leq \frac{1}{\alpha}$ (we use the Euclidean norm on $\mathbb{R}^n$). Let $C = \phi(\overline{\mathsf{B}}(\frac{1}{2}, 0_\mathsf{V}))$ and, by Lemma 1 from the proof of Theorem 3.4.8, let $z_1, \dots, z_m \in \mathrm{bd}(C)$ be such that

$$\Big\|\sum_{j=1}^{m} x_j z_j\Big\|_{\mathbb{R}^n} \leq \alpha\|(x_1, \dots, x_m)\|_{\mathbb{R}^m}$$

for every $x_1, \dots, x_m \in \mathbb{R}$. Defining $w_j = \phi(z_j)$, $j \in \{1, \dots, m\}$, we then have

$$\Big\|\sum_{j=1}^{m} x_j w_j\Big\| \leq \frac{1}{\alpha}\Big\|\sum_{j=1}^{m} x_j z_j\Big\|_{\mathbb{R}^n} \leq \|(x_1, \dots, x_m)\|_{\mathbb{R}^m},$$

giving the desired result. ▼

Now, using the lemma, take $m = \lfloor \frac{n}{2} \rfloor$ and let $w_1, \dots, w_m \in \mathsf{V}$ satisfy $\|w_j\| \geq \frac{1}{2}$, $j \in \{1, \dots, m\}$, and

$$\Big\|\sum_{j=1}^{m} x_j w_j\Big\| \leq \|(x_1, \dots, x_m)\|_{\mathbb{R}^m}$$

for each $(x_1, \dots, x_m) \in \mathbb{R}^m$. Define $f\colon \mathbb{R}^m \to \mathbb{R}$ by

$$f(x_1, \dots, x_m) = \Big\|\sum_{j=1}^{m} x_j w_j\Big\|.$$

We have

$$|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)| \leq \Big\|\sum_{j=1}^{m} (x_{1,j} - x_{2,j}) w_j\Big\| \leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathbb{R}^n},$$

using Exercise 3.1.3. Thus f is Lipschitz with Lipschitz constant 1. Let us prove a useful inequality for $\mathrm{mean}(f)$.

2 Lemma *There exists* $\mathrm{c} \in \mathbb{R}_{>0}$ *such that, given the following data:*

(i) $\mathrm{n} \in \mathbb{Z}_{>0}$ *and* $\mathrm{m} = \lfloor \frac{\mathrm{n}}{2} \rfloor$*;*

(ii) an n*-dimensional Banach space* $(\mathsf{V}, \|\cdot\|)$*;*

(iii) vectors $\mathrm{w}_1, \dots, \mathrm{w}_\mathrm{m} \in \mathsf{V}$ *such that* $\|\mathrm{w}_\mathrm{j}\| \geq \frac{1}{2}$, $\mathrm{j} \in \{1, \dots, \mathrm{m}\}$*;*

(iv) the function $\mathrm{f}\colon \mathbb{R}^\mathrm{m} \to \mathbb{R}$ *defined by*

$$\mathrm{f}(\mathrm{x}_1, \dots, \mathrm{x}_\mathrm{m}) = \Big\|\sum_{\mathrm{j}=1}^{\mathrm{m}} \mathrm{x}_\mathrm{j}\mathrm{w}_\mathrm{j}\Big\|,$$

we have $\mathrm{mean(f)} \geq \mathrm{c}\sqrt{\log(\mathrm{n})}$, *the mean being taken with respect to the Gaussian measure on* $\mathbb{R}^{\mathrm{m}}$.

Proof Let $\boldsymbol{\epsilon} \in \{-1,1\}^m$ and define the isomorphism $\boldsymbol{\phi}_{\boldsymbol{\epsilon}}\colon \mathbb{R}^m \to \mathbb{R}^m$ by

$$\boldsymbol{\phi}_{\boldsymbol{\epsilon}}(x_1,\dots,x_m) = (\epsilon_1 x_1,\dots,\epsilon_m x_m).$$

Note that, by the change of variable theorem, since $|\det \boldsymbol{\phi}_{\boldsymbol{x}}| = 1$,

$$\begin{aligned}\int_{\mathbb{R}^m} f(\boldsymbol{x})\,\mathrm{d}\gamma^m(\boldsymbol{x}) &= \frac{1}{(\sqrt{2\pi})^m}\int_{\mathbb{R}^m} f(\boldsymbol{x})\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{x}\|^2_{\mathbb{R}^m}}\,\mathrm{d}\boldsymbol{x}\\ &= \frac{1}{(\sqrt{2\pi})^m}\int_{\mathbb{R}^m} f(\epsilon_1x_1,\dots,\epsilon_m x_m)\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{\phi}_{\boldsymbol{\epsilon}}(\boldsymbol{x})\|^2_{\mathbb{R}^m}}\,\mathrm{d}\boldsymbol{x}\\ &= \frac{1}{(\sqrt{2\pi})^m}\int_{\mathbb{R}^m} f(\epsilon_1x_1,\dots,\epsilon_m x_m)\mathrm{e}^{-\frac{1}{2}\|\boldsymbol{x}\|^2_{\mathbb{R}^m}}\,\mathrm{d}\boldsymbol{x}\\ &= \int_{\mathbb{R}^m} f\circ\boldsymbol{\phi}_{\boldsymbol{\epsilon}}(\boldsymbol{x})\,\mathrm{d}\gamma^m(\boldsymbol{x}).\end{aligned}$$

Therefore, we have

$$\begin{aligned}\mathrm{mean}(f) = \int_{\mathbb{R}^m} f(\boldsymbol{x})\,\mathrm{d}\gamma^m(\boldsymbol{x}) &= \frac{1}{2^m}\sum_{\boldsymbol{\epsilon}\in\{-1,1\}^m}\int_{\mathbb{R}^m} f\circ\boldsymbol{\phi}_{\boldsymbol{\epsilon}}(\boldsymbol{x})\,\mathrm{d}\gamma^m(\boldsymbol{x})\\ &= \int_{\mathbb{R}^m}\frac{1}{2^m}\sum_{\boldsymbol{\epsilon}\in\{-1,1\}^m}\Big\|\sum_{j=1}^m \epsilon_j x_j w_j\Big\|\,\mathrm{d}\gamma^m(\boldsymbol{x}).\end{aligned}$$

Now let $r \in \{1,\dots,m\}$ and note that

$$\begin{aligned}\|x_r w_r\| &= \Big\|\frac{1}{2}\Big(x_r w_r + \frac{1}{2^{m-1}}\sum_{\substack{\boldsymbol{\epsilon}\in\{-1,1\}^m\\ \epsilon_r=1}}\sum_{\substack{j=1\\ j\neq r}}^m \epsilon_j x_j w_j\Big) + \frac{1}{2}\Big(x_r w_r - \frac{1}{2^{m-1}}\sum_{\substack{\boldsymbol{\epsilon}\in\{-1,1\}^m\\ \epsilon_r=1}}\sum_{\substack{j=1\\ j\neq r}}^m \epsilon_j x_j w_j\Big)\Big\|\\ &\leq \frac{1}{2}\Big\|x_r w_r + \frac{1}{2^{m-1}}\sum_{\substack{\boldsymbol{\epsilon}\in\{-1,1\}^m\\ \epsilon_r=1}}\sum_{\substack{j=1\\ j\neq r}}^m \epsilon_j x_j w_j\Big\| + \frac{1}{2}\Big\|\frac{1}{2^{m-1}}\sum_{\substack{\boldsymbol{\epsilon}\in\{-1,1\}^m\\ \epsilon_r=-1}}\sum_{\substack{j=1\\ j\neq r}}^m \epsilon_j x_j w_j - x_r w_r\Big\|\\ &= \frac{1}{2}\Big\|\frac{1}{2^{m-1}}\sum_{\substack{\boldsymbol{\epsilon}\in\{-1,1\}^m\\ \epsilon_r=1}}\sum_{j=1}^m \epsilon_j x_j w_j\Big\| + \frac{1}{2}\Big\|\frac{1}{2^{m-1}}\sum_{\substack{\boldsymbol{\epsilon}\in\{-1,1\}^m\\ \epsilon_r=-1}}\sum_{j=1}^m \epsilon_j x_j w_j\Big\|\\ &\leq \frac{1}{2^m}\Big(\sum_{\substack{\boldsymbol{\epsilon}\in\{-1,1\}^m\\ \epsilon_r=1}}\Big\|\sum_{j=1}^m \epsilon_j x_j w_j\Big\| + \sum_{\substack{\boldsymbol{\epsilon}\in\{-1,1\}^m\\ \epsilon_r=-1}}\Big\|\sum_{j=1}^m \epsilon_j x_j w_j\Big\|\Big)\\ &= \frac{1}{2^m}\sum_{\boldsymbol{\epsilon}\in\{-1,1\}^m}\Big\|\sum_{j=1}^m \epsilon_j x_j w_j\Big\|.\end{aligned}$$

As this holds for every $r \in \{1, \dots, m\}$ we have

$$\begin{aligned}\mathrm{mean}(f) &\geq \int_{\mathbb{R}^m} \max\{\|x_1 w_1\|, \dots, \|x_m w_m\|\}\, \mathrm{d}\gamma^m(\boldsymbol{x}) \\ &\geq \frac{1}{2} \int_{\mathbb{R}^m} \|\boldsymbol{x}\|_\infty \, \mathrm{d}\gamma^m(\boldsymbol{x}) \geq \frac{C}{2} \sqrt{\log(\lfloor \tfrac{n}{2} \rfloor)},\end{aligned}$$

where C is as in Lemma 5.2.2. Note that

$$\lim_{n\to\infty} \frac{\sqrt{\log(\lfloor \frac{n}{2} \rfloor)}}{\sqrt{\log(n)}} = 1.$$

Thus let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$\frac{\sqrt{\log(\lfloor \frac{n}{2} \rfloor)}}{\sqrt{\log(n)}} \geq \frac{1}{2}$$

for $n \geq N$. If we take

$$c = \min\Big\{\frac{C}{4}\Big\} \cup \left\{ \frac{C}{2} \frac{\sqrt{\log(\lfloor \frac{n}{2} \rfloor)}}{\sqrt{\log(n)}} \;\middle|\; n \in \{1, \dots, N\} \right\},$$

then we get the desired estimate. ▼

Next we prove a couple of lemmata concerning collections of points in the unit sphere. We let

$$\mathbb{S}^{k-1} = \{\boldsymbol{a} \in \mathbb{R}^k \mid \|\boldsymbol{a}\|_{\mathbb{R}^k} = 1\}$$

what

be the unit sphere in $\mathbb{R}^k$. As in , if $r \in \mathbb{R}_{>0}$, an r-net for $\mathbb{S}^{k-1}$ is a collection $(\boldsymbol{a}_j)_{j\in\{1,\dots,N\}}$ of points in $\mathbb{S}^{k-1}$ such that, if $\boldsymbol{a} \in \mathbb{S}^{k-1}$, then there exists $j \in \{1, \dots, N\}$ such that $\|\boldsymbol{a} - \boldsymbol{a}_j\|_{\mathbb{R}^k} \leq r$.

3 Lemma *For each $\epsilon \in \mathbb{R}_{>0}$ there exists $\delta \in \mathbb{R}_{>0}$ such that, given the following data:*

(i) $\mathrm{k} \in \mathbb{Z}_{>0}$*;*

(ii) a δ net $\mathscr{N}$ for $\mathbb{S}^{\mathrm{k}-1}$*;*

(iii) a Banach space $(\mathsf{V}, \|\cdot\|)$*;*

(iv) $\mathrm{v}_1, \dots, \mathrm{v}_\mathrm{k} \in \mathsf{V}$ *satisfying*

$$(1 - \delta) \leq \Big\|\sum_{\mathrm{j}=1}^{\mathrm{k}} \mathrm{a}_\mathrm{j} \mathrm{v}_\mathrm{j}\Big\| \leq (1 + \delta)$$

for all $\mathbf{a} \in \mathscr{N}$,

it holds that

$$(1 - \epsilon)\|\mathbf{a}\|_{\mathbb{R}^\mathrm{k}} \leq \Big\|\sum_{\mathrm{j}=1}^{\mathrm{k}} \mathrm{a}_\mathrm{j} \mathrm{v}_\mathrm{j}\Big\| \leq (1 + \epsilon)\|\mathbf{a}\|_{\mathbb{R}^\mathrm{k}}$$

for all $\mathbf{a} \in \mathbb{R}^\mathrm{k}$.

Proof Note that
$$\lim_{r\downarrow 0}\frac{1-3r}{1-r}=1.$$
Thus there exists $\bar{r}'\in\mathbb{R}_{>0}$ such that $\frac{1-3r}{1-r}\geq\frac{1}{2}$ for $r\in(0,\bar{r}']$. We take $\bar{r}=\min\{1,\bar{r}'\}$. Let $k\in\mathbb{Z}_{>0}$, let $r\in(0,\bar{r})$, let $\mathscr{N}$ be an r-net of $\mathbb{S}^{k-1}$, let $(\mathsf{V},\|\cdot\|)$ be a Banach space, and let $v_1,\dots,v_k\in\mathsf{V}$ be such that
$$(1-r)\leq\Big\|\sum_{j=1}^{k}a_jv_j\Big\|\leq(1+r)\tag{5.10}$$
for all $\boldsymbol{a}\in\mathscr{N}$. Let $\boldsymbol{a}\in\mathbb{S}^{k-1}$. Then, since $\mathscr{N}$ is an r-net, there exists $\boldsymbol{a}_0\in\mathscr{N}$ such that $\|\boldsymbol{a}-\boldsymbol{a}_0\|_{\mathbb{R}^k}\leq r$. Define
$$\boldsymbol{u}_1=\frac{\boldsymbol{a}-\boldsymbol{a}_0}{\|\boldsymbol{a}-\boldsymbol{a}_0\|_{\mathbb{R}^k}}$$
so that $\boldsymbol{a}=\boldsymbol{a}_0+r_1\boldsymbol{u}_1$ with $r_1\in(0,r)$. Now apply the same procedure to $\boldsymbol{u}_1\in\mathbb{S}^{k-1}$ to get $\boldsymbol{u}_1=\boldsymbol{a}_1+r_2'\boldsymbol{u}_2$ for $\boldsymbol{a}_1\in\mathscr{N}$, $\boldsymbol{u}_2\in\mathbb{S}^{k-1}$, and $r_2'\in(0,r)$. Then
$$\boldsymbol{a}=\boldsymbol{a}_0+r_1\boldsymbol{a}_1+r_2\boldsymbol{u}_2$$
where $r_2=r_1r_2'$. This procedure can be iterated to give $\boldsymbol{a}=\sum_{l=0}^{\infty}r_l\boldsymbol{x}_l$ for $\boldsymbol{x}_l\in\mathscr{N}$ and $r_l\in(0,r^l)$ for each $l\in\mathbb{Z}_{>0}$. Note that $r_0=1$. Note that this sum converges by Proposition 3.4.2 and Example I-2.4.2–1 since $\|\boldsymbol{a}_l\|_{\mathbb{R}^k}=1$ for each $l\in\mathbb{Z}_{>0}$.

Now note that, for $\boldsymbol{a}=\sum_{l=0}^{\infty}r_l\boldsymbol{a}_l\in\mathbb{S}^{k-1}$ (with $r_0=1$ as above),
$$\Big\|\sum_{j=1}^{k}a_jv_j\Big\|=\Big\|\sum_{l=0}^{\infty}r_l\sum_{j=1}^{k}a_{l,j}v_j\Big\|\leq\Big(\sum_{l=0}^{\infty}r^l\Big\|\sum_{j=1}^{k}a_{l,j}v_j\Big\|\Big)\leq\frac{1+r}{1-r},$$
using Example I-2.4.2–1 and (5.10).

Note that
$$\Big\|\sum_{j=1}^{k}x_{0,j}v_j\Big\|\geq 1-r$$
since $\boldsymbol{a}_1\in\mathscr{N}$. We also have
$$\Big\|\sum_{j=1}^{k}\sum_{l=1}^{\infty}r_la_{l,j}v_j\Big\|\leq\Big(\sum_{l=1}^{\infty}r_l\Big\|\sum_{j=1}^{k}a_{l,j}v_j\Big\|\Big)\leq\frac{r(1+r)}{1-r}.$$
Noting that $r_0=1$ and using these last two relations we compute
$$\begin{aligned}\Big\|\sum_{j=1}^{k}a_jv_j\Big\|&=\Big\|\sum_{j=1}^{k}\Big(a_{0,j}v_j+\sum_{l=1}^{\infty}r_la_{l,j}v_j\Big)\Big\|\\&\geq\Big|\Big\|\sum_{j=1}^{k}a_{0,j}v_j\Big\|-\Big\|\sum_{j=1}^{k}\sum_{l=1}^{\infty}r_la_{l,j}v_j\Big\|\Big|\\&\geq\Big|1-r-\frac{r(1+r)}{1-r}\Big|=\frac{1-3r}{1-r}.\end{aligned}$$

Now let $\epsilon \in \mathbb{R}_{>0}$ and, since

$$\lim_{r\downarrow 0} \frac{1+r}{1-r} = 1, \quad \lim_{r\downarrow 0} \frac{1-3r}{1-r} = 1,$$

choose $\delta \in \mathbb{R}_{>0}$ such that

$$\frac{1+\delta}{1-\delta} \leq (1+\epsilon), \quad \frac{1-3\delta}{1-\delta} \geq (1-\epsilon).$$

Then

$$(1-\epsilon) \leq \Big\| \sum_{j=1}^{k} a_j v_j \Big\| \leq (1+\epsilon).$$

This establishes the conclusions of the lemma for $\boldsymbol{a} \in \mathbb{S}^{k-1}$. If $\boldsymbol{a} \in \mathbb{R}^k$ then define $\boldsymbol{a}' = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_{\mathbb{R}^k}} \in \mathbb{S}^{k-1}$. Then

$$(1-\epsilon) \leq \Big\| \sum_{j=1}^{k} a'_j v_j \Big\| \leq (1+\epsilon)$$

$$\implies \quad (1-\epsilon)\|\boldsymbol{a}\|_{\mathbb{R}^k} \leq \Big\| \sum_{j=1}^{k} a_j v_j \Big\| \leq (1+\epsilon)\|\boldsymbol{a}\|_{\mathbb{R}^k},$$

as desired. ▼

4 Lemma *For* r $\in \mathbb{R}_{>0}$*, there exists an* r*-net* $\mathscr{N}$ *for* $\mathbb{S}^{k-1}$ *such that* $\mathrm{card}(\mathscr{N}) \leq \mathrm{e}^{\frac{2k}{\delta}}$.

Proof Let $\boldsymbol{a} \in \mathbb{S}^{k-1}$ and let $\boldsymbol{a}_2 \in \mathbb{S}^{k-1}$ satisfy $\|\boldsymbol{a}_1 - \boldsymbol{a}_2\|_{\mathbb{R}^k} > r$. Then choose $\boldsymbol{a}_3 \in \mathbb{S}^{k-1}$ such that

$$\|\boldsymbol{a}_1 - \boldsymbol{a}_3\|_{\mathbb{R}^k}, \|\boldsymbol{a}_2 - \boldsymbol{a}_3\|_{\mathbb{R}^k} > r.$$

Carrying on this way, we define a sequence $\boldsymbol{a}_1, \boldsymbol{a}_2, \dots$ in $\mathbb{S}^{k-1}$. We claim that this sequence is finite. First, note that $\mathbb{S}^{k-1}$ is closed (it is the preimage of the closed set $\{1\}$ under the continuous map $\|\cdot\|_{\mathbb{R}^k}$, cf.) and bounded (obviously). Next, suppose that we have a countable sequence $(\boldsymbol{a}_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{S}^{k-1}$ such that $\|\boldsymbol{a}_j - \boldsymbol{a}_l\|_{\mathbb{R}^k} > r$ for $j \neq l$. Then there can clearly be no convergent subsequence of this sequence since all terms are separated from the others by distance at least r. By the Bolzano–Weierstrass Theorem, this contradicts the compactness of $\mathbb{S}^{k-1}$. Thus the sequence above can be written as $\boldsymbol{a}_1, \dots, \boldsymbol{a}_N$. We claim that this sequence is an r-net. If not, there exists $\boldsymbol{a} \in \mathbb{S}^{k-1}$ such that $\|\boldsymbol{a} - \boldsymbol{a}_j\|_{\mathbb{R}^k} > r$ for each $j \in \{1, \dots, N\}$. This contradicts the definition of the points $\boldsymbol{a}_1, \dots, \boldsymbol{a}_N$.

what

Next we estimate N. Note that the balls $\overline{\mathsf{B}}^k(\frac{r}{2}, \boldsymbol{a}_j)$, $j \in \{1, \dots, N\}$, are disjoint. Moreover, these balls are contained in the ball $\overline{\mathsf{B}}^k(1 + \frac{r}{2}, \boldsymbol{0})$. By Example II-1.6.37 there exists $v_k \in \mathbb{R}_{>0}$ such that

$$\lambda(\overline{\mathsf{B}}^k(1 + \frac{r}{2}, \boldsymbol{0})) = v_k(1 + \tfrac{r}{2})^k, \quad \lambda(\overline{\mathsf{B}}^k(\frac{r}{2}, \boldsymbol{a}_j)) = v_k(\tfrac{r}{2})^k, \qquad j \in \{1, \dots, N\}.$$

Since the balls $\overline{\mathsf{B}}^k(\frac{r}{2}, \boldsymbol{a}_j)$, $j \in \{1, \dots, N\}$, are disjoint and contained in $\overline{\mathsf{B}}^k(1 + \frac{r}{2}, \boldsymbol{0})$, we must have

$$N v_k (\tfrac{r}{2})^k = \lambda(\cup_{j=1}^N \overline{\mathsf{B}}^k(\frac{r}{2}, \boldsymbol{a}_j)) < \lambda(\overline{\mathsf{B}}^k(1 + \frac{r}{2}, \boldsymbol{0})) = v_k(1 + \tfrac{r}{2})^k.$$

Thus $N(\frac{r}{2})^k < (1 + \frac{r}{2})^k$ which gives

$$N < (1 + \tfrac{2}{r})^k < 1 + \tfrac{2k}{r} < \mathrm{e}^{\frac{2k}{r}},$$

using Exercise I-2.2.1 and the definition in Section I-3.8.1 of the exponential function. ▼

Now we complete the proof. We let $\epsilon \in \mathbb{R}_{>0}$ and $n \in \mathbb{Z}_{>0}$ be given, and we let $(\mathsf{V}, \|\cdot\|)$ be a Banach space of dimension n. We let $m = \lfloor \frac{n}{2} \rfloor$ and let $w_1, \dots, w_m \in \mathsf{V}$ be such that $\|w_j\| \geq \frac{1}{2}$, $j \in \{1, \dots, m\}$, and

$$\Big\|\sum_{j=1}^m x_j w_j\Big\| \leq \|(x_1, \dots, x_m)\|_{\mathbb{R}^m}$$

for each $\boldsymbol{x} \in \mathbb{R}^m$. As above, define $f \colon \mathbb{R}^m \to \mathbb{R}$ by

$$f(x_1, \dots, x_m) = \Big\|\sum_{j=1}^m x_j w_j\Big\|,$$

noting, as we showed above, that f is Lipschitz with Lipschitz constant 1. Let $k \in \mathbb{Z}_{>0}$ and let $\boldsymbol{a} \in \mathbb{S}^{k-1}$. Let us define $\pi_{\boldsymbol{a}} \colon (\mathbb{R}^m)^k \to \mathbb{R}^m$ by

$$\pi_{\boldsymbol{a}}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \sum_{j=1}^k a_j \boldsymbol{x}_j.$$

Following the constructions of Section 2.7.6, let $\gamma^{mk} \circ \pi_{\boldsymbol{a}}^{-1}$ be the image measure on $\mathbb{R}^m$.

5 Lemma *For each* $\mathbf{a} \in \mathbb{S}^{k-1}$, $\gamma^{mk} \circ \pi_{\mathbf{a}}^{-1} = \gamma^{m}$.

Proof Let $\{\boldsymbol{f}_1, \dots, \boldsymbol{f}_k\}$ be an orthonormal (with respect to the Euclidean inner product) basis for $\mathbb{R}^k$ such that $\boldsymbol{f}_1 = \boldsymbol{a}$ (see Section 4.4.2 for a discussion of orthonormal bases). Let us write

$$\boldsymbol{f}_r = (f_{r1}, \dots, f_{rk}), \qquad r \in \{1, \dots, k\}.$$

Let $\boldsymbol{F} \in \mathsf{L}(\mathbb{R}^k; \mathbb{R}^k)$ be the matrix whose (r, j)th component is f_{rj}. By , $\boldsymbol{F} \in \mathsf{O}(k)$. For what $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in (\mathbb{R}^m)^k$ let us write

$$\boldsymbol{x}_j = (x_{j1}, \dots, x_{jm}), \qquad j \in \{1, \dots, k\}.$$

Then, denote $\hat{\boldsymbol{x}}_\alpha \in \mathbb{R}^k$, $\alpha \in \{1, \dots, m\}$, by

$$\hat{\boldsymbol{x}}_\alpha = (x_{1\alpha}, \dots, x_{k\alpha}).$$

Define $\boldsymbol{R}\colon (\mathbb{R}^m)^k \to (\mathbb{R}^m)^k$ by

$$\boldsymbol{R}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \Big(\sum_{j=1}^{k} f_{1j}\boldsymbol{x}_j, \dots, \sum_{j=1}^{k} f_{kj}\boldsymbol{x}_j\Big)$$

We claim that $\boldsymbol{R}$ is orthogonal with respect to the Euclidean inner product on $(\mathbb{R}^m)^k$. Indeed, we compute

$$\begin{aligned}
\|\boldsymbol{R}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|^2_{\mathbb{R}^{mk}} &= \sum_{r=1}^{k}\sum_{j=1}^{k}\sum_{l=1}^{k} f_{rj}f_{rl}\langle \boldsymbol{x}_j, \boldsymbol{x}_l\rangle_{\mathbb{R}^m} \\
&= \sum_{r=1}^{k}\sum_{j=1}^{k}\sum_{l=1}^{k} f_{rj}f_{rl}\sum_{\alpha=1}^{m}\sum_{\beta=1}^{m} x_{j\alpha}x_{l\beta}\langle \boldsymbol{e}_\alpha, \boldsymbol{e}_\beta\rangle_{\mathbb{R}^m} \\
&= \sum_{\alpha=1}^{m}\sum_{r=1}^{k}\Big(\sum_{j=1}^{k} f_{rj}x_{j\alpha}\Big)\Big(\sum_{l=1}^{k} f_{rl}x_{l\alpha}\Big) = \sum_{\alpha=1}^{m}\langle \boldsymbol{F}\hat{\boldsymbol{x}}_\alpha, \boldsymbol{F}\hat{\boldsymbol{x}}_\alpha\rangle_{\mathbb{R}^k} \\
&= \sum_{\alpha=1}^{m}\|\hat{\boldsymbol{x}}_\alpha\|_{\mathbb{R}^k} = \sum_{\alpha=1}^{m}\sum_{j=1}^{k} x_{j\alpha}^2 = \sum_{j=1}^{k}\|\boldsymbol{x}_j\|^2_{\mathbb{R}^m} = \|(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|^2_{\mathbb{R}^{mk}}.
\end{aligned}$$

Therefore, by Theorem II-1.3.18, $\boldsymbol{R} \in \mathsf{O}(mk)$ as claimed.

Let $\boldsymbol{y} \in \mathbb{R}^m$ and define

$$S_{\boldsymbol{y}} = \{(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in (\mathbb{R}^m)^k \mid \boldsymbol{R}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = (\boldsymbol{y}, \boldsymbol{y}_1, \dots, \boldsymbol{y}_k),\ \boldsymbol{y}_2, \dots, \boldsymbol{y}_k \in \mathbb{R}^m\}.$$

We claim that $\pi_{\boldsymbol{a}}^{-1}(\boldsymbol{y}) = S_{\boldsymbol{y}}$. From the definition of $\boldsymbol{R}$, if $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in S_{\boldsymbol{y}}$ we immediately have

$$\pi_{\boldsymbol{a}}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \boldsymbol{y}.$$

Thus $S_{\boldsymbol{y}} \subseteq \pi_{\boldsymbol{a}}^{-1}(\boldsymbol{y})$. For the converse inclusion, let $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_m) \in \pi_{\boldsymbol{a}}^{-1}(\boldsymbol{y})$. Then, by definition of $\boldsymbol{R}$ again,

$$\boldsymbol{R}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = (\boldsymbol{y}, \boldsymbol{y}_2, \dots, \boldsymbol{y}_k)$$

where

$$\boldsymbol{y}_j = \sum_{l=1}^{k} f_{jl}\boldsymbol{x}_l, \qquad j \in \{2, \dots, k\}.$$

Thus $\pi_{\boldsymbol{a}}^{-1}(\boldsymbol{y}) \subseteq S_{\boldsymbol{y}}$.

Now let $B \in \mathscr{B}(\mathbb{R}^m)$ and note that, by the preceding paragraph,

$$\pi_{\boldsymbol{a}}^{-1}(B) = \{(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \in (\mathbb{R}^m)^k \mid \boldsymbol{R}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = (\boldsymbol{y}, \boldsymbol{y}_2, \dots, \boldsymbol{y}_k),\ \boldsymbol{y} \in B,\ \boldsymbol{y}_2, \dots, \boldsymbol{y}_k \in \mathbb{R}^m\}.$$

By definition of image measure, Lemma 5.2.1(iii), and Fubini's Theorem,

$$\begin{aligned}
\int_B \mathrm{d}\gamma^{mk}\circ\pi_{\boldsymbol{a}}^{-1}(\boldsymbol{y}) &= \int_{\pi_{\boldsymbol{a}}^{-1}(B)} \mathrm{d}\gamma^{mk}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \int_{\boldsymbol{R}(\pi_{\boldsymbol{a}}^{-1}(B))} \mathrm{d}\gamma^{mk}(\boldsymbol{y}_1, \dots, \boldsymbol{y}_k) \\
&= \int_{(\mathbb{R}^m)^{k-1}}\Big(\int_B \mathrm{d}\gamma(\boldsymbol{y})\Big)\,\mathrm{d}\gamma^{m(k-1)}(\boldsymbol{y}_2, \dots, \boldsymbol{y}_m) = \int_B \mathrm{d}\gamma(\boldsymbol{y}),
\end{aligned}$$

as desired. ▼

Using the lemma and Theorem 5.2.3 we have

$$\gamma^{mk}\Big(\Big\{(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\in(\mathbb{R}^m)^k \;\Big|\; \Big|f\Big(\sum_{j=1}^k a_j\boldsymbol{x}_j\Big)-\mathrm{mean}(f)\Big|\geq r\Big\}\Big)\leq 2\mathrm{e}^{-\alpha r^2}$$

for every $\boldsymbol{a}\in\mathbb{S}^{k-1}$ and $r\in\mathbb{R}_{>0}$, where $\alpha=\frac{2}{\pi^2}$.

Now let $\epsilon\in\mathbb{R}_{>0}$ and, by Lemmata 3 and 4, let $\delta\in\mathbb{R}_{>0}$ be such that there exists a δ-net $\mathscr{N}$ of cardinality at most $\mathrm{e}^{\frac{2k}{\delta}}$ for $\mathbb{S}^{k-1}$ for which, if

$$(1-\delta)\leq\Big\|\sum_{j=1}^k a_j v_j\Big\|\leq(1+\delta)$$

for every $\boldsymbol{a}\in\mathscr{N}$, then

$$(1-\epsilon)\|\boldsymbol{a}\|_{\mathbb{R}^k}\leq\Big\|\sum_{j=1}^k a_j v_j\Big\|\leq(1+\epsilon)\|\boldsymbol{a}\|_{\mathbb{R}^k}$$

for every $\boldsymbol{a}\in\mathbb{R}^k$. Note that δ depends only on ϵ, and not on k. Define $r=\delta\mathrm{mean}(f)$ and note that

$$\Big\{(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\in(\mathbb{R}^m)^k\Big|\;\Big|f\Big(\sum_{j=1}^k a_j\boldsymbol{x}_j\Big)-\mathrm{mean}(f)\Big|\geq r\Big\}$$
$$=\Big\{(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\in(\mathbb{R}^m)^k\;\Big|\;\Big|\Big\|\sum_{j=1}^k\sum_{\alpha=1}^m a_j\frac{x_{j\alpha}w_\alpha}{\mathrm{mean}(f)}\Big\|-1\Big|\geq\delta\Big\}$$

for every $\boldsymbol{a}\in\mathscr{N}$. Thus let us define

$$S(k,\delta,\boldsymbol{a})=\Big\{(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\in(\mathbb{R}^m)^k\;\Big|\;\Big|\Big\|\sum_{j=1}^k\sum_{\alpha=1}^m a_j\frac{x_{j\alpha}w_\alpha}{\mathrm{mean}(f)}\Big\|-1\Big|\geq\delta\Big\}$$

and note that

$$\gamma^{mk}(S(k,\delta,\boldsymbol{a}))\leq 2\mathrm{e}^{-\alpha\delta^2\mathrm{mean}(f)^2},$$

where $c\in\mathbb{R}_{>0}$ is as specified by Lemma 2. By Lemma 4 and countable additivity of the Gaussian measure, this means that

$$\gamma^{mk}(\cup_{\boldsymbol{a}\in\mathscr{N}}S(k,\delta,\boldsymbol{a}))\leq 2\exp\Big(\frac{2k}{\delta}-\alpha\delta^2\mathrm{mean}(f)^2\Big)\leq 2\exp\Big(\frac{2k}{\delta}-\alpha\delta^2c^2\log(n)\Big).\tag{5.11}$$

Note that α and c are absolute constants, independent of any other data. Let

$$N_\epsilon=\min\Big\{m\geq 2\;\Big|\;\frac{1}{2}\delta\frac{\log(\frac{1}{2})}{\log(m)}+\frac{1}{2}\alpha\delta^3c^2>0\Big\}.$$

Note that N_ϵ indeed depends only on ϵ via its dependence on δ. Now define

$$\Delta_\epsilon = \frac{1}{2}\delta\frac{\log(\frac{1}{2})}{\log(N_\epsilon)} + \frac{1}{2}\alpha\delta^3c^2.$$

If $n > N_\epsilon$ then

$$\Delta\log(n) = \frac{1}{2}\delta\log(\tfrac{1}{2})\frac{\log(n)}{\log(N_\epsilon)} + \frac{1}{2}\alpha\delta^3c^2\log(n) < \frac{1}{2}\delta\log(\tfrac{1}{2}) + \frac{1}{2}\alpha\delta^3c^2\log(n).$$

Thus, if $k \in \mathbb{Z}_{>0}$ satisfies $k \le \lfloor\Delta_\epsilon\log(n)\rfloor$, then

$$\frac{2k}{\delta} - \alpha\delta^2c^2\log(n) < \log(\tfrac{1}{2}).$$

By (5.11), with $n > N_\epsilon$ and $k \le \lfloor\Delta_\epsilon\log(n)\rfloor$, we have $\gamma^{mk}(\cup_{\boldsymbol{a}\in\mathscr{N}}S(k,\delta,\boldsymbol{a})) \in \mathbb{R}_{>0}$. Thus there exists

$$(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k) \in ((\mathbb{R}^m)^k \setminus \cup_{\boldsymbol{a}\in\mathscr{N}}S(k,\delta,\boldsymbol{a})).$$

Now, given the definition of $S(k,\delta,\boldsymbol{a})$, this means that

$$\left|\left\|\sum_{j=1}^k\sum_{\alpha=1}^m a_j\frac{x_{j\alpha}w_\alpha}{\operatorname{mean}(f)}\right\| - 1\right| < \delta$$

for every $\boldsymbol{a} \in \mathscr{N}$. Thus, taking

$$v_j = \sum_{\alpha=1}^m \frac{x_{j\alpha}w_\alpha}{\operatorname{mean}(f)}, \qquad j \in \{1,\dots,k\},$$

we see that

$$(1-\delta) \le \left\|\sum_{j=1}^k a_jv_j\right\| \le (1+\delta)$$

for every $\boldsymbol{a} \in \mathscr{N}$. Since δ has been chosen according to Lemma 3, the theorem follows. ■

5.2.3 Consequences of Dvoretzky's Theorem

In this section we consider some corollaries of Dvoretzky's Theorem as we state it in Theorem 5.2.5. As we shall see, each of the results we present can be categorised in three different ways, depending on how one states the hypotheses. We, therefore, provide two restatements of Dvoretzky's Theorem to set the stage for how the remaining presentation will look.

The first restatement of Dvoretzky's Theorem is simply a slight rearrangement of the hypotheses.

5.2.6 Corollary (Alternative statement of Dvoretzky's Theorem) *If* $\epsilon \in \mathbb{R}_{>0}$ *and if* $k \in \mathbb{Z}_{>0}$, *then there exists* $N_{\epsilon,k} \in \mathbb{Z}_{>0}$ *such that,*

(i) if $n > N_\epsilon$ *and*

(ii) if $(V, \|\cdot\|)$ *is an* n*-dimensional Banach space*

then there are vectors $v_1, \dots, v_k \in V$ *for which*

$$(1-\epsilon)\|(a_1,\dots,a_k)\|_{\mathbb{R}^k} \le \left\|\sum_{j=1}^{k} a_j v_j\right\| \le (1+\epsilon)\|(a_1,\dots,a_k)\|_{\mathbb{R}^k}$$

for every $(a_1, \dots, a_k) \in \mathbb{R}^k$.

Proof Let $N_\epsilon \in \mathbb{Z}_{>0}$ and $\Delta_\epsilon \in \mathbb{R}_{>0}$ be as in Dvoretzky's Theorem. Taking $N_{\epsilon,k} = \max\{N_\epsilon, e^{\frac{k}{\Delta_\epsilon}}\}$, we see that if n and $(V, \|\cdot\|)$ satisfy the hypotheses of the corollary, then k, n, and $(V, \|\cdot\|)$ satisfy the hypotheses of Dvoretzky's Theorem. The corollary then follows immediately. ■

The second restatement of Dvoretzky's Theorem is for infinite-dimensional Banach spaces. In this case, the hypotheses are simpler since infinite-dimensional Banach spaces obviously contain subspaces of arbitrarily large finite dimension.

5.2.7 Corollary (Dvoretzky's Theorem for infinite-dimensional Banach spaces) *If* $\epsilon \in \mathbb{R}_{>0}$, *if* $k \in \mathbb{Z}_{>0}$, *and if* $(V, \|\cdot\|)$ *is an infinite-dimensional Banach space, then there exists* $v_1, \dots, v_k \in V$ *such that*

$$(1-\epsilon)\|(a_1,\dots,a_k)\|_{\mathbb{R}^k} \le \left\|\sum_{j=1}^{k} a_j v_j\right\| \le (1+\epsilon)\|(a_1,\dots,a_k)\|_{\mathbb{R}^k}$$

for every $(a_1, \dots, a_k) \in \mathbb{R}^k$.

Next let us consider what Dvoretzky's Theorem says about convex bodies. First we provide a precise notion of the phenomenon we observed in Example 5.2.4.

5.2.8 Definition ((ϵ-round) section of a convex body in Euclidean space) Let $C \subseteq \mathbb{R}^n$ be a convex body and let $\epsilon \in \mathbb{R}_{>0}$.

(i) A ***section*** of C is a set of the form $C \cap A$ where A is an affine subspace.

(ii) The ***dimension*** of a section $C \cap A$ of C is the dimension of A.

(iii) A section $C \cap A$ is a ***central section*** if A is a subspace.

(iv) A section $C \cap A$ of a convex body is ***ϵ-round*** if there exists $x_0 \in \mathbb{R}^n$ and $r \in \mathbb{R}_{>0}$ such that

$$\overline{B}^n((1-\epsilon)r, x_0) \cap A \subseteq C \cap A \subseteq \overline{B}^n((1+\epsilon)r, x_0) \cap A. \qquad \bullet$$

Our first statement concerning the application of Dvoretzky's Theorem to convex bodies is the following.

5.2.9 Corollary (High-dimensional balanced convex bodies admit almost round sections I) *If* $\epsilon \in \mathbb{R}_{>0}$ *then there exists* $\mathrm{N}_\epsilon \in \mathbb{Z}_{>0}$ *and* $\Delta_\epsilon \in \mathbb{R}_{>0}$ *such that,*

(i) if $\mathrm{n} > \mathrm{N}_\epsilon$*,*

(ii) if $\mathrm{k} \in \mathbb{Z}_{>0}$ *satisfies* $2\mathrm{k} - 1 \leq \lfloor \Delta_\epsilon \log(\mathrm{n}) \rfloor$*, and*

(iii) if $\mathsf{C} \subseteq \mathbb{R}^{\mathrm{n}}$ *is a balanced convex body,*

then there is an ϵ*-round central section of* C *of dimension* k*.*

Proof For $\epsilon \in \mathbb{R}_{>0}$, let $\epsilon' \in \mathbb{R}_{>0}$ be such that

$$(1-\epsilon) \leq \frac{1}{1+\epsilon'}, \quad \frac{1}{1-\epsilon'} \leq (1+\epsilon).$$

Let $N'_{\epsilon'}$ and $\Delta'_{\epsilon'}$ be as given in Dvoretzky's Theorem. Let $n > N'_{\epsilon'}$ and $k \in \mathbb{Z}_{>0}$ satisfy $2k - 1 \leq \lfloor \Delta'_{\epsilon'} \log(n) \rfloor$. Let $C \subseteq \mathbb{R}^n$ be a balanced convex body and let $\|\cdot\|$ be the gauge of C (see Definition 5.1.2). By Dvoretzky's Theorem, let $v_1, \dots, v_{2k-1} \in \mathbb{R}^n$ satisfy

$$(1-\epsilon')\|a\|_{\mathbb{R}^{2k-1}} \leq \left\| \sum_{j=1}^{2k-1} a_j v_j \right\| \leq (1+\epsilon')\|a\|_{\mathbb{R}^{2k-1}}$$

for every $a \in \mathbb{R}^{2k-1}$. Let $\mathsf{V} \subseteq \mathbb{R}^n$ be the subspace spanned by $\{v_1, \dots, v_{2k-1}\}$ and let $\mathsf{V}^\perp$ be the orthogonal complement of V with respect to the Euclidean inner product. Let $v_{2k}, \dots, v_n$ be an orthonormal basis for $\mathsf{V}^\perp$ and define a linear map $\mathsf{L}\colon \mathbb{R}^n \to \mathbb{R}^n$ by asking that $\mathsf{L}(e_j) = v_j$, $j \in \{1, \dots, n\}$, where $\{e_1, \dots, e_n\}$ is the standard basis (this defines L by Theorem I-4.5.24). Define the ellipsoid $E_r = \mathsf{L}(\overline{\mathsf{B}}(r, \mathbf{0}))$. Note that if $x \in E_r \cap \mathsf{V}$ then

$$(1-\epsilon')r \leq \|x\| \leq (1+\epsilon')r.$$

Thus, for every $r \in \mathbb{R}_{\geq 0}$,

$$\overline{\mathsf{B}}((1-\epsilon')r, \mathbf{0}) \cap \mathsf{V} \subseteq E_r \cap \mathsf{V} \subseteq \overline{\mathsf{B}}((1+\epsilon')r, \mathbf{0}) \cap \mathsf{V},$$

where $\overline{\mathsf{B}}(r, \mathbf{0})$ is the ball of radius r centred at $\mathbf{0}$ for the norm $\|\cdot\|$. Therefore,

$$E_{(1-\epsilon)r} \cap \mathsf{V} \subseteq E_{(1+\epsilon)^{-1}r} \cap \mathsf{V} \subseteq \overline{\mathsf{B}}(r, \mathbf{0}) \cap \mathsf{V} \subseteq E_{(1-\epsilon)^{-1}r} \cap \mathsf{V} \subseteq E_{(1+\epsilon)r} \cap \mathsf{V}$$

for every $r \in \mathbb{R}_{\geq 0}$.

Since $C = \overline{\mathsf{B}}(1, \mathbf{0})$ by Theorem 5.1.3, it follows that the central section $C \cap \mathsf{V}$ is approximated by the ellipsoids $E_{1-\epsilon} \cap \mathsf{V}$ (from within) and $E_{1+\epsilon} \cap \mathsf{V}$ (from without).

Now we prove a lemma.

1 **Lemma** *If* $\mathrm{E} \subseteq \mathbb{R}^{2\mathrm{k}-1}$ *is an ellipsoid with centre* $\mathbf{0}$*, then there exists a* k*-dimensional subspace* $\mathsf{U} \subseteq \mathbb{R}^{2\mathrm{k}-1}$ *such that* $\mathrm{E} \cap \mathsf{U} = \mathrm{B} \cap \mathsf{U}$ *where* B *is a ball with respect to the Euclidean norm.*

Proof By Proposition 5.1.5, let $\{f_1, \dots, f_{2k-1}\}$ be an orthonormal basis such that

$$E = \{y_1 f_1 + \dots + y_n f_{2k-1} \in \mathbb{R}^{2k-1} \mid \lambda_1 y_1^2 + \dots + \lambda_{2k-1} y_{2k-1}^2 \leq 1\}$$

for $\lambda_j \in \mathbb{R}_{>0}$, $j \in \{1, \dots, 2k-1\}$. Now we define a linear map $\mathsf{A}\colon \mathbb{R}^{2k-1} \to \mathbb{R}^{k-1}$. To do so, we write $x \in \mathbb{R}^{2k-1}$ as

$$x = y_1 f_1 + \dots + y_{2k-1} f_{2k-1}$$

and

$$\mathsf{A}(y_1\boldsymbol{f}_1 + \cdots + y_{2k-1}\boldsymbol{f}_{2k-1}) = \mathsf{A}_1(y_1, \dots, y_{2k-1})\boldsymbol{e}_1 + \cdots + \mathsf{A}_{k-1}(y_1, \dots, y_{2k-1})\boldsymbol{e}_{k-1}$$

where $\mathsf{A}_j\colon \mathbb{R}^{2k-1} \to \mathbb{R}$, $j \in \{1, \dots, k-1\}$, is a linear function of the components $y_1, \dots, y_{2k-1}$ and $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1}\}$ is the standard basis for $\mathbb{R}^{k-1}$. We define the linear functions A_j, $j \in \{1, \dots, k-1\}$, by

$$\mathsf{L}_j(y_1, \dots, y_{2k-1}) = \begin{cases} y_j\sqrt{\lambda_j - \lambda_k} - y_{k+j-1}\sqrt{\lambda_{k+j-1} - \lambda_k}, & \lambda_k \notin \{\lambda_j, \lambda_{k+j-1}\}, \\ y_j, & \lambda_k = \lambda_{k+j-1}, \\ y_{k+j-1}, & \lambda_k = \lambda_j, \\ y_j - y_{k+j-1}, & \lambda_k = \lambda_j = \lambda_{k+j-1}. \end{cases}$$

One easily sees that A has rank $k-1$ (by, for example, directly checking that A is surjective) and so, if $\mathsf{U} = \ker(\mathsf{A})$, then $\dim(\mathsf{U}) = k$ by the Rank–Nullity Theorem.

Note that if

$$y_1\boldsymbol{f}_1 + \cdots + y_{2k-1}\boldsymbol{f}_{2k-1} \in \ker(\mathsf{A})$$

we have

$$\lambda_j y_j^2 + \lambda_{k+j-1}y_{k+j-1}^2 = \lambda_k(y_j^2 + y_{k+j-1}^2), \qquad j \in \{1, \dots, k-1\}.$$

Therefore, if

$$y_1\boldsymbol{f}_1 + \cdots + y_{2k-1}\boldsymbol{f}_{2k-1} \in \mathsf{U}$$

then

$$\lambda_k(y_1^2 + \cdots + y_{2k-1}^2) = \lambda_1 y_1^2 + \cdots + \lambda_{2k-1}y_{2k-1}^2.$$

Since $\{\boldsymbol{f}_1, \dots, \boldsymbol{f}_{2k-1}\}$ is an orthonormal basis, if

$$\boldsymbol{x} = y_1\boldsymbol{f}_1 + \cdots + y_{2k-1}\boldsymbol{f}_{2k-1},$$

then

$$\|\boldsymbol{x}\|^2_{\mathbb{R}^{2k-1}} = y_1^2 + \cdots + y_{2k-1}^2.$$

Thus $\boldsymbol{x} \in E \cap \mathsf{U}$ if and only if

$$\lambda_k\|\boldsymbol{x}\|^2_{\mathbb{R}^{2k-1}} = \lambda_1 y_1^2 + \cdots + \lambda_{2k-1}y_{2k-1}^2 \le 1,$$

i.e., if and only if $\boldsymbol{x} \in \overline{\mathsf{B}}^{2k-1}(\lambda_k^{-1}, \boldsymbol{0}) \cap \mathsf{U}$, giving the lemma. ▼

Now let $\{\boldsymbol{f}_1, \dots, \boldsymbol{f}_{2k-1}\}$ be an orthonormal basis for V and, using Theorem I-4.5.24, define an isomorphism $\iota\colon \mathbb{R}^{2k-1} \to \mathsf{V}$ by asking that $\iota(\boldsymbol{e}_j) = \boldsymbol{f}_j$, $j \in \{1, \dots, 2k-1\}$, where $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{2k-1}\}$ is the standard basis for $\mathbb{R}^{2k-1}$. Note that $\iota^{-1}(E_r \cap \mathsf{V})$ is an ellipsoid. By the lemma, let $\mathsf{U} \subseteq \mathbb{R}^{2k-1}$ be a k-dimensional subspace for which $\mathsf{U} \cap \iota^{-1}(E_r \cap \mathsf{V}) = \mathsf{U} \cap B_r$, where B_r is a ball (indexed by r, but not necessarily of radius r) with respect to the Euclidean norm on $\mathbb{R}^{2k-1}$. Note that, since ι maps the orthonormal basis $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{2k-1}\}$ to the orthonormal basis $\{\boldsymbol{f}_1, \dots, \boldsymbol{f}_{2k-1}\}$, $\iota(B_r) = B'_r \cap \mathsf{V}$ where B'_r is a ball in $\mathbb{R}^n$ with respect to the Euclidean norm; again, note that B'_r is not necessarily of radius r. Taking $\mathsf{U}' = \iota(\mathsf{U})$ we then have $E_r \cap \mathsf{U}' = B'_r \cap \mathsf{U}'$. Thus

$$B'_{1-\epsilon} \cap \mathsf{U}' \subseteq C \cap \mathsf{U}' \subseteq B'_{1+\epsilon} \cap \mathsf{U}',$$

so showing that C has the ϵ-round central section $C \cap \mathsf{U}'$. The result follows by taking $N_\epsilon = N'_{\epsilon'}$ and $\Delta_\epsilon = \Delta'_{\epsilon'}$. ■

Now we have an alternative statement of the preceding corollary.

5.2.10 Corollary (High-dimensional balanced convex bodies admit almost round sections II) *If $\epsilon \in \mathbb{R}_{>0}$ and if $k \in \mathbb{Z}_{>0}$, then there exists $N_{\epsilon,k} \in \mathbb{Z}_{>0}$ such that,*

(i) if $n > N_\epsilon$ and

(ii) if $C \subseteq \mathbb{R}^n$ is a balanced convex body,

then there is an ϵ-round central section of C of dimension k.

Proof This follows from Corollary 5.2.9 as Corollary 5.2.6 follows from Theorem 5.2.5. ■

In order to provide an infinite-dimensional statement of the preceding results, we need to first generalise the them somewhat to Banach spaces. The idea here is similar to our above constructions in Euclidean space, but there is a difference because in a general Banach space, the notion of "roundness" is not naturally defined.

5.2.11 Definition ((ϵ-round) section of convex body in a Banach space) Let $(\mathsf{V}, \|\cdot\|)$ be a Banach space, let $C \subseteq \mathbb{R}^n$ be a convex body, and let $\epsilon \in \mathbb{R}_{>0}$.

(i) A ***section*** of C is a set of the form $C \cap \mathsf{A}$ where A is an affine subspace.

(ii) The ***dimension*** of a section $C \cap \mathsf{A}$ of C is the dimension of A.

(iii) A section $C \cap \mathsf{A}$ is a ***central section*** if A is a subspace.

(iv) An n-dimensional section $C \cap \mathsf{A}$ of a convex body is ***ϵ-round*** if there exists an injective affine map $\iota\colon \mathbb{R}^n \to \mathsf{V}$ with $\operatorname{image}(\iota) = \mathsf{A}$ such that

$$\iota(\overline{\mathsf{B}}^n((1-\epsilon), \mathbf{0})) \subseteq C \cap \mathsf{A} \subseteq \iota\overline{\mathsf{B}}^n((1+\epsilon), \mathbf{0}). \quad \bullet$$

With this definition we have the following result.

5.2.12 Corollary (Unit balls in high-dimensional Banach spaces admit almost round sections I) *If $\epsilon \in \mathbb{R}_{>0}$ then there exists $N_\epsilon \in \mathbb{Z}_{>0}$ and $\Delta_\epsilon \in \mathbb{R}_{>0}$ such that,*

(i) if $n > N_\epsilon$,

(ii) if $k \in \mathbb{Z}_{>0}$ satisfies $k \leq \lfloor \Delta_\epsilon \log(n) \rfloor$, and

(iii) if $(\mathsf{V}, \|\cdot\|)$ is a Banach space of dimension n,

then there is an ϵ-round central section of $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ of dimension k, where $\overline{\mathsf{B}}(1, 0_{\mathsf{V}})$ denotes the unit ball centred at 0_{V} with respect to the norm $\|\cdot\|$.

Proof Let $\iota'\colon \mathbb{R}^n \to \mathsf{V}$ be an isomorphism (e.g., by choosing a basis for V and using Theorem I-4.5.45) and define a norm $\|\cdot\|'$ on $\mathbb{R}^n$ by $\|x\|' = \|\iota'(x)\|$. Let $\overline{\mathsf{B}}(1, \mathbf{0})' \subseteq \mathbb{R}^n$ be the unit ball with respect to the norm $\|\cdot\|'$. Let $\epsilon \in \mathbb{R}_{>0}$ and define N_ϵ and Δ_ϵ as in the proof of Corollary 5.2.9. As we saw in the proof of Corollary 5.2.9, there exists a k-dimensional subspace $\mathsf{U} \subseteq \mathbb{R}^n$ and an ellipsoid $E \subseteq \mathbb{R}^n$ with centre $\mathbf{0}$ such that

$$(1-\epsilon)E \cap \mathsf{U} \subseteq \overline{\mathsf{B}}(1, \mathbf{0})' \cap \mathsf{U} \subseteq (1+\epsilon)E \cap \mathsf{U}.$$

Therefore, if $U' = \iota'(U)$, then

$$\iota'((1-\epsilon)E) \cap U' \subseteq \overline{B}(1, 0_V) \cap U' \subseteq \iota'((1+\epsilon)E) \cap U'.$$

By definition of an ellipsoid, there exists an isomorphism $\iota'' \colon \mathbb{R}^n \to \mathbb{R}^n$ such that $E = \iota''(\overline{B}^n(1, \mathbf{0}))$. Note that

$$\iota''((1-\epsilon)\overline{B}^n(1, \mathbf{0})) = (1-\epsilon)E, \quad \iota''((1+\epsilon)\overline{B}^n(1, \mathbf{0})) = (1+\epsilon)E.$$

If we take $\iota \colon \mathbb{R}^n \to U'$ defined by $\iota = \iota' \circ \iota''$, then we have

$$\iota((1-\epsilon)\overline{B}(1, \mathbf{0})) \cap U' \subseteq \overline{B}(1, 0_V) \cap U' \subseteq \iota((1+\epsilon)\overline{B}(1, \mathbf{0})) \cap U',$$

which is the result. ■

In the usual way, we have the following restatement.

5.2.13 Corollary (Unit balls in high-dimensional Banach spaces admit almost round sections II) *If $\epsilon \in \mathbb{R}_{>0}$ and if $k \in \mathbb{Z}_{>0}$, then there exists $N_{\epsilon,k} \in \mathbb{Z}_{>0}$ such that,*

(i) if $n > N_\epsilon$ and

(ii) if $(V, \|\cdot\|)$ is a Banach space of dimension n,

then there is an ϵ-round central section of $\overline{B}(1, 0_V)$ of dimension k.

Proof This follows from Corollary 5.2.12 as Corollary 5.2.6 follows from Theorem 5.2.5. ■

Also in the usual way, we have the following infinite-dimensional version of the preceding two results.

5.2.14 Corollary (Almost round sections of unit balls in infinite-dimensional Banach spaces) *If $\epsilon \in \mathbb{R}_{>0}$, if $k \in \mathbb{Z}_{>0}$, and if $(V, \|\cdot\|)$ is an infinite-dimensional Banach space, then there exists an ϵ-round central section of $\overline{B}(1, 0_V)$ of dimension k.*

Our final interpretation of Dvoretzky's Theorem deals with characterising the differences between Banach spaces that are isomorphic as vector spaces.

5.2.15 Definition (Banach–Mazur[1] distance) If $(V_1, \|\cdot\|_1)$ and $(V_2, \|\cdot\|_2)$ are $\mathbb{R}$-Banach spaces that are isomorphic as $\mathbb{R}$-vector spaces, the ***Banach–Mazur distance*** between these Banach spaces is

$$d_{BM}(V_1, V_2) = \inf\{\|L\|_{1,2}\|L^{-1}\|_{2,1} \mid L \in L(V_1; V_2) \text{ is an isomorphism}\},$$

where $\|\cdot\|_{1,2}$ and $\|\cdot\|_{2,1}$ are the induced norms on $L(V_1; V_2)$ and $L(V_2; V_1)$, respectively. •

With this definition we have the following consequence of Dvoretzky Theorem.

[1] Barry Charles Mazur (1937–) is an American topologist and number theorist.

5.2.16 Corollary (High-dimensional Banach spaces contain subspaces close to Euclidean spaces I) *If* $\epsilon \in \mathbb{R}_{>0}$ *then there exists* $N_\epsilon \in \mathbb{Z}_{>0}$ *and* $\Delta_\epsilon \in \mathbb{R}_{>0}$ *such that,*

(i) *if* $n > N_\epsilon$,

(ii) *if* $k \in \mathbb{Z}_{>0}$ *satisfies* $k \leq \lfloor \Delta_\epsilon \log(n) \rfloor$, *and*

(iii) *if* $(\mathsf{V}, \|\cdot\|)$ *is an* n*-dimensional Banach space,*

then there exists a k*-dimensional subspace* U *of* V *such that*

$$d_{BM}(\mathsf{U}, \mathbb{R}^k) \leq 1 + \epsilon,$$

where the norm on U *is the restriction of* $\|\cdot\|$ *and the norm on* $\mathbb{R}^k$ *is the Euclidean norm.*

Proof Let $\epsilon \in \mathbb{R}_{>0}$ and let $\epsilon' \in \mathbb{R}_{>0}$ be sufficiently small that $\frac{1+\epsilon'}{1-\epsilon'} \leq 1 + \epsilon$. (This is possible since $\lim_{\epsilon' \downarrow 0} \frac{1+\epsilon'}{1-\epsilon'} = 1$.) By Corollary 5.2.12, let $\iota\colon \mathbb{R}^k \to \mathsf{V}$ be an injective linear map such that

$$\iota(\overline{\mathsf{B}}^k((1-\epsilon'), \mathbf{0})) \subseteq \overline{\mathsf{B}}(1, 0_{\mathsf{V}}) \cap \mathsf{U} \subseteq \iota(\overline{\mathsf{B}}^k((1+\epsilon'), \mathbf{0})),$$

where $\mathsf{U} = \operatorname{image}(\iota)$. Using Theorem 3.5.14,

$$\|\iota\|_{\mathbb{R}^k, \mathsf{U}} = \sup\{\|\iota(\boldsymbol{x})\|_{\mathsf{U}} \mid \boldsymbol{x} \in \overline{\mathsf{B}}^k(1, \mathbf{0})\} \leq \sup\{u \mid u \in \overline{\mathsf{B}}((1-\epsilon')^{-1}, 0_{\mathsf{V}})\} \leq (1-\epsilon')^{-1}$$

(noting that $\iota(\overline{\mathsf{B}}^k(1, \mathbf{0})) \subseteq \overline{\mathsf{B}}((1-\epsilon')^{-1}, 0_{\mathsf{V}})$) and

$$\|\iota^{-1}\|_{\mathsf{U}, \mathbb{R}^k} = \sup\{\|\iota^{-1}(u)\|_2 \mid u \in \overline{\mathsf{B}}(1, 0_{\mathsf{V}})\} \leq \sup\{\|\boldsymbol{x}\|_2 \mid \boldsymbol{x} \in \overline{\mathsf{B}}^k((1+\epsilon'), \mathbf{0})\} \leq (1+\epsilon'),$$

noting that $\iota^{-1}(\overline{\mathsf{B}}(1, 0_{\mathsf{V}})) \subseteq \overline{\mathsf{B}}^k((1+\epsilon'), \mathbf{0})$. Thus

$$\|\iota\|_{\mathbb{R}^k, \mathsf{U}} \|\iota^{-1}\|_{\mathsf{U}, \mathbb{R}^k} \leq \frac{1+\epsilon'}{1-\epsilon'} \leq 1 + \epsilon,$$

giving the desired conclusion. ■

As always, we have the following restatement of the preceding corollary.

5.2.17 Corollary (High-dimensional Banach spaces contain subspaces close to Euclidean spaces II) *If* $\epsilon \in \mathbb{R}_{>0}$ *and if* $k \in \mathbb{Z}_{>0}$, *then there exists* $N_{\epsilon,k} \in \mathbb{Z}_{>0}$ *such that,*

(i) *if* $n > N_\epsilon$ *and*

(ii) *if* $(\mathsf{V}, \|\cdot\|)$ *is a Banach space of dimension* n,

then there exists a k*-dimensional subspace* U *of* V *such that*

$$d_{BM}(\mathsf{U}, \mathbb{R}^k) \leq 1 + \epsilon,$$

where the norm on U *is the restriction of* $\|\cdot\|$ *and the norm on* $\mathbb{R}^k$ *is the Euclidean norm.*

Proof This follows from Corollary 5.2.16 as Corollary 5.2.6 follows from Theorem 5.2.5. ■

Finally, we have an infinite-dimensional version of the preceding two results.

5.2.18 Corollary (Infinite-dimensional Banach spaces contain subspaces close to Euclidean spaces) *If* $\epsilon \in \mathbb{R}_{>0}$*, if* $k \in \mathbb{Z}_{>0}$*, and if* $(\mathsf{V}, \|\cdot\|)$ *is an infinite-dimensional Banach space, then there exists a* k*-dimensional subspace* U *of* V *such that*

$$d_{BM}(\mathsf{U}, \mathbb{R}^k) \leq 1 + \epsilon,$$

where the norm on U *is the restriction of* $\|\cdot\|$ *and the norm on* $\mathbb{R}^k$ *is the Euclidean norm.*

5.2.4 Notes

Concentration of measure phenomenon by Ledoux [2001].
Gaussian concentration from [Pisier 1986].
[Dvoretzky 1960].
Idea of probabilistic approach due to Milman [1971].

Chapter 6

Topological vector spaces

In Chapters 3 and 4 we considered some important classes of vector spaces with topological structure. However, these spaces are not adequate to give useful topological descriptions of all of the vector spaces one encounters in applications. One can see this by a consideration of examples of Banach spaces from Section 3.8. Many of these spaces were spaces of functions with certain properties, and one can see that there are many classes of functions omitted from this description. For example, the simple space $\mathsf{C}^0(\mathbb{R};\mathbb{R})$ of continuous $\mathbb{R}$-valued functions on the real line is not among the spaces of functions that we described as a Banach space. In this chapter we describe a topological structure for vector spaces that includes Banach and Hilbert spaces as special cases, but which also includes useful descriptions of vector spaces that do not naturally admit the structure of a Banach or Hilbert space.

Do I need to read this chapter? The material in this chapter is important and useful, and some of the examples in Section 6.5 will feature prominently in our future developments. However, the chapter can probably be skipped until it is subsequently needed. •

Contents

6.1 General topological vector spaces . . . 621
6.1.1 Why go beyond norms? . . . 621
6.1.2 Definitions and basic properties . . . 623
6.1.3 Boundedness in topological vector spaces . . . 626
6.1.4 Attributes of subsets of a topological vector space . . . 628
6.1.5 Completeness of topological vector spaces . . . 629
Exercises . . . 629
6.2 Locally convex topological vector spaces . . . 631
6.2.1 Seminorms and Minkowski functionals . . . 631
6.2.2 Seminorms and locally convex topologies . . . 634
6.2.3 Properties of locally convex topological vector spaces . . . 638
Exercises . . . 641
6.3 Special classes of topological vector spaces . . . 642
6.3.1 Finite-dimensional spaces . . . 642

6.3.2 Normed spaces 643
6.3.3 Metrisable spaces 643
6.3.4 Countable union spaces 643
6.4 Linear maps between locally convex topological vector spaces 645
6.4.1 Duals of locally convex topological vector spaces 645
6.4.2 Reflexivity in locally convex topological vector spaces 645
6.4.3 Duals of Fréchet spaces 645
6.4.4 Duals of countable union spaces 646
6.5 Examples of locally convex spaces 648
6.5.1 Locally convex spaces of sequences 648
6.5.2 Locally convex spaces of continuous functions on $\mathbb{R}$ 650
6.5.3 Locally convex spaces of continuous functions on metric spaces 652
6.5.4 Locally convex spaces of locally integrable functions on $\mathbb{R}$ 652
6.5.5 Distributions as elements of locally convex topological vector spaces . . 653
Exercises 655

Section 6.1

General topological vector spaces

In this section we consider the general problem of topologising a vector space. The idea is that one want a topology that respects the vector space structure. While we shall make principal use of very specific topologies for vector spaces (but still sometimes more general than the situations illustrated in Chapters 3 and 4), it is worth developing the theory in a little generality in order to see what are the consequences of the most general sorts of axioms one might demand for a topology on a vector space.

Do I need to read this section? This is background material for this chapter, so if one is reading this chapter, this is the place to start. •

6.1.1 Why go beyond norms?

Were we to launch immediately into the definitions needed to get started with locally convex topologies, it is possible that the reader would quickly be alienated by the seemingly pointless abstraction. However, it is also true that the examples that illustrate the ideas of locally convex topologies are somewhat complicated, and so it is difficult to present these alongside the general theory without making substantial detours. Therefore, in this section, without rigour, we illustrate via an example why norms are too restrictive to cover some interesting situations, and how one might go beyond norms in a reasonable way.

We consider the set $\mathsf{C}^0(\mathbb{R};\mathbb{R})$ of continuous $\mathbb{R}$-valued functions on $\mathbb{R}$. In Section 3.8.5 we reminded the reader that, if we restrict consideration to members of $\mathsf{C}^0(\mathbb{R};\mathbb{R})$ that are bounded, then, with the norm $\|\cdot\|_\infty$ defined by

$$\|f\|_\infty = \{|f(x)| \mid x \in \mathbb{R}\},$$

the resulting vector space is a Banach space. However, if one does not restrict consideration to bounded functions, then clearly the norm $\|\cdot\|_\infty$ cannot even be defined. However, it is still possible that one may wish to topologies $\mathsf{C}^0(\mathbb{R};\mathbb{R})$, since it is a vector space of perfectly nice functions that one may very well encounter in applications.

To think about how one might usefully topologise $\mathsf{C}^0(\mathbb{R};\mathbb{R})$ we note that, if $K \subseteq \mathbb{R}$ is a compact set, then, for any member $f \in \mathsf{C}^0(\mathbb{R};\mathbb{R})$, $f|K$ is bounded. Thus we can define $\|f\|_{\infty,K} \in \mathbb{R}$ by

$$\|f\|_{\infty,K} = \sup\{|f(x)| \mid x \in K\}.$$

One can check that

1. $\|\lambda f\|_{\infty,K} = |\lambda|\|f\|_{\infty,K}$ for all $\lambda \in \mathbb{R}$ and $f \in \mathsf{C}^0(\mathbb{R};\mathbb{R})$,

2. $\|f + g\|_{\infty,K} \leq \|f\|_{\infty,K} + \|g\|_{\infty,K}$ for all $f, g \in \mathsf{C}^0(\mathbb{R};\mathbb{R})$, and
3. $\|f\|_{\infty,K} \geq 0$ for all $f \in \mathsf{C}^0(\mathbb{R};\mathbb{R})$.

Thus $\|\cdot\|_{\infty,K}$ falls short of being a norm only by the fact that there exists nonzero functions f for which $\|f\|_{\infty,K} = 0$. Note, however, that $\|\cdot\|_{\infty,K}$ *is* a norm on the vector space $\{f|K \mid f \in \mathsf{C}^0(\mathbb{R};\mathbb{R})\}$. We shall call such objects as $\|\cdot\|_{\infty,K}$ "seminorms." In this case, we have a family of seminorms

$$\{\|\cdot\|_{\infty,K} \mid K \subseteq \mathbb{R} \text{ is compact}\}.$$

The question becomes, "How can these seminorms be used to define a topology, and is this topology meaningful?"

To see how one might define a topology, instead of defining the topology directly, let us rather indicate what a convergent sequence might look like in this topology. Let $f \in \mathsf{C}^0(\mathbb{R};\mathbb{R})$ and let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{C}^0(\mathbb{R};\mathbb{R})$. We say that the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ ***converges*** to f if, for every compact subset $K \subseteq \mathbb{R}$, the sequence $(f_j|K)_{j\in\mathbb{Z}_{>0}}$ converges to $f|K$ in the norm $\|\cdot\|_{\infty,K}$.

Let us consider an example of this sort of convergence.

6.1.1 Example We let $f \in \mathsf{C}^0(\mathbb{R};\mathbb{R})$ be defined by $f(x) = 0$ for all $x \in \mathbb{R}$, and, for $j \in \mathbb{Z}_{>0}$, we define $f_j \in \mathsf{C}^0(\mathbb{R};\mathbb{R})$ by

$$f_j(x) = \begin{cases} 0, & x \in [-j, j], \\ x - j, & x \in (j, \infty), \\ -x - j, & x \in (-\infty, -j). \end{cases}$$

We claim that the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges to f. Indeed, let $K \subseteq \mathbb{R}$ be compact, and choose $N \in \mathbb{Z}_{>0}$ sufficiently large that $K \subseteq [-N, N]$. Then, since $f_j|[-N, N]$ is zero for $j \geq N$, it follows that $f_j|K$ is zero for geN, giving the desired conclusion. It may be useful for the reader to sketch the graphs of the functions in the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ to understand why it converges to f, and what this means and does not mean relative to the sorts of convergence considered in Section I-3.6. •

This is all we intend to say about this example for the moment. However, let us raise some of the more immediate questions arising from our rather superficial discussion.

1. Is there a topology on $\mathsf{C}^0(\mathbb{R};\mathbb{R})$ in which convergence is exactly as we have defined it? The answer is, "yes," and we consider this in . where?
2. While the topology on $\mathsf{C}^0(\mathbb{R};\mathbb{R})$ implicitly considered above is not the norm topology for the norm $\|\cdot\|_\infty$, is it the norm topology for some other norm? The answer is. "no," and this is considered in . where?
3. While the topology on $\mathsf{C}^0(\mathbb{R};\mathbb{R})$ implicitly considered above is not a norm topology for *any* norm, is there a metric on $\mathsf{C}^0(\mathbb{R}; l\mathbb{R})$ for which the topology is the metric topology? The answer here is, "yes," and we show this in . The answer to this question is (non-obviously) related to the fact that the following two statements are equivalent: where?

(a) for $f \in \mathsf{C}^0(\mathbb{R};\mathbb{R})$ and a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$, the sequence $(f_j|K)_{j\in\mathbb{Z}_{>0}}$ converges to $f|K$ in the norm $\|\cdot\|_{\infty,K}$ for every compact set $K \subseteq \mathbb{R}$;

(b) for $f \in \mathsf{C}^0(\mathbb{R};\mathbb{R})$ and a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$, the sequence $(f_j|[-N,N])_{j\in\mathbb{Z}_{>0}}$ converges to $f|[-N,N]$ in the norm $\|\cdot\|_{\infty,[-N,N]}$ for every $N \in \mathbb{Z}_{>0}$.

With this as motivation, we get down to the business of formally defining the notion of a locally convex topology.

6.1.2 Definitions and basic properties

In this section we shall have a need to regard $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ as topological spaces, and we do this by giving them their usual topology, which can be described in a multitude of ways.

1. One can regard $\mathbb{F}$ as a metric space with metric $\mathrm{d}(a_1,a_2) = |a_1 - a_2|$. This gives the balls
$$\mathsf{B}(r,a) = \{a' \in \mathbb{F} \mid \mathrm{d}(a',a) < r\}, \qquad r \in \mathbb{R}_{>0},\ a \in \mathbb{F}.$$
In this case the metric topology is prescribed by declaring that a subset $U \subseteq \mathbb{F}$ is open when, for any point in U, there is a ball centred at that point contained in U.
2. We can regard $\mathbb{F}$ as a normed vector space with norm $|\cdot|$. In this case, this gives rise to the metric described just above, and so gives rise to the same open sets.

With this in mind, we make the following definition.

6.1.2 Definition (Topological vector space) Let $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$, A ***topological $\mathbb{F}$-vector space*** is a pair $(\mathsf{V},\mathscr{O})$ where V is an $\mathbb{F}$-vector space and where $\mathscr{O}$ is a topology on V such that the maps
$$\begin{aligned}\mu\colon \mathbb{F}\times\mathsf{V} &\to \mathsf{V}\\ (a,v) &\mapsto av\end{aligned}$$
and
$$\begin{aligned}\alpha\colon \mathsf{V}\times\mathsf{V} &\to \mathsf{V}\\ (v_1,v_2) &\mapsto v_1+v_2\end{aligned}$$
are continuous. •

This definition is pretty natural, but gives rise to a surprising amount of structure. As an initial illustration of this, we have the following result.

6.1.3 Proposition (Determination of topology from a neighbourhood base of zero) *Let $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$ and let $(\mathsf{V},\mathscr{O})$ be a topological $\mathbb{F}$-vector space. Then $\mathscr{O}$ is uniquely determined by any neighbourhood base of* 0.

Proof First of all, by Proposition 1.2.26 we know that $\mathscr{O}$ is uniquely determined by the choice of a neighbourhood base at every point in V. Thus, to prove the proposition, it suffices to show that a neighbourhood base for 0 uniquely determined a neighbourhood base at v for every $v \in \mathsf{V}$.

Let $v \in \mathsf{V}$. By Exercise 6.1.1 we know that

$$\alpha_v \colon \mathsf{V} \to \mathsf{V} v' v' + v$$

is an homeomorphism of V. It follows from this that, if $\mathscr{B}_{v'}$ is a neighbourhood base for $v' \in \mathsf{V}$, then

$$\{\alpha_v(B) \mid B \in \mathscr{B}_{v'}\}$$

is a neighbourhood base for $\alpha_v(v') = v' + v$. (This is just a matter of verifying the definitions of a neighbourhood base.) Taking $v' = 0$ gives the desired conclusion. ■

With the preceding result in mind, we make a definition.

6.1.4 Definition (Local base) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological $\mathbb{F}$-vector space. A ***local base*** for $\mathscr{O}$ is a neighbourhood base for 0. •

The proof of the preceding proposition shows that if $\mathscr{B}_0$ is a local base for the topology of a topological vector space, then

$$v + \mathscr{B}_0 = \{\alpha_v(B) \mid B \in \mathscr{B}_0\}$$

is a neighbourhood base for v.

An important property of local bases is the following result, which is a consequence of topological vector spaces being regular, in the terminology of .

what

6.1.5 Proposition (Property of local bases) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological $\mathbb{F}$-vector space. If $\mathscr{B}_0$ is a local base for $\mathscr{O}$ and if $\mathrm{B} \in \mathscr{B}_0$, then there exists $\mathrm{B}' \in \mathscr{B}_0$ such that $\mathrm{cl}(\mathrm{B}') \subseteq \mathrm{B}$.*

Proof We begin with a general technical lemma.

1 Lemma *If U is a neighbourhood of 0, then there exists a neighbourhood V of 0 such that*

(i) $\mathrm{V} = -\mathrm{V}$ *and*

(ii) $\mathrm{V} + \mathrm{V} \subseteq \mathrm{U}$.

Proof Since $\alpha(0,0) = 0 + 0 = 0$ there exist neighbourhoods V_1 and V_2 of 0 such that $V_1 + V_2 \subseteq U$. Taking

$$V = V_1 \cap V_2 \cap (-V_1) \cap (-V_2)$$

gives the result. ▼

Note that one can then apply the lemma to the neighbourhood V itself to give a neighbourhood V of U such that $V = -V$ and $V + V + V + V \subseteq U$. In particular, $V + V + V \subseteq U$.

Now we state another lemma, from which the result will follow easily.

2 Lemma *If* $\mathsf{K} \subseteq \mathsf{V}$ *is compact and if* $\mathsf{A} \subseteq \mathsf{V}$ *is closed with* $\mathsf{K} \cap \mathsf{C} \neq \varnothing$*, then there is a neighbourhood* U *of* 0 *such that*

$$(\mathsf{K} + \mathsf{V}) \cap (\mathsf{C} + \mathsf{V}) = \varnothing.$$

Proof We can suppose that $K \neq \varnothing$, since the result is obvious otherwise. Let $v \in K$. Since $v \notin C$ and since C is closed, there exists a neighbourhood U_x of 0 such that $(v + U_x) \cap C = \varnothing$. Applying the preceding lemma, there exists a neighbourhood V_x of 0 satisfying $V_x = -V_x$ and

$$(x + V_x + V_x + V_x) \cap C = \varnothing \quad \Longrightarrow \quad (x + V_x + V_x) \cap (C - V_x) = \varnothing$$
$$\Longrightarrow \quad (x + V_x + V_x) \cap (C + V_x). \quad (6.1)$$

By compactness of K, let $x_1, \dots, x_m \in K$ be such that

$$K \subseteq (x_1 + V_{x_1}) \cup \dots \cup (x_m + V_{x_m}).$$

Denote $V = \cap_{j=1}^{m} V_{x_j}$ so that we have

$$K \subseteq \bigcup_{j=1}^{m} (x_j + V_{x_j} + V) \subseteq \bigcup_{j=1}^{m} (x_j + V_{x_j} + V_{x_j}).$$

Thus $(K + V) \cap (C + V) = \varnothing$ by (6.1). ▼

Note that, if K, C, and V are as in the lemma, then

$$\mathrm{cl}(K + V) \cap C \neq \varnothing.$$

To see this, note that

$$C + V = \cup_{v \in C} (x + V)$$

is open. Suppose that $v \in \mathrm{cl}(K + V) \cap C$. Thus $(v + V) \cap (K + V) \neq \varnothing$ by definition of closure. Since $v + V \in C + V$ we arrive at a contradiction of the lemma.

Now, to prove the proposition, let $B \in \mathscr{B}_0$ and apply the lemma in the case of $K = \{0\}$ and $C = \mathsf{V} \setminus B$. The lemma then gives a neighbourhood V of 0 such that $V \cap (C + V) = \varnothing$. By our comments just preceding

$$\varnothing = \mathrm{cl}(V) \cap C = \mathrm{cl}(V) \cap (\mathsf{V} \setminus B) \quad \Longrightarrow \quad \mathrm{cl}(V) \subseteq B.$$

Since $\mathscr{B}_0$ is a neighbourhood base, there exists $B' \in \mathscr{B}_0$ such that $B' \subseteq B$. Then $\mathrm{cl}(B') \subseteq B$. ■

We can further refine the "shape" of the elements of a local base for a topological vector space using the following notions.

6.1.6 Definition (Balanced set, absorbing set) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be an $\mathbb{F}$-vector space.

(i) A subset $B \subseteq \mathsf{V}$ is ***balanced*** if $\mu_a(B) \subseteq B$ for every $\alpha \in \mathbb{F}$ satisfying $|a| \leq 1$.

(ii) A subset $A \subseteq \mathsf{V}$ is ***absorbing*** if, for every $v \in \mathsf{V}$, there exists $\lambda \in \mathbb{R}_{>0}$ such that $x \in aA$ for every $a \in \mathbb{F}$ satisfying $|a| \geq \lambda$. •

The reader can explore the meaning of balanced sets in Exercise 6.1.3. We note that the notions of balanced and absorbing set are not dependent on the topology of V (although they do depend on the topology of $\mathbb{F}$).

For our purposes, we have the following results.

6.1.7 Proposition (Topological vector spaces possess balanced local bases) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological vector space. Then there exists a local base for $(\mathsf{V}, \mathscr{O})$ comprised of balanced sets.*

Proof First we show that every neighbourhood of 0 contains a balanced neighbourhood of 0. Let U be a neighbourhood of 0. By continuity of the scalar multiplication map μ, there exists $\delta \in \mathbb{R}_{>0}$ and a neighbourhood V of 0 such that $av \in U$ for $|a| < \delta$ and $v \in V$. Define

$$W = \{av \mid v \in V, \ |a| < \delta\}.$$

Being a union of the open sets $\mu_a(V)$, $|a| < \delta$, W is open. We claim that W is balanced. Indeed, let $w \in W$ and let $a \in \mathbb{F}$ be such that $|a| \leq 1$. Then $w = a'v$ for $|a| < \delta$ and $v \in V$. Thus $aw = (aa')v \in W$ since $|aa'| < |a'| < \delta$.

Now let $\mathscr{B}_0$ be a local base and, for each $B \in \mathscr{B}_0$ let U_B be a balanced neighbourhood of 0 such that $U_B \subseteq B$. Then

$$\{U_B \mid B \in \mathscr{B}_0\}$$

is a local base comprised of balanced sets. ■

6.1.8 Proposition (Neighbourhoods of zero are absorbing) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological vector space. Then every neighbourhood of 0 is an absorbing set.*

Proof Let U be a neighbourhood of 0 and let $v \in \mathsf{V}$. Note that the mapping $\mathbb{F} \ni a \mapsto av \in \mathsf{V}$ is continuous. Since U is a neighbourhood of 0, this means that

$$\{a \in \mathbb{F} \mid av \in U\}$$

is an open subset of $\mathbb{F}$ containing 0. Thus there exists $\lambda \in \mathbb{R}_{>0}$ such that $av \in U$ for $|a| \in (0, \lambda]$. Thus $v \in a^{-1}U$ for $|a^{-1}| \geq \lambda$, which is the result. ■

6.1.3 Boundedness in topological vector spaces

Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological $\mathbb{F}$-vector space. Let us first introduce a notion in a topological vector space that arises out of an interplay of the topological and vector space structure.

6.1.9 Definition (Bounded set) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological $\mathbb{F}$-vector space. A subset $E \subseteq \mathsf{V}$ is ***bounded*** if, for any neighbourhood U of 0, there exists $\lambda \in \mathbb{R}_{>0}$ such that $E \subseteq \mu_a(U)$ for every $a \in \mathbb{F}$ satisfying $|a| \geq \lambda$. •

The following characterisation of boundedness is often useful.

6.1.10 Proposition (Sequential characterisation of bounded sets) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological $\mathbb{F}$-vector space. For a subset $\mathrm{E} \subseteq \mathsf{V}$, the following statements are equivalent:*

(i) *E is bounded;*

(ii) *if $(\mathrm{v_j})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ is a sequence in E and if $(\mathrm{a_j})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ is a nowhere zero sequence in $\mathbb{F}$ converging to 0, then the sequence $(\mathrm{a_j v_j})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ converges to 0.*

Proof (i) $\Longrightarrow$ (ii) Let $U \subseteq \mathsf{V}$ be a balanced neighbourhood of 0 and let $\lambda \in \mathbb{R}_{>0}$ be such that $E \subseteq (aU)$ for $|a| \geq \lambda$. Then, in particular, $U \subseteq \lambda^{-1}E$. Let $N \in \mathbb{Z}_{>0}$ be such that $|a_j\lambda| < 1$ for $j \geq N$. Then, by virtue of U being balanced,

$$a_j v_j = (a_j\lambda)\lambda^{-1}v_j \in a_j\lambda U \in U, \qquad j \geq N.$$

Thus $(a_j v_j)_{j\in\mathbb{Z}_{>0}}$ converges to 0 by Proposition 6.1.7.

(ii) $\Longrightarrow$ (i) Suppose that E is not bounded. Then there exists a neighbourhood U of 0 and a sequence $(r_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}_{>0}$ satisfying $\lim_{j\to\infty} r_j = \infty$ such that $E \not\subset r_j V$ for every $j \in \mathbb{Z}_{>0}$. For $j \in \mathbb{Z}_{>0}$, let $v_j \in E$ be such that $v_j \notin r_j V$. Thus $r_j^{-1}v_j \notin V$ for every $j \in \mathbb{Z}_{>0}$. Thus we have a sequence $(r_j^{-1})_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}_{>0}$ converging to zero and sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ in E such that $(r_j^{-1}v_j)_{j\in\mathbb{Z}_{>0}}$ does not converge to 0. ■

The following result generalises the well known fact about normed vector spaces.

6.1.11 Proposition (Compact sets are bounded) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological $\mathbb{F}$-vector space. If $\mathrm{K} \subseteq \mathsf{V}$ is compact, then it is bounded.*

Proof Let U be a neighbourhood of 0 and let $V \subseteq U$ be a balanced neighbourhood, by Proposition 6.1.7. By Proposition 6.1.8, V is absorbing, and so

$$K \subseteq \bigcup_{j\in\mathbb{Z}_{>0}} jV.$$

By compactness of K, let $j_1, \dots, j_k \in \mathbb{Z}_{>0}$ be such that $j_1 < \dots < j_k$ and such that

$$K \subseteq j_1 U \cup \dots \cup j_k V.$$

Since V is balanced,

$$j_l j_k^{-1} V \subseteq V \quad \Longrightarrow \quad j_l V \subseteq j_k V, \qquad l \in \{1, \dots, k\}.$$

Similarly, if $a \in \mathbb{F}$ satisfies $|a| \geq j_k$, $j_k V \subseteq aV$. Thus, if $a \in \mathbb{F}$ satisfies $|a| \geq j_k$,

$$K \subseteq j_k V \subseteq aV \subseteq aU,$$

showing that K is bounded. ■

Recall from Theorem 3.6.15 that closed and bounded subsets of a normed vector space $(\mathsf{V}, \|\cdot\|)$ are compact if and only if V is finite-dimensional. One imagines that this carries over to topological vector spaces, and that closed and bounded subsets of a topological vector space $(\mathsf{V}, \mathscr{O})$ are compact if and only if V is finite-dimensional. This is a failing of imagination, however. Examples of infinite-dimensional topological vector spaces where closed and bounded subsets are compact are not necessarily trivial, so here we content ourselves with making a definition, saving for the presentation of examples having the stated property.

what

6.1.12 Definition (Heine–Borel property) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. A topological $\mathbb{F}$-vector space $(\mathsf{V}, \mathscr{O})$ has the ***Heine–Borel property*** if every closed and bounded subset of V is compact. •

6.1.4 Attributes of subsets of a topological vector space

Note that since V is a vector space, the notion of a subspace of V makes sense, as do the myriad other notions associated with the vector space structure. Also, the notion of a closed subset of V makes sense, as do the myriad other notions that come with a topology. The following result indicates how the vector space structure and the topology interact.

6.1.13 Proposition (Interaction of topological and vector space structure) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \mathscr{O})$ *be a topological* $\mathbb{F}$*-vector space. Let* $\mathrm{A}, \mathrm{A}' \subseteq \mathsf{V}$*, let* $\mathsf{U} \subseteq \mathsf{V}$ *be a subspace, let* $\mathrm{B} \subseteq \mathsf{V}$ *be balanced, let* $\mathsf{C} \subseteq \mathsf{V}$ *be convex, and let* $\mathrm{E} \subseteq \mathsf{V}$ *be bounded. Then the following statements hold:*

(i) $\mathrm{cl}(\mathrm{A}) = \cap\{\mathrm{A} + \mathrm{U} \mid \mathrm{U}$ *is a neighbourhood of* $0\}$*;*

(ii) $\mathrm{cl}(\mathrm{A}) + \mathrm{cl}(\mathrm{A}') \subseteq \mathrm{cl}(\mathrm{A} + \mathrm{A}')$*;*

(iii) $\mathrm{cl}(\mathsf{U})$ *is a subspace;*

(iv) $\mathrm{cl}(\mathrm{B})$ *is balanced and* $\mathrm{int}(\mathrm{B})$ *is balanced if* $0 \in \mathrm{int}(\mathrm{B})$*;*

(v) $\mathrm{cl}(\mathrm{C})$ *and* $\mathrm{int}(\mathrm{C})$ *are convex;*

(vi) $\mathrm{cl}(\mathrm{E})$ *is bounded.*

Proof (i) Following the proof of Proposition 6.1.3, $v \in \mathrm{cl}(A)$ if and only if $(v+U) \cap A \neq \emptyset$ for every neighbourhood U of 0. Thus $v \in \mathrm{cl}(A)$ if and only if $x \in A - U$ for every neighbourhood U of 0. Thus $x \in \mathrm{cl}(A)$ if and only if $x \in A + U$ for every neighbourhood U of 0, since $-U$ is a neighbourhood of 0 if U is a neighbourhood of 0.

(ii) Let $v \in \mathrm{cl}(A)$ and $v' \in \mathrm{cl}(A')$ and let U be a neighbourhood of $v + v'$. By continuity of addition, let U and U' be neighbourhood of v and v', respectively, such that $U + U' \subseteq U$. Then let $u \in A \cap U$ and $u' \in A' \cap V'$ since $v \in \mathrm{cl}(A)$ and $v' \in \mathrm{cl}(A')$. Then $u + u' \in (A + A') \cap U$, showing that $(A + A') \cap U \neq \emptyset$. Thus $v + v' \in \mathrm{cl}(A + A')$.

(iii) By Exercise 6.1.2, $a\,\mathrm{cl}(\mathsf{U}) = \mathrm{cl}(a\mathsf{U})$ for every $a \in \mathbb{F}$. We can now use part (ii):

$$a\,\mathrm{cl}(\mathsf{U}) + b\,\mathrm{cl}(\mathsf{U}) = \mathrm{cl}(a\mathsf{U}) + \mathrm{cl}(b\mathsf{U}) \subseteq \mathrm{cl}(a\mathsf{U} + b\mathsf{U}) = \mathrm{cl}(\mathsf{U}),$$

for $a, b \in \mathbb{F}$, and this gives this part of the proof.

(iv) That cl(B) is balanced follows in a manner entirely similar to that used to prove part (iii). Since μ_a is an homeomorphism for $a \neq 0$, we have $a\,\mathrm{int}(B) = \mathrm{int}(aB)$ for $|a| \in (0, 1]$. Thus

$$a\,\mathrm{int}(B) \subseteq a\,\mathrm{int}(B) \subseteq B, \qquad |a| \in (0, 1].$$

Since $a\,\mathrm{int}(B)$ is open, this then gives $a\,\mathrm{int}(B) \subseteq \mathrm{int}(B)$ for $|a| \in (0, 1]$. Since we clearly have $0\,\mathrm{int}(B) \subseteq \mathrm{int}(B)$ if $0 \in \mathrm{int}(B)$, this part of the result follows.

(v) That cl(C) is convex follows in a manner entirely similar to that used to prove part (iii). For $\lambda \in (0, 1)$ we have

$$\lambda\,\mathrm{int}(C) + (1 - \lambda)\,\mathrm{int}(C) \subseteq C,$$

just since $\mathrm{int}(C) \subseteq C$. Moreover, the set $\lambda\,\mathrm{int}(C) + (1 - \lambda)\,\mathrm{int}(C)$ is open by continuity of the vector space operations. Since int(C) is the union of all open subsets of C, this gives

$$\lambda\,\mathrm{int}(C) + (1 - \lambda)\,\mathrm{int}(C) \subseteq \mathrm{int}(C),$$

which is this part of the result.

(vi) Let U be a neighbourhood of 0. By Proposition 6.1.5, let V be a neighbourhood of 0 for which $\mathrm{cl}(V) \subseteq U$. Since E is bounded, there exists $\lambda \in \mathbb{R}_{>0}$ such that $E \subseteq (aV)$ for $|a| \geq \lambda$. Thus $E \subseteq (aU)$ for $|a| \geq \lambda$. ■

6.1.5 Completeness of topological vector spaces

Completeness featured prominently in our presentation of normed vector spaces, and we dedicated some time to understanding why it is an essential feature of these spaces in practice. It is, of course, similarly true that completeness is an essential feature of topological vector spaces. However, because of the increased generality of topological vector spaces as compared to normed vector spaces, it turns out that sequences are not adequate for describing completeness. Instead we use nets as explained in what?

Exercises

6.1.1 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological $\mathbb{F}$-vector space. Let $a_0 \in \mathbb{F} \setminus \{0\}$ and let $v_0 \in \mathsf{V}$. Show that the maps

$$\begin{aligned} \mu_{a_0}\colon \mathsf{V} &\to \mathsf{V} \\ v &\mapsto a_0 v' \end{aligned} \qquad \begin{aligned} \alpha_{v_0}\colon \mathsf{V} &\to \mathsf{V} \\ v &\mapsto v + v_0 \end{aligned}$$

are homeomorphisms of V.

6.1.2 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological $\mathbb{F}$-vector space. Let $A \subseteq \mathsf{V}$ and let $a \in \mathbb{F}$. Show that $\mathrm{cl}(\mu_a(A)) = \mu_a(\mathrm{cl}(A))$.

6.1.3 Which of the following subsets of the $\mathbb{R}$-vector space $\mathbb{R}^2$ is balanced?

(a)

Answer the same question, now identifying $\mathbb{R}^2$ with $\mathbb{C}$ and thinking of it as a $\mathbb{C}$-vector space.

6.1.4 Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be an $\mathbb{F}$-vector space. Let $v \in \mathsf{V}$, $a \in \mathbb{F}$, and $\lambda \in \mathbb{R}_{>0}$. If $B \subseteq \mathsf{V}$ is balanced, show that

$$av \in \lambda B \quad \Longleftrightarrow \quad |a|v \in \lambda B.$$

Section 6.2

Locally convex topological vector spaces

We introduce in this section an important and fairly easily described In Chapters 3 and 4 we considered the most common means by which a vector space may be topologised, namely by a norm. In Definition 3.1.2 we presented the idea of a seminorm on an $\mathbb{F}$-vector space, which is a $\mathbb{R}_{\geq 0}$-valued function with all of the properties of a norm, except that is lacks the property that the only vector of (semi)norm zero is the zero vector. Relative to our discussion for Banach spaces, we showed in Theorem 3.1.8 that every vector space with a seminorm can be turned into a related space with a norm by "quotienting out by the vectors with zero norm," as if these vectors were an embarrassment. Here we make use of multiple seminorms for the same vector space to develop a topology. The "locally convex" topologies we introduce here are less general than the sort of generality we discussed in Section 6.1, but the sorts of spaces we introduce here are commonly encountered in applications, and indeed cover many interesting and useful applications that are not covered by norms.

Do I need to read this section? If you are reading this chapter, this section is probably the most important one, since we shall make use of the structure we introduce here in various places, including in our discussion of systems in Chapter V-6. •

6.2.1 Seminorms and Minkowski functionals

In this section we work entirely with algebraic constructions, leaving for the next section the topological connections associated with the objects we introduce.

We first make a construction that generalises what we have already seen in Definition 5.1.2.

6.2.1 Definition (Minkowski functional) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a topological $\mathbb{F}$-vector space. If $A \subseteq \mathsf{V}$ is absorbing, then the ***Minkowski functional*** of A is the map $p_A \colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ given by

$$p_A(v) = \inf\{\lambda \in \mathbb{R}_{>0} \mid v \in \lambda A\}.$$ •

We wish to explore the relationships between Minkowski functions of convex, balanced, absorbing sets and seminorms. Our first result in this direction is the following. In Definition 3.1.2 we denoted a norm by the function $v \mapsto \|v\|$. Here we shall change the notation to the more common and flexible "p" to denote a typical seminorm.

6.2.2 Proposition (Seminorms are Minkowski functionals) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* V *be an* $\mathbb{F}$*-vector space, and let* p *be a seminorm on* V*. Then the subset*

$$\mathrm{B_p} = \{\mathrm{v} \in \mathsf{V} \mid \mathrm{p(v)} < 1\}$$

is convex, balanced, and absorbing, and $\mathrm{p} = \mathrm{p}_{\mathrm{B_p}}$*.*

Proof If $v \in B_p$ and if $a \in \mathbb{F}$ satisfies $|a| \leq 1$, then $p(av) = |a|p(v) < 1$. This shows that B_p is balanced. If $v_1, v_2 \in B_p$ and if $\lambda \in (0,1)$, we have

$$p(\lambda v_1 + (1-\lambda)v_2) \leq \lambda p(v_1) + (1-\lambda)p(v_2) < 1,$$

showing that B_p is convex. Let $v \in \mathsf{V}$ and note that, for all $a \in \mathbb{F}$ satisfying $|a| \geq p(v)$ we have

$$p(a^{-1}v) = |a|^{-1}p(v) < 1 \quad \Longrightarrow \quad a^{-1}v \in B_p \quad \Longrightarrow \quad v \in aB_p.$$

Thus B_p is absorbing.

Our preceding computation gives

$$\{\lambda \in \mathbb{R}_{>0} \mid \lambda > p(v)\} \subseteq \{\lambda \in \mathbb{R}_{>0} \mid v \in \lambda B_p\}.$$

Therefore,

$$\begin{aligned} p_{B_p}(v) = &\inf\{\lambda \in \mathbb{R}_{>0} \mid v \in \lambda B_p\} \\ \leq &\inf\{\lambda \in \mathbb{R}_{>0} \mid \lambda > p(v)\} = p(v). \end{aligned}$$

Now let $\lambda \in (0, p(v)]$ so that $p(\lambda^{-1}v) \geq 1$. Therefore, $\lambda^{-1}v \notin B_p$ and so $v \notin \lambda B_p$. Thus

$$p(v) \leq \inf\{\lambda \in \mathbb{R}_{>0} \mid v \in \lambda B_p\} = p_{B_p}(v),$$

which gives the result. ■

Thus seminorms give rise to sets whose Minkowski functional is equal to the seminorm we started with. Now let us consider the other construction, from subset to seminorm via the Minkowski functional.

6.2.3 Proposition (Minkowski functionals are sometimes seminorms) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, let* V *be an* $\mathbb{F}$*-vector space, and let* $\mathrm{B} \subseteq \mathsf{V}$ *be a convex, balanced, absorbing set. Then* $\mathrm{p_B}$ *is a seminorm. Moreover, if we define*

$$\mathrm{B}_< = \{\mathrm{v} \in \mathsf{V} \mid \mathrm{p_B(v)} < 1\}, \quad \mathrm{B}_\leq = \{\mathrm{v} \in \mathsf{V} \mid \mathrm{p_B(v)} \leq 1\},$$

then

$$\mathrm{B}_< \subseteq \mathrm{B} \subseteq \mathrm{B}_\leq$$

and $\mathrm{p}_{\mathrm{B}_<} = \mathrm{p_B} = \mathrm{p}_{\mathrm{B}_\leq}$*.*

Proof Let $v_1, v_2 \in \mathsf{V}$. Since B is absorbing, we can write $v_1 = \lambda_1 u_1$ and $v_2 = \lambda_2 u_2$ for $u_1, u_2 \in B$ and $\lambda_1, \lambda_2 \in \mathbb{R}_{>0}$. Then

$$\frac{\lambda_1}{\lambda_1+\lambda_2}u_1 + \frac{\lambda_2}{\lambda_1+\lambda_2}u_1 \in B \quad\Longrightarrow\quad \lambda_1 u_1 + \lambda_2 u_2 \in (\lambda_1+\lambda_2)B$$
$$\Longrightarrow\quad v_1 + v_2 \in (\lambda_1+\lambda_2)B.$$

Thus

$$\{\lambda_1, \lambda_2 \in \mathbb{R}_{>0} \mid v_1 \in \lambda_1 B,\ v_2 \in \lambda_2 B\} \subseteq \{\lambda_1, \lambda_2 \in \mathbb{R}_{>0} \mid v_1 + v_2 \in (\lambda_1+\lambda_2)B\}.$$

Thus we have

$$\begin{aligned} p_B(v_1+v_2) &= \inf\{\lambda \in \mathbb{R} \mid v_1+v_2 \in \lambda B\} \\ &= \inf\{\lambda_1+\lambda_2 \mid v_1+v_2 \in (\lambda_1+\lambda_2)B\} \\ &\le \inf\{\lambda_1+\lambda_2 \mid v_1 \in \lambda_1 B,\ v_2 \in \lambda_2 B\} \\ &= \inf\{\lambda_1 \mid v_1 \in \lambda_1 B\} + \inf\{\lambda_2 \mid v_2 \in \lambda_2 B\} = p_B(v_1) + p_B(v_2), \end{aligned}$$

using Proposition I-2.2.28. This gives the triangle inequality for p_B.

Now note that $0 \in B$ since B is absorbing. If $a = 0$ and $v \in B$, then

$$p_B(av) = p_B(0) = \inf\{\lambda \in \mathbb{R}_{>0} \mid 0 \in \lambda B\} = \inf \mathbb{R}_{>0} = 0,$$

since $0 \in B$. Now let $a \in \mathbb{R}_{>0}$. Then

$$\begin{aligned} p_B(av) &= \inf\{\lambda \in \mathbb{R}_{>0} \mid av \in \lambda B\} \\ &= \inf\{\lambda \in \mathbb{R}_{>0} \mid v \in \tfrac{\lambda}{a} B\} \\ &= \inf\{a\lambda \in \mathbb{R}_{>0} \mid v \in \lambda B\} \\ &= a \inf\{\lambda \in \mathbb{R}_{>0} \mid v \in \lambda B\} = a p_B(v), \end{aligned}$$

using Proposition I-2.2.28. Finally, let $a \in \mathbb{F} \setminus \{0\}$. Then, using Exercise 6.1.4 since B is balanced,

$$\begin{aligned} p_B(av) &= \inf\{\lambda \in \mathbb{R}_{>0} \mid av \in \lambda B\} \\ &= \inf\{\lambda \in \mathbb{R}_{>0} \mid |a|v \in \lambda B\} \\ &= |a| \inf\{\lambda \in \mathbb{R}_{>0} \mid v \in \lambda B\} = |a| p_B(v), \end{aligned}$$

using the fact that this part of the result has been proved for $a \in \mathbb{R}_{>0}$, even when B is not balanced. Thus p_B has the homogeneity property of a seminorm.

Suppose that $v \in B_<$. Then

$$\inf\{\lambda \in \mathbb{R}_{>0} \mid v \in \lambda B\} < 1.$$

Thus $v \in \lambda B$ for some $\lambda \in (0, 1)$. Thus $v \in B$ since B is balanced. This gives $B_< \subseteq B$. Now, if $v \in \mathsf{V} \setminus B_\le$, then

$$\inf\{\lambda \in \mathbb{R}_{>0} \mid v \in \lambda B\} > 1.$$

Thus $v \in \lambda B$ for some $\lambda > 1$. Since B is balanced, this means that $v \in \mathsf{V} \setminus B$. Thus $B \subseteq B_\le$.

The preceding inclusions give $p_{B_\leq} \leq p_B \leq p_{B_<}$. For the other inequalities, first let $v \in \mathsf{V}$ and let $\lambda, \mu \in \mathbb{R}_{>0}$ be such that $p_{B_\leq}(v) < \lambda < \mu$. Then

$$p_{B_\leq}(v/\lambda) < 1 \quad \Longrightarrow \quad v/\lambda \in B_\leq.$$

Thus $p_B(v/\lambda) \leq 1$ and $p_B(v/\mu) \leq \lambda/\mu < 1$. Thus $v/\mu \in B_<$ and so $p_{B_<}(v) \leq \mu$. This holds for every $t > p_{B_\leq}(v)$ and so $p_{B_<}(v) \leq p_{B_\leq}(v)$, and the result follows from this. ■

The punchline of the preceding two propositions is the following essential correspondence:

$$\{\text{seminorms on } \mathsf{V}\} \leftrightarrow \{\text{convex, balanced, absorbing sets}\}.$$

6.2.2 Seminorms and locally convex topologies

Now we introduce topology into our discussion of seminorms by a consideration of a—probably the most—important class of topological vector spaces, those that are "locally convex." We shall see that the property of a topological vector space being locally convex has a concrete realisation in terms of seminorms. We shall two this in three steps. First we show that a local base of convex balanced sets for a locally convex topology has associated with it gives rise to a family of seminorms, one for each convex balanced neighbourhood coming from the local base. Then we show that a family of seminorms gives rise to a natural collection of convex balanced neighbourhoods of zero, and these are a base for a locally convex topology. Finally, we show that the two constructions commute, giving the equivalence of two descriptions of a locally convex topological vector space.

We get started with the following definition.

6.2.4 Definition (Locally convex topological vector space) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. A topological $\mathbb{F}$-vector space $(\mathsf{V}, \mathscr{O})$ is ***locally convex*** if there exists a local base for $\mathscr{O}$ comprised of convex sets. •

This notion of a locally convex topological vector space is "geometric" in the sense that it gives a geometric property of a local base.

Before we get rolling with our programme of realising locally convex topologies using seminorms, let us establish the analogue of Proposition 6.1.7 for locally convex topological vector spaces.

6.2.5 Proposition (Locally convex topological vector spaces possess convex balanced local bases) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a locally convex topological $\mathbb{F}$-vector space. Then there is a local base for $(\mathsf{V}, \mathscr{O})$ comprised of convex balanced sets.*

Proof First we show that every convex neighbourhood of 0 contains a balanced convex neighbourhood of 0. We let U be a convex neighbourhood of 0. Denote

$$B = \cap\{aU \mid |a| = 1\}.$$

By continuity of scalar multiplication, let V be a neighbourhood of 0 and let $\delta \in \mathbb{R}_{>0}$ such that $aV \subseteq U$ for every $a \in \mathbb{F}$ such that $|a| < \delta$. Define

$$W = \{av \mid v \in U, |a| < \delta\}.$$

As we saw in the proof of Proposition 6.1.7, W is balanced and $W \subseteq U$. Then, if $a \in \mathbb{F}$ satisfies $|a| = 1$, we have $a^{-1}W = W$. Thus, if $|a| = 1$,

$$W \subseteq \alpha W \subseteq \alpha U \subseteq B.$$

Therefore, since W is a neighbourhood of 0, $\mathrm{int}(B)$ is a neighbourhood of 0. The sets aU, $|a| = 1$, are convex, and so B is the intersection of convex sets, and so is convex, cf. Exercise II-1.9.3. Therefore, by Proposition 6.1.13(v), $\mathrm{int}(B)$ is convex. It remains to show that $\mathrm{int}(B)$ is balanced, which will follow from Proposition 6.1.13(iv) if we can show that B is balanced. To this end, let $a \in \mathbb{F}$ satisfy $|a| \leq 1$ and write $a = rb$ where $r \in [0, 1]$ and $|b| = 1$. Then

$$\begin{aligned} aB = rbB &= \cap \{rba'U \mid a' \in \mathbb{F}, |a'| = 1\} \\ &= \cap \{ra'U \mid a' \in \mathbb{F}, |a'| = 1\}. \end{aligned}$$

Since $a'U$ is a convex set containing 0 and since $r \in [0, 1]$, $ra'U \subseteq a'U$ and so, if $|a| \leq 1$,

$$\begin{aligned} aB &= \cap \{ra'U \mid a' \in \mathbb{F}, |a'| = 1\} \\ &\subseteq \cap \{a'U \mid a' \in \mathbb{F}, |a'| = 1\} = B, \end{aligned}$$

and so we have demonstrated that every convex neighbourhood of 0 contains a balanced convex neighbourhood of 0.

Now let $\mathscr{B}_0$ be a local base of convex neighbourhoods of 0 and, for each $B \in \mathscr{B}_0$ let U_B be a convex balanced neighbourhood of 0 such that $U_B \subseteq B$. Then

$$\{U_B \mid B \in \mathscr{B}_0\}$$

is a local base comprised of convex balanced sets. ■

Now we consider how seminorms can be used to describe locally convex topologies. We begin by assigning to a locally convex topological vector space a family of seminorms.

6.2.6 Proposition (Family of seminorms associated to a locally convex topological vector space) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a locally convex topological $\mathbb{F}$-vector space. If $\mathscr{B}_0$ is a local base of convex balanced neighbourhoods of 0, then the family*

$$\{p_U \mid U \in \mathscr{B}_0\}$$

of seminorms has the following properties:

(i) *for each $U \in \mathscr{B}_0$, $U = \{v \in \mathsf{V} \mid p_U(v) < 1\}$;*

(ii) *the functions $p_U \colon \mathsf{V} \to \mathbb{R}$, $U \in \mathscr{B}_0$, are continuous.*

Proof First of all, since each $U \in \mathscr{B}_0$ is absorbing by Proposition 6.1.8, p_U is a seminorm by Proposition 6.2.3.

(i) This was shown during the proof of the last assertion of Proposition 6.2.3.

(ii) Let $\epsilon \in \mathbb{R}_{>0}$. Then, for $v_1, v_2 \in \epsilon U$ for $U \in \mathscr{B}_0$,

$$|p_U(v_1) - p_U(v_2)| \leq p_U(v_1 - v_2) < \frac{\epsilon}{2} + \frac{\epsilon}{2} < \epsilon$$

by Exercise 3.1.3 and the triangle inequality. This gives continuity of p_U. ■

Now we turn this around and show how to define a locally convex topology from a given family of seminorms.

6.2.7 Proposition (Locally convex topological vector space associated with a family of seminorms) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let V be an $\mathbb{F}$-vector space. Let $\mathscr{P}$ be a family of seminorms on V. For $p \in \mathscr{P}$ and for $k \in \mathbb{Z}_{>0}$, denote*

$$U(p, k) = \{v \in \mathsf{V} \mid p(v) < k^{-1}\}$$

and let $\mathscr{B}_0$ be the collection of all finite intersections of the sets $U(p, k)$, $p \in \mathscr{P}$, $k \in \mathbb{Z}_{>0}$. Then $\mathscr{B}_0$ is a convex balanced local base for a locally convex topology for V with the property that each $p \in \mathscr{P}$ is continuous.

Proof Of course, the topology $\mathscr{O}$ for V is defined by asking that

$$\{v + U \mid U \in \mathscr{B}_0\}$$

be a neighbourhood base for $v \in \mathsf{V}$. To show that this renders $(\mathsf{V}, \mathscr{O})$ a locally convex topological $\mathbb{F}$-vector space, we need only show that the operations μ of scalar multiplication and α of vector addition are continuous.

First we show continuity of addition. First let V be a neighbourhood of 0. Then, since $\mathscr{B}_0$ is a neighbourhood base for 0, there exists $p_1, \dots, p_m \in \mathscr{P}$ and $k_1, \dots, k_m \in \mathbb{Z}_{>0}$ such that

$$\bigcap_{j=1}^{m} U(p_j, k_j) \subseteq V.$$

Define

$$W = \bigcap_{j=1}^{m} U(p_j, 2k_j),$$

and note that, by the triangle inequality, $W + W \subseteq V$. With this computation at hand, let $v_1, v_2 \in \mathsf{V}$ and let $v_1 + v_2 + V$ be a neighbourhood of $v_1 + v_2$ with V a neighbourhood of 0. Then let W be as above and consider the neighbourhood $(v_1 + W) \times (v_2 + W)$ of $(v_1, v_2) \in \mathsf{V} \times \mathsf{V}$. Then

$$\alpha((v_1 + W) \times (v_2 + W)) \subseteq v_1 + v_2 + V,$$

and this proves continuity of addition.

To prove continuity of scalar multiplication, let $a \in \mathbb{F}$ and let $v \in \mathsf{V}$. Let U be a neighbourhood of 0 and let W be the neighbourhood of zero as constructed in the preceding paragraph. Let $\lambda \in \mathbb{R}_{>0}$ be such that $v \in \lambda W$ and define

$$\mu = \frac{\lambda}{1 + |a|\lambda}.$$

Let $v' \in v + \mu W$ and let $a' \in \mathbb{F}$ satisfy $|a' - a| < \lambda^{-1}$. Note that

$$||a'| - |a|| \leq |a' - a| < \frac{1}{\lambda}.$$

If $|a'| - |a| \geq 0$ this gives

$$|a'| - |a| < \frac{1}{\lambda} \quad \Longrightarrow \quad |a'|\lambda \leq 1 + |a|\lambda \quad \Longrightarrow \quad |a'|\mu \leq 1.$$

We similarly get this conclusion when $|a| - |a'| \geq 0$. We then have

$$a'v' - av = a'(v - v') + (a' - a)v \in |a'|\mu W + |a' - a|\lambda W \subseteq W + W \subseteq U$$

by virtue of the fact that W is balanced (being the intersection of balanced sets). This give continuity of scalar multiplication.

To show that $p \in \mathscr{P}$ is continuous, let $\epsilon \in \mathbb{R}_{>0}$ and let $k \in \mathbb{Z}_{>0}$ be such that $k^{-1} < \frac{\epsilon}{2}$. Then, for $v_1, v_2 \in U(p, k)$ we have

$$|p(v_1) - p(v_2)| \leq p(v_1 - v_2) \leq p(v_1) + p(v_2) < \frac{2}{k} < \epsilon,$$

giving the desired continuity. ■

6.2.8 Remark ("Commuting of descriptions of locally convex topologies) Now let us see that these descriptions of a locally convex topological vector space by (1) the prescription of a convex balanced local base and (2) a collection of seminorms really agree in the sense that the prescriptions "commute."

1. First suppose that we start with a convex balanced local base $\mathscr{B}_0$ for a locally convex topological vector space $(\mathsf{V}, \mathscr{O})$ and that we produce a collection of seminorms according to Proposition 6.2.6. Now, given this collection of seminorms, define a locally convex topology according to Proposition 6.2.7. The local base for this topology is then comprised of the sets

$$\bigcap_{j=1}^{m} \frac{1}{k_j} U_j \tag{6.2}$$

for $k_1, \dots, k_m \in \mathbb{Z}_{>0}$ and $U_1, \dots, U_m \in \mathscr{B}_0$. However, $U_j = p_{U_j}^{-1}((-\infty, 1))$ and so each of the sets $\frac{1}{k_j} U_j$ in is $\mathscr{O}$ by continuity of p_{U_j} and of scalar multiplication. Thus the sets (6.2) are in $\mathscr{O}$ and so the local base defined by Proposition 6.2.7 is the same as $\mathscr{O}$.

2. Now suppose that we start with a collection $\mathscr{P}$ of seminorms on a vector space V and define, according to Proposition 6.2.7, a locally convex topology $\mathscr{O}$. Then, starting from $\mathscr{O}$, we define a collection of seminorms according to Proposition 6.2.6. Among the seminorms will be those from $\mathscr{P}$ as we have $p = p_{U(p,1)}$. However, there will also be other seminorms. However, the seminorms obtained will define the same topology, just as we argued above. •

6.2.3 Properties of locally convex topological vector spaces

The presence of seminorms allows for characterisations of many topological concepts for locally convex topological vector spaces that resemble those for normed spaces. This concreteness is what makes these spaces especially easy to use.

6.2.9 Proposition (Seminorm characterisation of continuity) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{U}, \mathscr{O}_{\mathsf{U}})$ and $(\mathsf{V}, \mathscr{O}_{\mathsf{V}})$ be locally convex topological $\mathbb{F}$-vector spaces. Then a linear map $\mathsf{L} \in \mathrm{Hom}_{\mathbb{F}}(\mathsf{U}; \mathsf{V})$ is continuous if and only if, for every continuous seminorm p on V, there exists a continuous seminorm q on U such that $\mathrm{p}(\mathsf{L}(\mathrm{u})) \leq \mathrm{q}(\mathrm{u})$.*

Proof First suppose that L is continuous, and so continuous at 0, and let p be a continuous seminorm on V. Consider the neighbourhood $O \subseteq \mathsf{V}$ of 0 given by

$$O = \{v \in \mathsf{V} \mid p(v) < 1\}.$$

Continuity of L at 0 gives a neighbourhood $N = \mathsf{L}^{-1}(O) \subseteq \mathsf{U}$ of 0. Let $q_1, \dots, q_k$ be continuous seminorms for U and let $r_1, \dots, r_k \in \mathbb{R}_{>0}$ be such that

$$\{u \in \mathsf{U} \mid q_j(u) < r_j,\ j \in \{1, \dots, k\}\} \subseteq N.$$

Then

$$\mathsf{L}(\{u \in \mathsf{U} \mid q_j(u) < r_j,\ j \in \{1, \dots, k\}\}) \subseteq O.$$

Let $r = \min\{r_1, \dots, r_k\}$. Thus, if

$$u \in \{u' \in \mathsf{U} \mid q_j(u') < 1,\ j \in \{1, \dots, k\}\},$$

then

$$ru \in \{u' \in \mathsf{U} \mid q_j(u) < r,\ j \in \{1, \dots, k\}\} \subseteq N.$$

Thus

$$p(\mathsf{L}(ru)) = p(r\mathsf{L}(u)) = rp(\mathsf{L}(u)) < 1.$$

Therefore,

$$\sup\{p(\mathsf{L}(u)) \mid u \in \{u' \in \mathsf{U} \mid q_j(u') < 1,\ j \in \{1, \dots, k\}\}\} < r^{-1} < \infty.$$

Now let

$$s = \sup\{p(\mathsf{L}(u)) \mid u \in \{u' \in \mathsf{U} \mid q_j(u') < 1,\ j \in \{1, \dots, k\}\}\}.$$

For $u \in \mathsf{U}$ and for $\epsilon \in \mathbb{R}_{>0}$, denote

$$u_\epsilon = (q_1(u) + \cdots + q_k(u) + \epsilon)^{-1} u.$$

Note that

$$q_j(u_\epsilon) = \frac{q_j(u)}{q_1(u) + \cdots + q_k(u) + \epsilon} < 1, \qquad j \in \{1, \ldots, k\}.$$

Thus

$$u_\epsilon \in \{u' \in \mathsf{U} \mid q_j(u') < 1,\ j \in \{1, \ldots, k\}\},$$

whence

$$p(\mathsf{L}(u_\epsilon)) \le s,$$

and so

$$p(\mathsf{L}(u)) \le s(q_1(u) + \cdots + q_k(u) + \epsilon).$$

As this is valid for any $\epsilon \in \mathbb{R}_{>0}$, we have

$$p(\mathsf{L}(u)) \le s(q_1(u) + \cdots + q_k(u)).$$

Note that $s(q_1 + \cdots + q_k)$ is a continuous seminorm on U.

For the converse, let $u_0 \in \mathsf{U}$ and let $O \subseteq \mathsf{V}$ be a neighbourhood of $\mathsf{L}(u_0)$. Let $p_1, \ldots, p_k$ be continuous seminorms for V and let $r_1, \ldots, r_k \in \mathbb{R}_{>0}$ be such that

$$\{v \in \mathsf{V} \mid p_j(v - \mathsf{L}(u_0)) < r_j,\ j \in \{1, \ldots, k\}\} \subseteq O.$$

Let $q_1, \ldots, q_k$ be continuous seminorms for U satisfying

$$p_j(\mathsf{L}(u)) \le q_j(u), \qquad j \in \{1, \ldots, k\},\ u \in \mathsf{U}.$$

Then, if $u \in \mathsf{U}$ satisfies $q_j(u - u_0) < r_j$, then

$$p_j(\mathsf{L}(u - u_0)) \le q_j(u - u_0), \qquad j \in \{1, \ldots, k\},$$

and so

$$\mathsf{L}(u) \in \{v \in \mathsf{V} \mid p_j(v - \mathsf{L}(u_0)) < r_j,\ j \in \{1, \ldots, k\}\} \subseteq O,$$

giving continuity of L at u_0. ■

6.2.10 Proposition (Seminorm characterisation of boundedness) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and let $(\mathsf{V}, \mathscr{O})$ be a locally convex topological $\mathbb{F}$-vector space. A subset $\mathrm{B} \subseteq \mathsf{V}$ is bounded if and only if $\mathrm{p}|\mathrm{B}$ is bounded for every continuous seminorm p on V.*

Proof Suppose that B is bounded and let p be a continuous seminorm for V. Let $\lambda \in \mathbb{R}_{>0}$ be such that

$$B \subseteq a\{v \in \mathsf{V} \mid p(v) < 1\}, \qquad |a| \ge \lambda.$$

Then, for $|a| \ge \lambda$,

$$B \subseteq \{av \in \mathsf{V} \mid p(v) < 1\} = \{v \in \mathsf{V} \mid p(\tfrac{v}{a}) < 1\} = \{v \in \mathsf{V} \mid p(v) < \lambda\},$$

giving boundedness of p on B.

Next let B be such that $p|B$ is bounded for every continuous seminorm p. Let U be a neighbourhood of 0 and let $p_1, \dots, p_k$ be continuous seminorms and $r_1, \dots, r_k \in \mathbb{R}_{>0}$ be such that

$$\{v \in \mathsf{V} \mid p_j(v) < r_j,\ j \in \{1, \dots, k\}\} \subseteq U.$$

Let $r = \min\{1, \dots, k\}$. Let

$$m_j = \sup\{p_j(v) \mid b \in B\}, \qquad j \in \{1, \dots, k\},$$

and let $m = \max\{m_1, \dots, m_k\}$. Let $\lambda = \frac{m}{r}$, let $a \in \mathbb{F}$ satisfy $|a| > \lambda$, and let $v \in B$. Then

$$p_j(v) \le m \implies r^{-1}p_j(v) \le \lambda \implies p_j\left(\frac{v}{a}\right) < r < r_j,$$

and so

$$\frac{v}{a} \in \{v \in \mathsf{V} \mid p_j(v) < r_j,\ j \in \{1, \dots, k\}\} \subseteq U,$$

whence $v \in aU$, showing that B is bounded. ■

6.2.11 Proposition (Seminorm characterisation of convergence) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *let* $(\mathsf{V}, \mathscr{O})$ *be a locally convex topological* $\mathbb{F}$*-vector space, and let* $(\mathrm{I}, \preceq)$ *be a directed set. A net* $(\mathrm{v_i})_{\mathrm{i} \in \mathrm{I}}$ *in* V *converges to* $\mathrm{v_0} \in \mathsf{V}$ *if and only if the sequence* $(\mathrm{p}(\mathrm{v_0} - \mathrm{v_i}))_{\mathrm{i} \in \mathrm{I}}$ *converges to* 0 *for every continuous seminorm* p *on* V.

Proof First suppose that $(v_i)_{i \in I}$ converges to v_0, let $\epsilon \in \mathbb{R}_{>0}$, and let p be a continuous seminorm on V. Then there exists $i_0 \in I$ such that, for $i_0 \preceq i$,

$$v_i \in \{v \in \mathsf{V} \mid p(v_0 - v_i) < \epsilon\},$$

which proves convergence of the sequence $(p(v_0 - v_i))_{i \in I}$ to zero.

Now suppose that $(p(v_0 - v_i))_{i \in I}$ converges to zero for every continuous seminorm p. Let U be a neighbourhood of 0 and let $p_1, \dots, p_k$ be continuous seminorms and $r_1, \dots, r_k \in \mathbb{R}_{>0}$ be such that

$$\{v \in \mathsf{V} \mid p_j(v) < r_j,\ j \in \{1, \dots, k\}\} \subseteq U.$$

Then there exists $i_0 \in I$ such that, for $i_0 \preceq i$,

$$v_0 - v_i \in \{v \in \mathsf{V} \mid p_j(v) < r_j\}, \qquad j \in \{1, \dots, k\}.$$

Thus $v_0 - v_i \in U$ for $i_0 \preceq i$, and so we have convergence of $(v_i)_{i \in I}$ to v_0. ■

6.2.12 Proposition (Seminorm characterisation of Cauchy sequences) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *let* $(\mathsf{V}, \mathscr{O})$ *be a locally convex topological* $\mathbb{F}$*-vector space, and let* $(\mathrm{I}, \preceq)$ *be a directed set. A net* $(\mathrm{v_i})_{\mathrm{i} \in \mathrm{I}}$ *in* V *is Cauchy if and only if, for every continuous seminorm* p *on* V *and every* $\epsilon \in \mathbb{R}_{>0}$, *there exists* $\mathrm{i_0} \in \mathrm{I}$ *such that* $\mathrm{p}(\mathrm{v_v} - \mathrm{v_j}) < \epsilon$ *for* $\mathrm{i_o} \preceq \mathrm{i}, \mathrm{j}$.

Proof ■

6.2.13 Proposition (Seminorm characterisation of Hausdorffness) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \mathscr{O})$ *be a locally convex topological* $\mathbb{F}$*-vector space. Then* $(\mathsf{V}, \mathscr{O})$ *is Hausdorff if and only if, for every* $v \in \mathsf{V}$*, there exists a continuous seminorm* p *on* V *such that* $p(v) \in \mathbb{R}_{>0}$*.*

Exercises

6.2.1 Give an interpretation of Theorem 3.1.14 in terms of locally convex topologies and seminorms.

6.2.2 Use Proposition 6.2.9 to prove the equivalence of parts (i) and (iv) of Theorem 3.5.8.

Section 6.3

Special classes of topological vector spaces

6.3.1 Finite-dimensional spaces

$(\mathbb{R}^n, (\gamma_j)_{j\in\{1,\dots,n\}})$ is a finite multinormed space. Indeed, it is clear that $(\gamma_j)_{j\in\{1,\dots,n\}}$ is separating. Some examples of balloons at $(0,0) \in \mathbb{R}^2$ are shown in Figure 6.1. Note that it is generally true that the intersection of two balloons (and therefore any

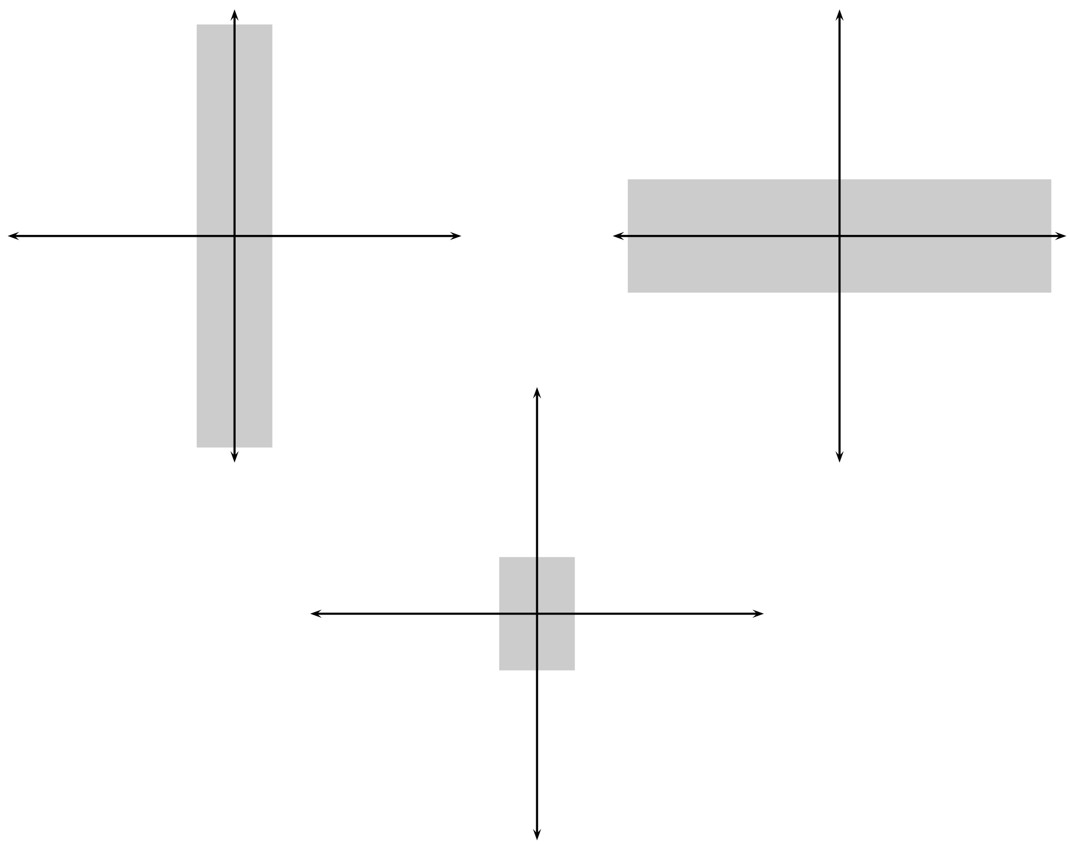


Figure 6.1 Balloons in $\mathbb{R}^2$. The bottom balloon is the intersection of the top two.

finite number of balloons) is itself a balloon. Note that open sets have the property that around any point in the set there is an open rectangle contained in the set. This quite clearly matches the usual definition of an open subset of $\mathbb{R}^n$. One also readily verifies that convergence in the multinorm is the same as convergence in the usual sense in $\mathbb{R}^n$. Since with this notion of convergence it is true that Cauchy sequences converge, it follows that $\mathbb{R}^n$ is a Fréchet space.

6.3.2 Normed spaces

6.3.1 Theorem (Characterisation of normed locally convex spaces) *Let* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ *and let* $(\mathsf{V}, \mathscr{O})$ *be a topological* $\mathbb{F}$*-vector space. Then* $(\mathsf{V}, \mathscr{O})$ *is normable if and only if it possesses a bounded convex neighbourhood of* 0*.*

Proof ■

6.3.3 Metrisable spaces

The essential structure that leads to the myriad useful properties of generalised signals is the structure of convergence on the sets of test signals. In the chapter to this point we merely defined directly the sort of convergence we use. Note, for example, that convergence in $\mathscr{D}(\mathbb{R};\mathbb{F})$ is *not* that of convergence with respect to a norm. Here we describe the general structure that gives rise to convergence in $\mathscr{D}(\mathbb{R};\mathbb{F})$ and our other spaces of test signals. What follows is a sequence of definitions and simple examples that lead us to our objective.

For countable multinorms we have the following characterisation of their structure, which is useful.

6.3.2 Proposition *If* $(\mathsf{V}, (\gamma_{\mathrm{j}})_{\mathrm{j}\in\mathbb{Z}_{>0}})$ *is a countable multinormed vector space then, for each* $\mathrm{v}_0 \in \mathsf{V}$*, there exists a sequence* $(\mathrm{B}_{\mathrm{k}})_{\mathrm{k}\in\mathbb{Z}_{>0}}$ *of balloons at* v_0 *with the properties*

(i) $\mathrm{B}_{\mathrm{k}+1} \subseteq \mathrm{B}_{\mathrm{k}}$, $\mathrm{k} \in \mathbb{Z}_{>0}$*, and*

(ii) every neighbourhood of v_0 *contains at least one of the balloons from* $(\mathrm{B}_{\mathrm{k}})_{\mathrm{k}\in\mathbb{Z}_{>0}}$*.*

Proof Define a collection of balloons by

$$C_{j,q} = \{v \in \mathsf{V} \mid \gamma_j(v - v_0) < q\}$$

where $j \in \mathbb{Z}_{>0}$ and where q is a positive rational number. This forms a countable set of balloons, and so may be enumerated as $(C_m)_{m\in\mathbb{Z}_{>0}}$. Now define $B_1 = C_1$ and inductively define $B_{k+1} = B_k \cap C_{k+1}$. Note now that the collection $(B_k)_{k\in\mathbb{Z}_{>0}}$ clearly satisfies the first condition of the proposition. Note that every balloon is contained in one of the balloons $(C_m)_{m\in\mathbb{Z}_{>0}}$. Thus if any neighbourhood must contain one of the balloons $(C_m)_{m\in\mathbb{Z}_{>0}}$. The result follows immediately. ■

6.3.4 Countable union spaces

In the preceding section we saw that Fréchet spaces generalised the notion of Banach spaces by allowing the topological notions of openness, convergence, etc., to be defined using a collection of seminorms rather then a norm. This setup was, as we saw in Example 6.5.8–1, not adequate to define convergence in $\mathscr{D}(\mathbb{R};\mathbb{F})$. In this section we generalise this idea even further by consider nested families of Fréchet spaces. It is this sort of idea that we need to complete our discussion of the topology on spaces of test signals.

With this definition we provide a new sort of topological structure for a vector space. In the following definition we use the fact that if $(\mathsf{V}, (\gamma_a)_{a\in\mathbb{Z}_{>0}})$ is a multinormed vector space and if U is a subspace of V, then the multinorm $(\gamma_a)_{a\in\mathbb{Z}_{>0}}$ can be thought of on U by simple restriction.

6.3.3 Definition Let V be an $\mathbb{F}$-vector space and let $((\mathsf{V}_j, (\gamma_{j,k})_{k\in\mathbb{Z}_{>0}}))_{j\in\mathbb{Z}_{>0}}$ be a collection of countable multinormed vector spaces with the following properties:

(i) V_j is a subspace of V for each $j \in \mathbb{Z}_{>0}$;

(ii) $\mathsf{V}_j \subseteq \mathsf{V}_{j+1}$ for $j \in \mathbb{Z}_{>0}$;

(iii) $\mathsf{V} = \cup_{j\in\mathbb{Z}_{>0}} \mathsf{V}_j$;

(iv) for each $j \in \mathbb{Z}_{>0}$, the multinorm induced on V_j by the multinorm $(\gamma_{j+1,k})_{k\in\mathbb{Z}_{>0}}$ on V_{j+1} is weaker that the multinorm $(\gamma_{j,k})_{k\in\mathbb{Z}_{>0}}$.

Then V, equipped with the collection of countable multinormed vector spaces $((\mathsf{V}_j, (\gamma_{j,k})_{k\in\mathbb{Z}_{>0}}))_{j\in\mathbb{Z}_{>0}}$, is a ***countable union space***. •

The notion of convergence in a countable union space is something that is not immediately clear from the definition, so we state it explicitly.

6.3.4 Definition Let V be a countable union space defined by a collection of countable multinormed vector spaces $((\mathsf{V}_j, (\gamma_{j,k})_{k\in\mathbb{Z}_{>0}}))_{j\in\mathbb{Z}_{>0}}$.

(i) A sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ ***converges*** to v_0 if there exists $m \in \mathbb{Z}_{>0}$ so that $v_j \in \mathsf{V}_m$, $j \in \mathbb{Z}_{>0}$, if $v_0 \in \mathsf{V}_m$, and if $(v_j)_{j\in\mathbb{Z}_{>0}}$ converges to v_0 in $(\mathsf{V}_m, (\gamma_{m,k})_{k\in\mathbb{Z}_{>0}})$.

(ii) A sequence $(v_j)_{j\in\mathbb{Z}_{>0}}$ is a ***Cauchy sequence*** if there exists $m \in \mathbb{Z}_{>0}$ so that $v_j \in \mathsf{V}_m$, $j \in \mathbb{Z}_{>0}$, and if $(v_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $(\mathsf{V}_m, (\gamma_{m,k})_{k\in\mathbb{Z}_{>0}})$.

(iii) The countable union space V is ***complete*** if every Cauchy sequence converges. •

Section 6.4

Linear maps between locally convex topological vector spaces

6.4.1 Duals of locally convex topological vector spaces

6.4.2 Reflexivity in locally convex topological vector spaces

6.4.3 Duals of Fréchet spaces

The dual space of an $\mathbb{F}$-normed vector space is simply the set of continuous linear $\mathbb{F}$-valued maps. The similar statement holds for Fréchet spaces, only now one has a notion of continuity adapted to the collection of seminorms.

6.4.1 Definition Let $(\mathsf{V}, (\gamma_j)_{j\in\mathbb{Z}_{>0}})$ be a countable multinormed $\mathbb{F}$-vector space.

(i) A linear map $\alpha\colon \mathsf{V} \to \mathbb{F}$ is ***continuous*** at $v_0 \in \mathsf{V}$ if for each $\epsilon > 0$ there exists a neighbourhood O of v_0 so that if $v \in O$ then $|\alpha(v) - \alpha(v_0)| < \epsilon$.

(ii) The set of continuous linear maps on V is the ***dual space*** for V and is denoted V'. •

Since $\mathscr{S}(\mathbb{R};\mathbb{F})$ and $\mathscr{E}(\mathbb{R};\mathbb{F})$ are Fréchet spaces, we may ask whether $\mathscr{S}'(\mathbb{R};\mathbb{F})$ and $\mathscr{E}'(\mathbb{R};\mathbb{F})$ are their duals, as defined above. The answer is, of course, yes, and is provided by the following general result.

6.4.2 Proposition *Let $(\mathsf{V}, (\gamma_j)_{j\in\mathbb{Z}_{>0}})$ be a countable multinormed $\mathbb{F}$-vector space and let $\alpha\colon \mathsf{V} \to \mathbb{F}$ be $\mathbb{F}$-linear. The following statements are equivalent:*

(i) α is continuous at v_0 for every $v_0 \in \mathsf{V}$;

(ii) α is continuous at $0 \in \mathsf{V}$;

(iii) if $(v_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence converging to $v_0 \in \mathsf{V}$ then the sequence $(\alpha(v_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\alpha(v_0)$.

Proof The implication (i) $\Longrightarrow$ (ii) is obvious. Now let $(v_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence converging to v_0 and suppose that α is continuous at 0. It is clear that the sequence $(v_j - v_0)_{j\in\mathbb{Z}_{>0}}$ converges to zero. This means that for every neighbourhood O of 0 there exists N sufficiently large that $v_j - v_0 \in O$ for $j \geq N$. Thus, for $\epsilon > 0$ we can choose N sufficiently large that $|\alpha(v_j - v_0)| < \epsilon$ for $j \geq N$, which shows that (ii) $\Longrightarrow$ (iii).

Finally, suppose that α is not continuous at $v_0 \in \mathsf{V}$. Then there exists $\epsilon > 0$ so that one cannot find a neighbourhood O of v_0 with the property that if $v \in O$ then $|\alpha(v) - \alpha(v_0)| < \epsilon$. By Proposition 6.3.2 let $(B_j)_{j\in\mathbb{Z}_{>0}}$ be a decreasing sequence of balloons with the property that every neighbourhood contains at least one of these balloons. Then take $v_j \in B_j$, noting that $|\alpha(v_j - v_0)| \geq \epsilon$. However, we also have the convergence of $(v_j)_{j\in\mathbb{Z}_{>0}}$ to v_0. This shows that (iii) cannot hold, and so completes the proof. ■

This shows that continuity of elements of $\mathscr{S}'(\mathbb{R};\mathbb{F})$ and $\mathscr{E}'(\mathbb{R};\mathbb{F})$ corresponds to a general notion of continuity associated with elements of the dual of a Fréchet space. Let us similarly put the notion of convergence in the spaces $\mathscr{S}'(\mathbb{R};\mathbb{F})$ and

$\mathscr{E}'(\mathbb{R};\mathbb{F})$ can be cast in our current general framework. We do this by making the dual of a countable multinormed vector space itself a multinormed vector space. We first make the fairly obvious observation that V' is an $\mathbb{F}$-vector space if one defines

$$(\alpha_1 + \alpha_2)(v) = \alpha_1(v) + \alpha_1(v), \quad (a\alpha)(v) = a(\alpha(v)) \tag{6.3}$$

for $\alpha, \alpha_1, \alpha_2 \in \mathsf{V}'$ and $a \in \mathbb{F}$.

6.4.3 Definition Let $(\mathsf{V}, (\gamma_j)_{j\in\mathbb{Z}_{>0}})$ be a countable multinormed vector space.

(i) For $v \in \mathsf{V}$ define a seminorm γ_v on V' by $\gamma_v(\alpha) = |\alpha(v)|$.

(ii) The ***weak multinorm*** on V' is the collection of seminorms $(\gamma_v)_{v\in\mathsf{V}}$. •

This gives on V' the same sort of structure as one has on V. One difference is that even though V possesses a countable multinorm, it is not the case that the weak multinorm is itself countable. Nevertheless, it is a multinorm, and all of the concepts associated with a multinorm apply to V'. Among these is the notion of convergence, Cauchy sequences, and completeness. The following result is a fundamental one in the theory of Fréchet spaces.

6.4.4 Theorem *If* $(\mathsf{V}, (\gamma_{\mathrm{j}})_{\mathrm{j}\in\mathbb{Z}_{>0}})$ *is a Fréchet space then* $(\mathsf{V}', (\gamma_{\mathrm{v}})_{\mathrm{v}\in\mathsf{V}})$ *is a complete multinormed space.*

Proof The proof here has essentially been done for the proof of Theorem IV-3.2.22. One need only transform the notation from $\mathscr{D}(\mathbb{R};\mathbb{F})$ with its seminorms $\delta_k(\phi) = \|\phi^{(k)}\|_\infty$, $k \in \mathbb{Z}_{>0}$, to the general notation of V with its seminorms γ_j, $j \in \mathbb{Z}_{>0}$. ■

From this result we see why it is that Theorems IV-3.3.15 and IV-3.7.13 essentially follow in the same manner as Theorem IV-3.2.22. Indeed, a general systematic treatment of distribution theory will begin with a treatment of Fréchet spaces. However, since it is possible to understand distributions without the general background, we take the approach of presenting the general theory as an "aside."

6.4.4 Duals of countable union spaces

Next we turn to defining continuous linear mappings from countable union spaces to $\mathbb{F}$. This will provide a systematic framework for understanding the space $\mathscr{D}'(\mathbb{R};\mathbb{F})$ of distributions.

6.4.5 Definition Let V be a countable union space defined by a collection $((\mathsf{V}_j, (\gamma_{j,k})_{k\in\mathbb{Z}_{>0}}))_{j\in\mathbb{Z}_{>0}}$ of countable multinormed vector spaces.

(i) A linear map $\alpha\colon \mathsf{V} \to \mathbb{F}$ is ***continuous*** if for every $j \in \mathbb{Z}_{>0}$ the restriction of α to V_j is continuous with respect to the multinorm $(\gamma_{j,k})_{k\in\mathbb{Z}_{>0}}$.

(ii) The set of continuous linear maps on V is the ***dual space*** for V and is denoted V'. •

Directly from Proposition 6.4.2 and the definition of convergence in a countable union space, we have the following result.

6.4.6 Proposition *Let* V *be a countable union space defined by a collection* $((\mathsf{V}_{\mathrm{j}}, (\gamma_{\mathrm{j,k}})_{\mathrm{k}\in\mathbb{Z}_{>0}}))_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *of countable multinormed vector spaces. For a linear map* $\alpha\colon \mathsf{V} \to \mathbb{F}$ *the following statements are equivalent:*

(i) *α is continuous;*

(ii) *if* $(\mathrm{v}_{\mathrm{j}})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *is a sequence converging in* V *then the sequence* $(\alpha(\mathrm{v}_{\mathrm{j}}))_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *converges in* $\mathbb{R}$.

Note that this then gives the notion of continuity of elements of $\mathscr{D}'(\mathbb{R};\mathbb{C})$ as continuity as elements in the dual of a countable union space. For a general countable union space V defined by a collection $((\mathsf{V}_j, (\gamma_{j,k})_{k\in\mathbb{Z}_{>0}}))_{j\in\mathbb{Z}_{>0}}$ of countable multinormed vector spaces, the dual V′ comes equipped with the natural notion of vector addition and scalar multiplication (see (6.3)). Also, we may render V′ a multinormed space in exactly the same manner as was done for the dual of a multinormed space.

6.4.7 Definition Let V be a countable union space defined by a collection $((\mathsf{V}_j, (\gamma_{j,k})_{k\in\mathbb{Z}_{>0}}))_{j\in\mathbb{Z}_{>0}}$ of countable multinormed vector spaces.

(i) For $v \in \mathsf{V}$ define a seminorm γ_v on V′ by $\gamma_v(\alpha) = |\alpha(v)|$.

(ii) The ***weak multinorm*** on V′ is the collection of seminorms $(\gamma_v)_{v\in\mathsf{V}}$. •

With this multinorm, one has the usual notions of convergence, Cauchy sequences, and convergence in V′. One also has the following result which follows directly from Theorem 6.4.4 and the definition of convergence in a countable union space.

6.4.8 Theorem *Let* V *be a countable union space defined by a collection* $((\mathsf{V}_{\mathrm{j}}, (\gamma_{\mathrm{j,k}})_{\mathrm{k}\in\mathbb{Z}_{>0}}))_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *of complete countable multinormed vector spaces. Then* V′ *is also complete.*

This, applied to the countable union space $\mathscr{D}(\mathbb{R};\mathbb{C})$, gives the completeness of $\mathscr{D}'(\mathbb{R};\mathbb{C})$.

Section 6.5

Examples of locally convex spaces

In Section 3.8 we gave a multitude of examples of Banach spaces, many of which will play an essential rôle in subsequent developments. However, there are also many cases where we will encounter natural spaces of signals that do not fall under any of the examples of Banach spaces. In this section we give some of the common locally convex spaces we will encounter subsequently. We shall restrict ourselves to a consideration only of spaces that are not already normed spaces, since we already have many of these at hand. However, it goes without saying that all of the normed vector spaces we have considered are also locally convex spaces.

Do I need to read this section? Some of the constructions here will be used in an essential way in some of our developments of system theory, especially in Chapter V-6. Thus it will be important to refer here for some of the facts we shall use later. •

6.5.1 Locally convex spaces of sequences

In Section 3.8.2 we introduced particular spaces of sequences, which we denoted by $\mathsf{c}_0(\mathbb{F})$ and $\ell^p(\mathbb{F})$, $p \in [1, \infty]$. These spaces of sequences were distinguished by their possessing natural norms that render them Banach spaces. Here we extend the analysis to the space $\mathbb{F}^\infty$ of all sequences (see Example I-4.5.2–3 for notation).

On the space F^∞ is all sequences in $\mathbb{F}$, we consider a few different families of seminorms, all indexed by a finite subset $K \subseteq \mathbb{Z}_{>0}$. The seminorms are

$$\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_{K,\infty} = \sup\{|a_j| \mid j \in K\},$$
$$\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_{K,p} = \left(\sum_{j\in K} |a_j|^p\right)^{1/p}.$$

This defines a family, one for each $p \in [1, \infty]$, of locally convex topologies for $\mathbb{F}^\infty$. The first observation we make is that all of the topologies defined by these norms as p varies are actually the same. To see this, we note that for a fixed finite subset $K \subseteq \mathbb{Z}_{>0}$, we have, for any $p, q \in [1, \infty]$,

$$c\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_{K,p} \le \|(a_j)_{j\in\mathbb{Z}_{>0}}\|_{K,q} \le c^{-1}\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_{K,p}$$

for some $c \in \mathbb{R}_{>0}$. This is just a consequence of the fact that all norms on finite-dimensional vector spaces agree by Theorem 3.1.15. Thus we do not need to make reference to p in the discussion of these topologies.

Let us give some properties of this topology for the space of sequences.

6.5.1 Theorem (Properties of the locally convex topological vector space $\mathbb{F}^\infty$) *For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, the locally convex topological vector space $\mathbb{F}^\infty$ has the following properties:*

(i) *Hausdorff;*
(ii) *metrisable;*
(iii) *complete;*
(iv) *separable;*
(v) *not normable.*

Proof For $m \in \mathbb{Z}_{>0}$, denote $K_m = \{1, \dots, m\}$. We claim that the topology of $\mathbb{F}^\infty$ is defined by the seminorms $\|\cdot\|_{K_m,p}$, $m \in \mathbb{Z}_{>0}$. To see this, let $K \subseteq \mathbb{Z}_{>0}$ be finite and let $m \in \mathbb{Z}_{>0}$ be large enough that $K \subseteq K_m$. Then, for any $(a_j)_{j\in\mathbb{Z}_{>0}} \in \mathbb{F}^\infty$, we have

$$\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_{K,p} \leq \|(a_j)_{j\in\mathbb{Z}_{>0}}\|_{K_m,p}.$$

This gives continuity of the identity map from $\mathbb{F}^\infty$ with the topology defined by the seminorms $(K_m)_{m\in\mathbb{Z}_{>0}}$ to $\mathbb{F}^\infty$ with its usual topology. We also have, for every $m \in \mathbb{Z}_{>0}$,

$$\|(a_j)_{j\in\mathbb{Z}_{>0}}\|_{K_m,p} \leq \|(a_j)_{j\in\mathbb{Z}_{>0}}\|_{K_m,p},$$

and this gives continuity of the identity map from $\mathbb{F}^\infty$ with its usual topology to $\mathbb{F}^\infty$ with the topology defined by the seminorms $(K_m)_{m\in\mathbb{Z}_{>0}}$. Thus the two topologies agree.

By Proposition 6.2.13 we easily conclude that $\mathbb{F}^\infty$ is Hausdorff. From we immediately conclude that $\mathbb{F}^\infty$ is metrisable. metrisation

To show that $\mathbb{F}^\infty$ is complete, it suffices to check that Cauchy sequences converge since the space it metrisable. Thus let $((a_{k,j})_{j\in\mathbb{Z}_{>0}})_{k\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathbb{F}^\infty$. Let $K \subseteq \mathbb{Z}_{>0}$ be finite. Then, by Proposition 6.2.12, we have that $((a_{k,j})_{j\in K})_k$ is a Cauchy sequence. This means that, for each $j \in K$, $(a_{k,j})_{k\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathbb{F}$ and so converges. Thus $((a_{k,j})_{j\in K})_k$ also converges, and so we conclude from Proposition 6.2.11 that $((a_{k,j})_{j\in\mathbb{Z}_{>0}})_{k\in\mathbb{Z}_{>0}}$ converges.

To see that $\mathbb{F}^\infty$ is separable, let $D \subseteq \mathbb{F}$ be a dense subset. Let $(a_j)_{j\in\mathbb{Z}_{>0}}$. For $m \in \mathbb{Z}_{>0}$, let $d_{m,1}, \dots, d_{m,m} \in D$ be such that

$$|a_j - d_{m,j}| < \frac{1}{m}, \qquad j \in \{1, \dots, m\}.$$

For $j > m$, take $d_{m,j} = 0$. We claim that the sequence $((d_{m,j})_{j\in\mathbb{Z}_{>0}})_{m\in\mathbb{Z}_{>0}}$ converges to $(a_j)_{j\in\mathbb{Z}_{>0}}$. Let $K \subseteq \mathbb{Z}_{>0}$ be finite and let $\epsilon \in \mathbb{R}_{>0}$. Let $m \in \mathbb{Z}_{>0}$ be such that $K \subseteq K_m$ and $\frac{1}{m} < \epsilon$. Then, for $k \geq m$,

$$|a_j - d_{k,j}| < \frac{1}{k} \leq \frac{1}{m} < \epsilon, \qquad j \in K_k \supseteq K_m \supseteq K.$$

By Proposition 6.2.11, this gives the claimed convergence. This shows that the set of sequences with entries in D are dense in $\mathbb{F}^\infty$. Moreover, the set of all such sequences is countable by Proposition I-1.7.16.

Finally, to show that $\mathbb{F}^\infty$ is not normable, by Theorem 6.3.1 it suffices to show that every convex neighbourhood of 0 is unbounded. Let U be a convex neighbourhood of 0 and let $j_1, \dots, j_m \in \mathbb{Z}_{>0}$ and let $r_1, \dots, r_m \in \mathbb{R}_{>0}$ be such that

$$\bigcap_{l=1}^{m} \{(a_j)_{j\in\mathbb{Z}_{>0}} \in \mathbb{F}^\infty \mid p_{K_{j_l}}((a_j)_{j\in\mathbb{Z}_{>0}}) < r_{j_l}\} \subseteq U.$$

Let $n \in \mathbb{Z}_{>0}$ be large enough that

$$K_{j_l} \subseteq K_n, \qquad l \in \{1, \dots, m\},$$

and let $r < r_j$, $j \in \{1, \dots, m\}$. Let $M \in \mathbb{R}_{>0}$ and let $(a_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence such that $a_1, \dots, a_n < r$ and such that $a_{n+1} \geq M$. Then $(a_j)_{j \in \mathbb{Z}_{>0}} \in U$ and

$$\|(a_j)_{j \in \mathbb{Z}_{>0}}\|_{K_{n+1},p} \geq M.$$

As this construction can be made for every $M \in \mathbb{R}_{>0}$, we conclude from Proposition 6.2.10 that U is not bounded. Thus $\mathbb{F}^\infty$ is not normable. ■

Let us relate the preceding topology on the space of all sequences to that for the special classes of sequences considered in Section 3.8.2.

6.5.2 Proposition (Continuity of inclusions of normed vector spaces of sequences) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. The inclusions of the normed vector spaces $\mathbb{F}_0^\infty$, $\mathsf{c}_0(\mathbb{F})$, and $\ell^{\mathrm{p}}(\mathbb{F})$, $\mathrm{p} \in [1, \infty]$, in $\mathbb{F}^\infty$ are continuous.*

Proof Let us consider first the inclusions of $\mathbb{F}_0^\infty$, $\mathsf{c}_0(\mathbb{F})$, and $\ell^\infty(\mathbb{F})$, as the proof of the continuity of the inclusions is the same for all cases. We shall use the seminorms $\|\cdot\|_{K,\infty}$ to prove continuity, noting that the choice of seminorm is immaterial in this case. Let $K \subseteq \mathbb{Z}_{>0}$ be finite and let $(a_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in one of the spaces $\mathbb{F}_0^\infty$, $\mathsf{c}_0(\mathbb{F})$, or $\ell^\infty(\mathbb{F})$. Then

$$\|(a_j)_{j \in \mathbb{Z}_{>0}}\|_{K,\infty} = \sup\{|a_j| \mid j \in K\} \leq \|(a_j)_{j \in \mathbb{Z}_{>0}}\|_\infty,$$

which gives the desired continuity by Proposition 6.2.9. The proof of continuity of the inclusion of $\ell^p(\mathbb{F})$, $p \in [1, \infty)$, follows similarly, using the seminorms $\|\cdot\|_{K,p}$. ■

6.5.2 Locally convex spaces of continuous functions on $\mathbb{R}$

We next consider general classes of continuous functions defined on an interval $I \subseteq \mathbb{R}$. Thus we consider the space

$$\mathsf{C}^0(I; \mathbb{F}) = \{f \in \mathbb{F}^I \mid f \text{ is continuous}\}.$$

We consider the locally convex topology for $\mathsf{C}^0(I; \mathbb{F})$ associated with the seminorms

$$\|f\|_{K,\infty} = \sup\{|f(x)| \mid x \in K\},$$

for a compact subinterval $K \subseteq I$.

Let us give some properties of this topology for the space of sequences.

6.5.3 Theorem (Properties of the locally convex topological vector space $\mathsf{C}^0(\mathrm{I}; \mathbb{F})$) *For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, the locally convex topological vector space $\mathsf{C}^0(\mathrm{I}; \mathbb{F})$ has the following properties:*

(i) *Hausdorff;*

(ii) *metrisable;*

(iii) *complete;*

(iv) *separable;*

(v) *normable if and only if* I *is compact.*

Proof In the case that I is compact, then we claim that $\mathsf{C}^0(I;\mathbb{F})$ is the same, as a topological vector space, as the normed vector space $(\mathsf{C}^0(I;\mathbb{F}), \|\cdot\|_\infty)$ discussed in Section 3.8.4. To see this, let $K \subseteq I$ be compact and note that $\|f\|_{K,\infty} \le \|f\|_\infty$. By Proposition 6.2.9, this gives continuity of the identity map as a map from the normed vector space into the locally convex topological vector space. On the other hand, the obvious inequality $\|f\|_\infty \le \|f\|_{I,\infty}$ gives continuity of the identity map from the locally convex topological vector space into the normed vector space. Thus the two topologies agree.

We now have that, when I is compact, the space $\mathsf{C}^0(I;\mathbb{F})$ is Hausdorff (because it is normed), metrisable (because it is normed), complete (Theorem 3.8.31), separable (Proposition 3.8.34) and normable (it is normed).

Now, for the remainder of the proof we assume that I is not compact.

Let $(K_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of compact subintervals in I such that $K_j \subseteq \operatorname{int}(K_{j+1})$, $j \in \mathbb{Z}_{>0}$, and $I = \cup_{j\in\mathbb{Z}_{>0}} K_j$. We claim that the seminorms $(\|\cdot\|_{K_j,\infty})_{j\in\mathbb{Z}_{>0}}$ define the topology of $\mathsf{C}^0(I;\mathbb{F})$. Let $K \subseteq I$ and let K_j be such that $K \subseteq K_j$. The inequality $\|f\|_{K,p} \le \|f\|_{K_j,p}$ for $f \in \mathsf{C}^0(I;\mathbb{F})$ shows that the identity mapping from $\mathsf{C}^0(I;\mathbb{F})$ with the claimed topology into the same space with its usual topology is continuous. The inequality $\|f\|_{K_j,\infty} \le \|f\|_{K_j,\infty}$ for $f \in \mathsf{C}^0(I;\mathsf{V})$ gives the continuity of the identity mapping from $\mathsf{C}^0(I;\mathsf{V})$ with its usual topology to the same space with the claimed topology. Thus the seminorms $(\|\cdot\|_{K_j,\infty})_{j\in\mathbb{Z}_{>0}}$ define the topology of $\mathsf{C}^0(I;\mathbb{F})$.

By Proposition 6.2.13 we easily conclude that $\mathsf{C}^0(I;\mathbb{F})$ is Hausdorff. From we immediately conclude that $\mathsf{C}^0(I;\mathbb{F})$ is metrisable. metrisation

To show that $\mathsf{C}^0(I;\mathbb{F})$ is complete, it suffices to check that Cauchy sequences converge since the space it metrisable. Thus let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathsf{C}^0(I;\mathbb{F})$. Let $K \subseteq I$ be compact. Then, by Proposition 6.2.12, we have that $(f_j|K)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathsf{C}^0(K;\mathbb{F})$. By Theorem 3.8.31, this means that $(f_j|K)_{j\in\mathbb{Z}_{>0}}$ converges to $f_K \in \mathsf{C}^0(K;\mathbb{F})$. Moreover, for two compact subintervals $K, L \subseteq I$, we have $f_K|K\cap L = f_L|K\cap L$. Therefore, there exists $f \in \mathsf{C}^0(I;\mathbb{F})$ with the property that $(f_j|K)_{j\in\mathbb{Z}_{>0}}$ converges uniformly to $f|K$ for every compact interval $K \subseteq I$. That is, $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges to f in the locally convex topology.

To see that $\mathsf{C}^0(I;\mathbb{F})$ is separable, for $m \in \mathbb{Z}_{>0}$, let $D_m \subseteq \mathsf{C}^0(K_m;\mathbb{F})$ be a dense subset by Proposition 3.8.34. We can think of $f \in D_m$ as a function on I by extending it arbitrarily to a continuous function on all of I. For $f \in \mathsf{C}^0(I;\mathbb{F})$ and $m \in \mathbb{Z}_{>0}$, let $f_m \in D_m$ be such that

$$\|f - f_m\|_{K_m,\infty} < \frac{1}{m}.$$

We claim that $(f_m)_{m\in\mathbb{Z}_{>0}}$ converges to f. Let $K \subseteq I$ and $\epsilon \in \mathbb{R}_{>0}$. Let $m \in \mathbb{Z}_{>0}$ be such that $K \subseteq K_m$ and $\frac{1}{m} < \epsilon$. Then, for $k \ge m$,

$$\|f - f_k\|_{K,\infty} < \frac{1}{k} < \frac{1}{m} < \epsilon,$$

since $K \subseteq K_m \subseteq K_k$. This gives convergence of $(f_m)_{m\in\mathbb{Z}_{>0}}$ to f. Since $\cup_{m\in\mathbb{Z}_{>0}} D_m$ is countable by Proposition I-1.7.16, we conclude that $\mathsf{C}^9(I;\mathbb{F})$ is separable, Finally, to show that $\mathbb{F}^\infty$ is not normable when I is not compact, by Theorem 6.3.1 it suffices to show that every convex neighbourhood of 0 is unbounded. Let U be a convex neighbourhood of 0 and let $j_1,\dots,j_m \in \mathbb{Z}_{>0}$ and let $r_1,\dots,r_m \in \mathbb{R}_{>0}$ be such that

$$\bigcap_{l=1}^{m}\{f \in \mathsf{C}^0(I;\mathbb{F}) \mid p_{K_{j_l}}(f) < r_{j_l}\} \subseteq U.$$

Let $n \in \mathbb{Z}_{>0}$ be large enough that

$$K_{j_l} \subseteq K_n, \qquad l \in \{1,\dots,m\},$$

and let $r < r_j$, $j \in \{1,\dots,m\}$. Let $M \in \mathbb{R}_{>0}$ and let $f \in \mathsf{C}^0(I;\mathbb{F})$ be such that $|f(x)| < r$ for $x \in K_n$ and such that there exists $x \in K_{n+1} \setminus K_n$ such that $|f(x)| \geq M$. Then $f \in U$ and

$$\|f\|_{K_{n+1},\infty} \geq M.$$

As this construction can be made for every $M \in \mathbb{R}_{>0}$, we conclude from Proposition 6.2.10 that U is not bounded. Thus $\mathsf{C}^0(I;\mathbb{F})$ is not normable. ■

Let us relate the preceding topology on the space of all sequences to that for the special classes of continuous functions sequences considered in Section 3.8.4.

6.5.4 Proposition (Continuity of inclusions of normed vector spaces of continuous functions) *Let* $\mathbb{F} \in \{\mathbb{R},\mathbb{C}\}$*. The inclusions of the normed vector spaces* $\mathsf{C}^0_{\mathrm{cpt}}(\mathrm{I};\mathbb{F})$*,* $\mathsf{C}^0_0(\mathrm{I};\mathbb{F})$*, and* $\mathsf{C}^0_{\mathrm{bdd}}(\mathrm{I};\mathbb{F})$ *in* $\mathsf{C}^0(\mathrm{I};\mathbb{F})$ *are continuous.*

Proof The proof of the continuity of the inclusions looks the same for all cases. Let $K \subseteq I$ be compact and let $f\colon I \to \mathbb{F}$ be a function in one of the spaces named in the proposition. Then

$$\|f\|_{K,\infty} = \sup\{|f(x)| \mid x \in K\} \leq \|f\|_\infty,$$

and this gives continuity of the inclusion by Proposition 6.2.9. ■

6.5.3 Locally convex spaces of continuous functions on metric spaces

6.5.4 Locally convex spaces of locally integrable functions on $\mathbb{R}$

Now we discuss spaces of locally integrable functions, extending our consideration of Banach spaces of such functions in Section 3.8.7. For $p \in [1,\infty]$ and for an interval $I \subseteq \mathbb{R}$, we denote

$$\mathsf{L}^{(p)}_{\mathrm{loc}}(I;\mathbb{F}) = \{f \in \mathbb{F}^I \mid f|K \in \mathsf{L}^{(p)}(K;\mathbb{F}) \text{ for every compact subinterval } K \subseteq I\}.$$

Just as we did for the L^p-spaces, we can quotient out the subspace of functions that are almost everywhere zero:

$$Z^p(I;\mathbb{F}) = \{f \in \mathsf{L}^p_{\mathrm{loc}}(I;\mathbb{F}) \mid \lambda(\{x \in I \mid f(x) \neq 0\}) = 0\}$$

and define

$$\mathsf{L}^p_{\mathrm{loc}}(I;\mathbb{F}) = \mathsf{L}^{(p)}_{\mathrm{loc}}(I;\mathbb{F})/Z^p(I;\mathbb{F}).$$

We render these spaces locally convex topological vector spaces by defining seminorms, one for each compact subinterval $\mathbb{K} \subseteq \mathbb{T}$:

$$\|f\|_{K,p} = \left(\int_{\mathbb{K}} |f|^p \, d\lambda\right)^{1/p}, \qquad p \in [1,\infty),$$
$$\|f\|_{K,\infty} = \operatorname{ess\,sup}\{|f(x)| \mid x \in K\}.$$

Let us record the properties of these topological vector spaces.

6.5.5 Theorem (Character of signal spaces) *For $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ and for $p \in [1,\infty]$, the locally convex topological vector space $\mathsf{L}^p_{\mathrm{loc}}(I;\mathbb{F})$ has the following properties:*

(i) *Hausdorff;*
(ii) *metrisable;*
(iii) *complete;*
(iv) *separable if $p \neq \infty$;*
(v) *not separable if $p = \infty$;*
(vi) *normable if I is compact;*
(vii) *not normable if I is not compact.*

Proof The proof of Theorem 6.5.3 applies to prove all assertions except that that $\mathsf{L}^\infty_{\mathrm{loc}}(I;\mathbb{F})$ is not separable. When I is compact, this assertion is Proposition 3.8.49. When I is not compact, the assertion follows since $\mathsf{L}^\infty(I;\mathbb{F})$ contained the nonseparable subspace $\mathsf{L}^\infty(K;\mathbb{F})$ for any compact subinterval $K \subseteq I$. ■

Let us relate the preceding topology on the space of all sequences to that for the special classes of integrable functions considered in Section 3.8.7.

6.5.6 Proposition (Continuity of inclusions of normed vector spaces of integrable functions) *Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. The inclusion of the normed vector spaces $\mathsf{L}^p(I;\mathbb{F})$ in $\mathsf{L}^p_{\mathrm{loc}}(I;\mathbb{F})$ are continuous for every $p \in [1,\infty]$.*

Proof The proof of the continuity of the inclusions looks the same for all cases. Thus we fix $p \in [1,\infty]$. Let $K \subseteq I$ be compact and let $f \in \mathsf{L}^p(I;\mathbb{F})$. Then

$$\|f\|_{K,p} = \begin{cases} \left(\int_K |f|^p \, d\lambda\right)^{1/p}, & p \in [1,\infty), \\ \sup\{|f(x)| \mid x \in K\}, & p = \infty \end{cases} \leq \|f\|_p,$$

and this gives continuity of the inclusion by Proposition 6.2.9. ■

6.5.5 Distributions as elements of locally convex topological vector spaces

Throughout this chapter we have made use of some structure on our sets $\mathscr{D}$, $\mathscr{S}$, and $\mathscr{E}$ of test signals, but without really saying what this structure is. That there is some underlying structure here can be seen by the manner in which one can set us the proofs of Theorems IV-3.3.15 and IV-3.7.13 for convergence in $\mathscr{S}'(\mathbb{R};\mathbb{F})$ and $\mathscr{E}'(\mathbb{R};\mathbb{F})$ to follow is the same manner as the proof of Theorem IV-3.2.22, that for convergence in $\mathscr{D}'(\mathbb{R};\mathbb{F})$. In this section we explore the structure that leads to these sorts of developments. Readers not interested in the details of the topology of spaces of distributions can easily forgo this section on a first reading.

6.5.7 Examples

1. Let $I \subseteq \mathbb{R}$ be a continuous time-domain. On $\mathscr{D}(I;\mathbb{F})$ define seminorms $\delta_{I,k}(\cdot)$ by
$$\delta_{I,k}(\phi) = \sup_{t\in I}|\phi^{(k)}(t)|.$$
2. On $\mathscr{S}(\mathbb{R};\mathbb{F})$ we define seminorms $\sigma_{m,k}$, $m,k \in \mathbb{Z}_{\geq 0}$, by
$$\sigma_{m,k}(\phi) = \sup_{t\in\mathbb{R}}|(1+t^2)^m\phi^{(k)}(t)|.$$
3. On $\mathscr{E}(\mathbb{R};\mathbb{C})$ and for $K \subseteq \mathbb{R}$ a compact set, define seminorms $\varepsilon_{K,k}(\cdot)$, $k \in \mathbb{Z}_{\geq 0}$, by
$$\varepsilon_{K,k}(\phi) = \sup_{t\in K}|\phi^{(k)}(t)|.$$ •

6.5.8 Examples (Example 6.5.7 cont'd)

1. On $\mathscr{D}(I;\mathbb{F})$ the sequence of seminorms $(\delta_{I,k})_{k\in\mathbb{Z}_{\geq 0}}$ is clearly separating since $\delta_{I,0}$ is a norm. Thus we have a countable multinorm on $\mathscr{D}(I;\mathbb{F})$. If I is compact, then convergence in the multinorm is the same as the notion of convergence defined in $\mathscr{D}(\mathbb{R};\mathbb{F})$. However, if I is not compact, this is no longer the case. Indeed, suppose that $I = \mathbb{R}$ and consider the sequence $(\phi_j)_{j\in\mathbb{Z}_{>0}}$ defined by
$$\phi_j(t) = \begin{cases} e^{\frac{j^2}{t^2-j^2}}e^{1-t^2}, & |t| < j \\ 0, & \text{otherwise.}\end{cases}$$
One may verify that the sequence converges uniformly to $\phi(t) = e^{-t^2}$, and that all derivatives similarly converge uniformly to those of ϕ. Note that $\phi \notin \mathscr{D}(\mathbb{R};\mathbb{F})$. This shows two things: (1) convergence in $\mathscr{D}(\mathbb{R};\mathbb{F})$ with respect to the multinorm $(\delta_{\mathbb{R},k})_{k\in\mathbb{Z}_{\geq 0}}$ is not the same as convergence in $\mathscr{D}(\mathbb{R};\mathbb{F})$ as per Definition IV-3.2.4; (2) the multinormed space $(\mathscr{D}(\mathbb{R};\mathbb{F}),(\delta_{\mathbb{R},j})_{j\in\mathbb{Z}_{\geq 0}})$ is not a Fréchet space.
2. Now let us consider the multinormed vector space $(\mathscr{S}(\mathbb{R};\mathbb{F});(\sigma_{m,k})_{m,k\in\mathbb{Z}_{\geq 0}})$. It is easily verified that the collection $(\sigma_{m,k})_{m,k\in\mathbb{Z}_{\geq 0}}$ of seminorms is separating, so this is indeed a multinormed vector space. In this case it is evident from Proposition 6.2.11 that convergence in the multinorm is equivalent to the usual notion of convergence in $\mathscr{S}(\mathbb{R};\mathbb{F})$. One can also show easily that the resulting space is a Fréchet space.

3. If $K \subseteq \mathbb{R}$ is compact, note that the family $(\varepsilon_{K,k})_{k\in\mathbb{Z}_{\geq 0}}$ is not separating on $\mathscr{E}(\mathbb{R};\mathbb{F})$. Indeed, if $\phi \in \mathscr{E}(\mathbb{R};\mathbb{F})$ has support contained in $\mathbb{R}\setminus K$ then we have $\varepsilon_{K,k}(\phi) = 0$ although ϕ is nonzero. However, if we take the collection of seminorms allowed by varying K over all compact subsets, the resulting collection *is* separating. Thus the set $(\varepsilon_{K,k})_{k\in\mathbb{Z}_{\geq 0}, K \text{ compact}}$ is a multinorm for $\mathscr{E}(\mathbb{R};\mathbb{F})$ that is uncountable. Furthermore, one can verify that convergence in this multinorm is exactly the usual notion of convergence in $\mathscr{E}(\mathbb{R};\mathbb{C})$. One can show that the Cauchy sequences converge, and so the resulting space is complete, but not Fréchet since the multinorm is not countable.
4. One can, however, regard $\mathscr{E}(\mathbb{R};\mathbb{F})$ as a Fréchet space as follows. Define $I_m = [-m, m]$ and define a countable collection of seminorms on $\mathscr{E}(\mathbb{R};\mathbb{F})$ by $(\varepsilon_{I_m,k})_{m\in\mathbb{Z}_{>0},k\in\mathbb{Z}_{\geq 0}}$. This is a separating collection of seminorms, and so is a countable multinorm. We also claim that a sequence converges in this multinorm if and only if it converges in the usual sense in $\mathscr{E}(\mathbb{R};\mathbb{F})$. It is obvious that if a sequence $(\phi_j)_{j\in\mathbb{Z}_{>0}}$ converges in $\mathscr{E}(\mathbb{R};\mathbb{F})$ then it converges in the given multinorm. Suppose then that a sequence converges in the given multinorm and let K be a compact set. By choosing m sufficiently large that $K \subseteq I_m$ we then have

$$\lim_{j\to\infty} \sup_{t\in K} |\phi_j^{(k)}| \leq \lim_{j\to\infty} \sup_{t\in I_m} |\phi_j^{(k)}| = 0,$$

giving convergence in $\mathscr{E}(\mathbb{R};\mathbb{C})$. •

6.5.9 Example (Example 6.5.7 cont'd) Define $I_j = [-j, j]$ and note the following facts:

(a) for each $j \in \mathbb{Z}_{>0}$, $\mathscr{D}(I_j;\mathbb{F})$ can be thought of as a subspace of $\mathscr{D}(\mathbb{R};\mathbb{F})$ by extending the signals to be zero outside I_j;

(b) $\mathscr{D}(I_j;\mathbb{F}) \subseteq \mathscr{D}(I_{j+1};\mathbb{F})$;

(c) $\mathscr{D}(\mathbb{R};\mathbb{F}) = \bigcup_{j\in\mathbb{Z}_{>0}} \mathscr{D}(I_j,\mathbb{F})$;

(d) the multinorm $(\delta_{I_{j+1},k})_{k\in\mathbb{Z}_{\geq 0}}$ restricted to $\mathscr{D}(I_j;\mathbb{F})$ is weaker then the multinorm $(\delta_{I_{j+1},k})_{k\in\mathbb{Z}_{\geq 0}}$.

This shows that $\mathscr{D}(\mathbb{R};\mathbb{C})$ is a countable union space defined by the collection $((\mathscr{D}_{I_j}, (\delta_{I_j,k})_{k\in\mathbb{Z}_{\geq 0}}))_{j\in\mathbb{Z}_{>0}}$ of countable multinormed spaces. •

Exercises

6.5.1 Show that a sequence $((a_{k,j})_{j\in\mathbb{Z}_{>0}})_{k\in\mathbb{Z}_{>0}}$ converges to $(a_j)_{j\in\mathbb{Z}_{>0}}$ if and only if $\lim_{k\to\infty} a_{k,j} = a_j$ for each $j \in \mathbb{Z}_{>0}$.

6.5.2 Determine whether the following sequences $((a_{k,j})_{j\in\mathbb{Z}_{>0}})_{k\in\mathbb{Z}_{>0}}$ of sequences converge in $\mathbb{R}^\infty$ and determine the limit sequence when it exists. Here are the sequences of sequences:

(a) $a_{k,j} = \begin{cases} 1, & j \leq k, \\ 0, & j > k; \end{cases}$

(b) $a_{k,j} = (-1)^{k+j}$;
(c) $a_{k,j} = j^k$;
(d) $a_{k,j} = k^j$;
(e) $a_{k,j} = \begin{cases} \frac{k-j+1}{k}, & j \leq k, \\ 0, & j > k. \end{cases}$

6.5.3 Determine whether the following sequences $(f_j)_{j\in\mathbb{Z}_{>0}}$ of functions in $\mathsf{C}^0(\mathbb{R};\mathbb{R})$ converge and determine the limit function when it exists. Here are the sequences of functions:

(a) $f_j(x) = \begin{cases} 0, & x \in [-j, j], \\ x - j, & x > j, \\ -x - j, & x < -j; \end{cases}$
(b) $f_j(x) = x^j$;
(c) $f_j(x) = \begin{cases} \sin(x), & x \in [0, 2j\pi], \\ 0, & \text{otherwise}; \end{cases}$
(d) $f_j(x) = |x|^{1/2j}$.

6.5.4 Determine for which, if any, $p \in [1, \infty]$ the following sequences $(f_j)_{j\in\mathbb{Z}_{>0}}$ of functions in $\mathsf{L}^p_{\mathrm{loc}}(\mathbb{R};\mathbb{R})$ converge and determine the limit function when it exists. Here are the sequences of functions:

(a) $f_j(x) = \begin{cases} 0, & x \in [-j, j], \\ x - j, & x > j, \\ -x - j, & x < -j; \end{cases}$
(b) $f_j(x) = x^j$;
(c) $f_j(x) = \begin{cases} \sin(x), & x \in [0, 2j\pi], \\ 0, & \text{otherwise}; \end{cases}$
(d) $f_j(x) = |x|^{1/2j}$;
(e) $f_j(x) = \begin{cases} (t - \frac{1}{j})^{-1/2}, & t \in (\frac{1}{j}, \infty), \\ 0, & \text{otherwise}. \end{cases}$

6.5.5 Let $I \subseteq \mathbb{R}$ be an interval.

(a) Show that a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{C}^0(I;\mathbb{R})$ converges pointwise if it converges with respect to the family of seminorms $\|\cdot\|_{K,\infty}$, $K \subseteq I$ a compact subinterval.
(b) Show that a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{C}^0(I;\mathbb{R})$ converges with respect to the family of seminorms $\|\cdot\|_{K,\infty}$, $K \subseteq I$ a compact subinterval, if it converges uniformly.
(c) Give an example of an interval $I \subseteq \mathbb{R}$ and a sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathsf{C}^0(I;\mathbb{R})$ that converges with respect to the family of seminorms $\|\cdot\|_{K,\infty}$, $K \subseteq I$ a compact subinterval, but does not converge uniformly.

6.5.6 Show that the following inclusions are continuous:

(a) the inclusion of $\mathsf{c}_0(\mathbb{F})$ in $\mathbb{F}^\infty$;

(b) the inclusion of $\ell^p(\mathbb{F})$ in $\mathbb{F}^\infty$, $p \in [1, \infty]$.

6.5.7 Let $I \subseteq \mathbb{R}$ be an interval. Show that the following inclusions are continuous:

(a) the inclusion of $\mathsf{C}^0_0(I; \mathbb{F})$ in $\mathsf{C}^0(I; \mathbb{F})$;

(b) the inclusion of $\mathsf{C}^0_{\mathrm{bdd}}(I; \mathbb{F})$ in $\mathsf{C}^0(I; \mathbb{F})$.

6.5.8 Let $I \subseteq \mathbb{R}$ be an interval and let $p \in [1, \infty]$. Show that the inclusion of $\mathsf{L}^p(I; \mathbb{F})$ in $\mathsf{L}^p_{\mathrm{loc}}(I; \mathbb{F})$ is continuous.

Chapter 7

Hardy spaces

In Chapter IV-9 we will present the Laplace transform, which can, in some sense, be thought of as a generalisation of the CCFT where the transformed signal is not a function of a real variable, but of a complex variable. One of the difficulties of the Laplace transform is to understand it as a map between spaces, as one has to give properties to the space of complex functions in which the Laplace transform takes its values. An entirely related issue arises in Chapter V-7 when we consider transfer functions for linear systems. A transfer function will be a function of a complex variable, and one wishes to have at hard spaces where these transfer functions live. It is with these objectives in mind that we discuss in this chapter a collection of spaces of functions of a complex variable known as "Hardy spaces.[1]" These spaces are tightly connected with the various L^p-spaces considered in Section 3.8.7.

We consider Hardy spaces in two flavours, one flavour defined on vertical strips in the complex plane and the other defined on annuli in the complex plane. Both will be useful to us; the first for continuous-time systems the second for discrete-time systems.

Do I need to read this chapter? This is a chapter that can be read when one needs to understand deeper properties of Laplace transforms and transfer functions in Chapters IV-9 and V-7. •

Contents

7.1 Preliminary constructions and results . . . 661
7.1.1 The Hardy–Littlewood maximal function . . . 661
7.1.2 Nontangential limits . . . 668
7.1.3 Constructions with subharmonic functions . . . 670
7.2 Hardy spaces of harmonic functions defined on vertical strips . . . 671
7.2.1 Definitions . . . 671

[1]After Godfrey Harold Hardy (1877–1947). Hardy was a leading mathematician of his time, and made substantial contributions to analysis and number theory. Apart from his mathematical work, he is known for a number of other contributions and aspects of his life, including having Srinivasa Ramanujan as protégé, writing *A Mathematician's Apology* describing the inner life of a mathematician, and his love for cricket.

7.2.2 Poisson integral representations of harmonic functions on vertical strips 672
7.2.3 Nontangential limits for the harmonic Hardy spaces on vertical strips . 686
Exercises 693
7.3 Hardy spaces of harmonic functions defined on annuli 694
7.3.1 Definitions 694
7.3.2 Poisson integral representations of harmonic functions on annuli 696
7.3.3 Nontangential limits for the harmonic Hardy spaces on annuli 696
7.4 Hardy spaces of holomorphic functions defined on vertical strips 697
7.4.1 Definitions 697
7.4.2 Poisson integral representations of holomorphic functions on vertical strips 698
7.4.3 Banach spaces of holomorphic functions on vertical strips 698
7.4.4 Locally convex topological vector spaces of holomorphic functions on vertical strips 701
7.5 Hardy spaces of holomorphic functions defined on annuli 704
7.5.1 Definitions 704
7.5.2 Poisson integral representations of holomorphic functions on annuli . . 705
7.5.3 Banach spaces of holomorphic functions on annuli 705
7.5.4 Locally convex topological vector spaces of holomorphic functions on annuli 705
7.5.5 Hardy spaces of vector-valued holomorphic functions defined on annuli 705

Section 7.1

Preliminary constructions and results

Many of the constructions we shall make for Hardy spaces require some substantial buildup of material that is not directly related to the principal constructions. In this section we shall gather this preliminary material.

Do I need to read this section? If you want to understand the material in this chapter in depth and detail, you will have, at some point, to wrestle with the material in this section. However, at a first reading, the section can be skipped, and then referred to at the points where the results are required. •

7.1.1 The Hardy–Littlewood maximal function

Our discussion begins with the following definition; if X is a set and if $\mathbb{F} \in \mathbb{R}, \mathbb{C}\}$, we denote by $\mathscr{M}(X;\mathbb{F})$ the set of $\mathbb{F}$-valued (not necessarily positive) measures on X.

7.1.1 Definition (Hardy–Littlewood maximal function) Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

(i) If $f \in \mathsf{L}^1_{\mathrm{loc}}(\mathbb{R};\mathbb{F})$, the ***Hardy–Littlewood maximal function*** of f is the function $\mathsf{M}f\colon \mathbb{R} \to \mathbb{R}$ defined by

$$\mathsf{M}f(t) = \sup\left\{\frac{1}{\lambda(I)}\int_I |f(\tau)|\,\mathrm{d}\tau \;\middle|\; I \in \mathscr{I},\ t \in I\right\}.$$

(ii) If $\mu \in \mathscr{M}(\mathbb{R};\mathbb{F})$, the ***Hardy–Littlewood maximal function*** of μ is the function $\mathsf{M}\mu\colon \mathbb{R} \to \mathbb{R}$ defined by

$$\sup\left\{\frac{|\mu|(I)}{\lambda(I)} \;\middle|\; I \in \mathscr{I},\ t \in I\right\}$$ •

We shall be interested in the properties of the assignment $f \mapsto \mathsf{M}f$. To understand this in the case when $f \in \mathsf{L}^1(\mathbb{R};\mathbb{F})$, we introduce the notion of "weak-L^p spaces." This we do with the following quite general definitions.

7.1.2 Definition (Distribution function, weak-L^p space) Let $(X, \mathscr{A}, \mu)$ be a measure space and let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

(i) If $f \in \mathsf{L}^0(X;\mathbb{F})$, the ***distribution function*** of f is

$$\begin{aligned} m_f\colon \mathbb{R}_{>0} &\to \overline{\mathbb{R}}_{\geq 0} \\ \sigma &\mapsto \mu(\{x \in X \mid |f(x)| > \sigma\}). \end{aligned}$$

(ii) A function $f\colon X \to \mathbb{F}$ is ***weak*-L^p** if there exists $C \in \mathbb{R}_{>0}$ such that

$$m_f(\sigma) \le \frac{C}{\sigma^p}, \qquad \sigma \in \mathbb{R}_{>0}.$$

By $\mathsf{L}^{\mathrm{w},p}(X;\mathbb{F})$ we denote the set of weak-L^p functions.

(iii) For $f \in \mathsf{L}^{\mathrm{w},p}(X;\mathbb{F})$, we denote

$$\|f\|_{\mathrm{w},p} = \sup\{\sigma m_f(\sigma)^{1/p} \mid \sigma \in \mathbb{R}_{>0}\}. \qquad \bullet$$

We comment that $\|\cdot\|_{\mathrm{w},p}$ is not a norm as it does not satisfy the triangle inequality. Nevertheless, it is common practice to treat it as one would treat a norm. It may not be surprising to know that L^p-functions are weak-L^p. This follows from a slight generalisation of Chebychev's inequality, presented as Corollary 2.7.36.

7.1.3 Lemma (Generalisation of Chebychev's inequality) *Let* $(X, \mathscr{A}, \mu)$ *be a measure space and let* $\mathbb{F} \in \{\mathbb{R}; \mathbb{C}\}$*. If* $f \in \mathsf{L}^p(X;\mathbb{F})$*, then* $f \in \mathsf{L}^{\mathrm{w},p}(X;\mathbb{F})$*; moreover,*

$$m_f(\sigma) \le \frac{\|f\|_p^p}{\sigma^p}.$$

Proof For $\sigma \in \mathbb{R}_{>0}$, denote

$$E_{f,\sigma} = \{x \in X \mid |f(x)| > \sigma\}$$

and note that

$$\sigma^p \mu(E_{f,\sigma}) \le \int_{E_{f,\sigma}} |f(x)|^p \,\mathrm{d}\mu(x) \le \|f\|_p^p. \qquad \blacksquare$$

The following important result will be used subsequently in this section, and is also useful in Chapter IV-6 in our discussion of the continuous-continuous Fourier transform.

7.1.4 Theorem (Marcinkiewicz[2] Interpolation Theorem) *Let* $(X, \mathscr{A}, \mu)$ *and* $(Y, \mathscr{B}, \nu)$ *be measure spaces, let* $\bar{p} \in (1, \infty]$*, let* $\mathbb{F} \in \{\mathbb{R}; \mathbb{C}\}$*, and let*

$$A\colon \mathsf{L}^1(X;\mathbb{F}) + \mathsf{L}^{\bar{p}}(X;\mathbb{F}) \to \mathsf{L}^0(Y;\mathbb{F})$$

be a mapping with the following properties:

(i) $|A(f+g)(y)| \le |A(f)(y)| + |A(g)(y)|$ *for all* $f, g \in \mathsf{L}^1(X;\mathbb{F}) + \mathsf{L}^{\bar{p}}(X;\mathbb{F})$ *and* $y \in Y$*;*

(ii) there exists $C_0 \in \mathbb{R}_{>0}$ *such that*

$$m_{A(f)}(\sigma) \le \frac{C_0}{\sigma}\|f\|_1, \qquad f \in \mathsf{L}^1(X;\mathbb{F}),\ \sigma \in \mathbb{R}_{>0};$$

(iii) we have the following cases:

---

[2] Józef Marcinkiewicz (1910-1940) was Polish mathematician who made contributions to analysis. He died as a prisoner of war during the Second World War.

(a) $p \in (1,\infty)$: *there exists* $C_1 \in \mathbb{R}_{>0}$ *such that*

$$m_{A(f)}(\sigma) \le \left(\frac{C_1}{\sigma}\|f\|_{\bar{p}}\right)^{\bar{p}}, \qquad f \in L^{\bar{p}}(X;\mathbb{F}),\ \sigma \in \mathbb{R}_{>0};$$

(b) $p = \infty$: *there exists* $C_1 \in \mathbb{R}_{>0}$ *such that*

$$\|A(f)\|_\infty \le C_1\|f\|_\infty, \qquad f \in L^\infty(X;\mathbb{F}).$$

Then, if $p \in (1,\bar{p})$, *there exists* $C_p \in \mathbb{R}_{>0}$ *such that*

$$\|A(f)\|_p \le C_p\|f\|_p, \qquad f \in L^p(X;\mathbb{F}).$$

Proof First let us show that, if $f \in L^p(X;\mathbb{F})$, $p \in (1,\bar{p})$, then $f \in L^1(X;\mathbb{F}) + L^{\bar{p}}(X;\mathbb{F})$. Indeed, let us denote

$$E_0 = \{x \in X \mid |f(x)| \le 1\}, \quad E_1 = \{x \in X \mid |f(x)| > 1\}$$

and note that $f = f\chi_{E_0} + f\chi_{E_1}$. We then have

$$|f\chi_{E_0}(x)| \le |f(x)|^p, \quad |f\chi_{E_1}(x)| \le |f(x)|^{p/\bar{p}}, \qquad x \in X.$$

Thus $f\chi_{E_0} \in L^1(X;\mathbb{F})$ and $f\chi_{E_1} \in L^{\bar{p}} \in L^{\bar{p}}(X;\mathbb{F})$.

Now let $p \in (1,\bar{p})$ and let $f \in L^p(X;\mathbb{F})$. For $\sigma \in \mathbb{R}_{>0}$, denote

$$F_\sigma = \{y \in Y \mid |A(f)(y)| > \sigma\}.$$

Also denote

$$E_{\sigma,0} = \{x \in X \mid |f(x)| \le \tfrac{\sigma}{2C_1}\}, \quad E_{\sigma,1} = \{x \in X \mid |f(x)| > \tfrac{\sigma}{2C_1}\}$$

and define $f_{\sigma,0} = f\chi_{E_{\sigma,0}}$ and $f_{\sigma,1} = f\chi_{E_{\sigma,1}}$. Note that $f_{\sigma,1} \in L^1(X,\mathbb{F})$ and $f_{\sigma,0} \in L^{\bar{p}}(X;\mathbb{F})$. We also denote

$$G_{\sigma,0} = \{y \in Y \mid |A(f_{\sigma,0})(y)| > \tfrac{\sigma}{2}\}, \quad G_{\sigma,1} = \{y \in Y \mid |A(f_{\sigma,1})(y)| > \tfrac{\sigma}{2}\}$$

and note that, by hypothesis (i),

$$\begin{aligned} F_\sigma &= \{y \in Y \mid |A(f)(y)| > \sigma\} \\ &= \{y \in Y \mid |A(f_{sigma,0} + f_{\sigma,1})(y)| > \sigma\} \\ &\subseteq y \in Y | A(f_{\sigma,0})(y)| + |A(f_{\sigma,1})(y)| > \sigma \\ &\subseteq \{y \in Y \mid |A(f_{\sigma,0})(y)| > \tfrac{\sigma}{2},\ |A(f_{\sigma,0})(y)| > \tfrac{\sigma}{2}\} \\ &\subseteq G_{\sigma,0} \cup G_{\sigma,1}. \end{aligned}$$

According to hypothesis (ii), we have

$$\nu(G_{\sigma,1}) \le \frac{2C_0}{\sigma}\|f_{\sigma,1}\|_1 \le \frac{2C_0}{\sigma}\int_{E_{\sigma,1}} |f(x)|\,d\mu(x).$$

Thus we compute, using Fubini's Theorem,

$$\begin{aligned}\int_0^\infty p\sigma^{p-1}\nu(G_{\sigma,1})\,\mathrm{d}\sigma &\le \int_0^\infty p\sigma^{p-1}\left(\frac{2C_0}{\sigma}\int_{E_{\sigma,1}}|f(x)|\,\mathrm{d}\mu(x)\right)\mathrm{d}\sigma\\ &\le 2C_0p\int|f(x)|\int_0^{2C_1|f(x)|}\sigma^{p-2}\mathrm{d}\sigma\,\mathrm{d}\mu(x)\\ &= \frac{2^pC_0C_1^{p-1}p}{p-1}\|f\|_p^p,\end{aligned}$$

noting that $p-2>-1$.

To estimate $\nu(G_{\sigma,0})$, and so estimate $\nu(F_\sigma)$, and so estimate $\|A(f)\|_p$ using Lemma 2, we consider two cases.

1. $\bar{p}\in(1,\infty)$: By the hypothesis (a) we have

$$\nu(G_{\sigma,0})\le\left(\frac{2C_1}{\sigma}\|f_{\sigma,0}\|_{\bar{p}}\right)^{\bar{p}}\le\frac{(2C_1)^{\bar{p}}}{\sigma^{\bar{p}}}\int_{E_{\sigma,0}}|f(x)|^{\bar{p}}\,\mathrm{d}\mu(x).$$

   Therefore, using Fubini's Theorem,

$$\begin{aligned}\int_0^\infty p\sigma^{p-1}\nu(G_{\sigma,0})\,\mathrm{d}\sigma &\le \int_0^\infty p\sigma^{p-1}\left(\frac{(2C_1)^{\bar{p}}}{\sigma^{\bar{p}}}\int_{E_{\sigma,0}}|f(x)|^{\bar{p}}\,\mathrm{d}\mu(x)\right)\mathrm{d}\sigma\\ &\le (2C_1)^{\bar{p}}p\int|f(x)|^{\bar{p}}\int_{2C_1|f(x)|}^\infty\sigma^{p-\bar{p}-1}\,\mathrm{d}\sigma\,\mathrm{d}\mu(x)\\ &= \frac{(2C_1)^pp}{\bar{p}-p}\|f\|_p^p,\end{aligned}$$

   noting that $p-\bar{p}-1<-1$.
2. $\bar{p}=\infty$: Here, directly by hypothesis (b), we have $\|f_{\sigma,0}\|_\infty<\frac{\sigma}{2C_1}$ and so $G_{\sigma,0}=\varnothing$ and so $\nu(G_{\sigma,0})=0$. In this case,

$$\int_0^\infty p\sigma^{p-1}\nu(G_{\sigma,1})\,\mathrm{d}\sigma=0$$

   Now we use Lemma 2 and our estimates above to calculate

$$\begin{aligned}\|A(f)\|_p^p &= \int_0^\infty p\sigma^{p-1}\nu(F_\sigma)\,\mathrm{d}\sigma\\ &\le \int_0^\infty p\sigma^{p-1}\nu(G_{\sigma,0})\,\mathrm{d}\sigma+\int_0^\infty p\sigma^{p-1}\nu(G_{\sigma,1})\,\mathrm{d}\sigma\\ &\le \begin{cases}(\frac{(2C_1)^pp}{\bar{p}-p}+\frac{2^pC_0C_1^{p-1}p}{p-1})\|f\|_p^p, & \bar{p}\in(1,\infty),\\ \frac{2^pC_0C_1^{p-1}p}{p-1}\|f\|_p^p, & \bar{p}=\infty.\end{cases}\end{aligned}$$

Thus the result follows by taking

$$C_p^p=\begin{cases}(\frac{(2C_1)^pp}{\bar{p}-p}+\frac{2^pC_0C_1^{p-1}p}{p-1})\|f\|_p^p, & \bar{p}\in(1,\infty),\\ \frac{2^pC_0C_1^{p-1}p}{p-1}\|f\|_p^p, & \bar{p}=\infty.\end{cases}$$

■

Let us now characterise the Hardy–Littlewood maximal function in terms of weak-L^p functions. For a set X and for $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, we denote by $\mathscr{M}_{\mathrm{fin}}(X;\mathbb{F})$ the set of finite $\mathbb{F}$-valued (not necessarily positive) measures on X.

7.1.5 Theorem (Hardy–Littlewood[3] Maximal Theorem) *If* $p \in [1,\infty]$, *if* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, *and if* $f \in \mathsf{L}^p(\mathbb{R};\mathbb{F})$, *then* $\mathsf{M}f(t)$ *is finite for almost every* $t \in \mathbb{R}$ *and we have the following statements:*

(i) if $p = 1$, *then* $\mathsf{M}f \in \mathsf{L}^{w,1}(\mathbb{R};\mathbb{R})$ *and, moreover,* $m_{\mathsf{M}f}(\sigma) \le \frac{2}{\sigma}\|f\|_1$ *for every* $f \in \mathsf{L}^1(\mathbb{R};\mathbb{F})$*;*

(ii) if $p \in (1,\infty]$, *then* $\mathsf{M}f \in \mathsf{L}^p(\mathbb{R};\mathbb{R})$ *and, moreover, there exists* $C_p \in \mathbb{R}_{>0}$ *such that* $\|\mathsf{M}f\|_p \le C_p\|f\|_p$ *for every* $f \in \mathsf{L}^p(\mathbb{R};\mathbb{F})$.

Also, if $\mu \in \mathscr{M}_{\mathrm{fin}}(\mathbb{R};\mathbb{F})$, *then*

(iii) $\mathsf{M}\mu \in \mathsf{L}^{w,1}(\mathbb{R};\mathbb{R})$ *and, moreover,* $m_{\mathsf{M}\mu}(\sigma) \le \frac{2}{\sigma}\|\mu\|_{\mathrm{TV}}$.

Proof In order to prove the theorem we will first prove a few lemmata.

1 Lemma (Wiener Covering Lemma) *Let* μ *be a finite positive Borel measure on* $\mathbb{R}$ *and let* $(I_1,\dots,I_n)$ *be a family of open intervals. Then there exist* $j_1,\dots,j_s \in \{1,\dots,n\}$ *such that the intervals* $(I_{j_1},\dots,I_{j_s})$ *are pairwise disjoint and*

$$\sum_{r=1}^{s} \mu(I_{j_r}) \ge \frac{1}{2}\mu\left(\bigcup_{j=1}^{n} I_j\right).$$

Proof We first claim that there exist $k_1,\dots,k_m \in \{1,\dots,n\}$ such that, for each $l \in \{1,\dots,m\}$,

$$I_{k_l} \not\subset \bigcup_{\substack{l'\in\{1,\dots,m\}\\ l'\neq l}} I_{k_{l'}}$$

and such that

$$\bigcup_{l=1}^{m} I_{k_l} = \bigcup_{j=1}^{n} I_j.$$

This claim we prove by induction on n, it being clearly true for $n = 1$. So suppose it true for $n = N$ and let $(I_1,\dots,I_{N+1})$ be a family of open intervals. By the induction hypothesis, there exists $k_1,\dots,k_m \in \{1,\dots,N\}$ such that, for each $l \in \{1,\dots,m\}$,

$$I_{k_l} \not\subset \bigcup_{\substack{l'\in\{1,\dots,m\}\\ l'\neq l}} I_{k_{l'}}$$

and such that

$$\bigcup_{l=1}^{m} I_{k_l} = \bigcup_{j=1}^{N} I_j.$$

[3] John Edensor Littlewood (1885–1977) was an English mathematician who made contributions to analysis and number theory. He was a frequent collaborator of Hardy.

We have two possibilities. The first possibility is that

$$I_{N+1} \subseteq \bigcup_{l=1}^{m} I_{k_l},$$

in which case the indices $k_1, \dots, k_m \in \{1, \dots, N+1\}$ are as claimed. The second possibility is that

$$I_{N+1} \not\subset \bigcup_{l=1}^{m} I_{k_l},$$

in which case the indices $k_1, \dots, k_m, k_{m+1} \in \{1, \dots, N+1\}$, with $k_{m+1} = N+1$, are as claimed.

Now, given the preceding claim, we can without loss of generality assume that the intervals $(I_1, \dots, I_n)$ have the property that, for any $j \in \{1, \dots, n\}$,

$$I_j \not\subset \bigcup_{\substack{j' \in \{1,\dots,n\} \\ j' \neq j}} I_{j'}.$$

Now write $I_j = (a_j, b_j)$ and assume the ordering of the intervals is such that $a_j \leq a_{j+1}$ for $j \in \{1, \dots, n-1\}$. We claim that $b_{j+1} > b_j$. Indeed, if this were not the case, then $I_{j+1} \subseteq I_j$. We also claim that $a_{j+1} > b_{j-1}$. Indeed, if this were not the case, then $I_j \subseteq I_{j-1} \cup I_{j+1}$. From this we can conclude that, if $j_1, j_2 \in \{1, \dots, n\}$ are distinct and either both even or both odd, then $I_{j_1} \cap I_{j_2} = \varnothing$. Moreover,

$$\sum_{j \text{ odd}} \mu(I_j) + \sum_{j \text{ even}} \mu(I_j) \geq \mu\left(\bigcup_{j=1}^{n} I_j\right).$$

We can conclude that at least one of the families of intervals

$$(I_j \mid j \text{ odd}), \quad (I_j \mid j \text{ even})$$

will give the desired collection of intervals. ▼

2 Lemma *For a measure space* $(\mathsf{X}, \mathscr{A}, \mu)$*, for* $\mathrm{p} \in [1, \infty]$*, for* $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$*, and for* $\mathrm{f} \in \mathsf{L}^0(\mathsf{X}; \mathbb{F})$

$$\int |\mathrm{f}(\mathrm{x})|^{\mathrm{p}} \, \mathrm{d}\mu(\mathrm{x}) = \int_0^\infty \mathrm{p}\sigma^{\mathrm{p}-1} \mathrm{m}_{\mathrm{f}}(\sigma) \, \mathrm{d}\sigma.$$

Proof If $\operatorname{supp}(f)$ does not have σ-finite measure, then both sides of the equality to be proved are ∞, and so the lemma holds in this case. This we suppose that $\operatorname{supp}(f)$ has σ-finite measure, i.e., $\operatorname{supp}(f)$ is a countable union of sets of finite measure. (We have now finished the abuse of notation of using "σ" in two different ways.) For $\sigma \in \mathbb{R}_{>0}$, denote

$$E_{f,\sigma} = \{x \in X \mid |f(x)| > \sigma\}.$$

Also note the obvious equality

$$|\alpha|^p = \int_0^\alpha p\sigma^{p-1} \, \mathrm{d}\sigma, \qquad \alpha \in \mathbb{R}_{\geq 0}.$$

We then calculate

$$\begin{aligned}\int |f(x)|^p \, d\mu(x) &= \int \int_0^{|f(x)|} p\sigma^{p-1} d\sigma \, d\mu(x) \\ &= \int_0^{\infty} \int p\sigma^{p-1} \chi_{E_{f,\sigma}}(x) \, d\mu(x) \, d\sigma \\ &= \int_0^{\infty} p\sigma^{p-1} m_f(\sigma) \, d\sigma,\end{aligned}$$

using Fubini's Theorem. ▼

(i) For $f \in \mathsf{L}^1(X; \mathbb{F})$ and $\sigma \in \mathbb{R}_{>0}$, denote

$$E_\sigma = \{t \in \mathbb{R} \mid \mathsf{M}f(t) > \lambda\}.$$

We claim that E_σ is open. Indeed, if $t \in E_\sigma$, then there exists an open interval $I \ni t$ and such that

$$\frac{1}{\lambda(I)} \int_O |f(\tau)| \, d\tau > \sigma.$$

Thus $I \subseteq E_\sigma$, giving openness of E_σ, as claimed. Now let $K \subseteq E_\sigma$ be compact. For each $t \in K$ there exists an open interval I_t such that

$$\lambda(I_t) < \frac{1}{\sigma} \int_{I_t} |f(\tau)| \, d\tau.$$

Choose a finite subcover of the open cover $(I_t)_{t \in K}$ of K so we have a finite collection $I_1, \dots, I_n$ covering K. By Lemma 1, let $j_1, \dots, j_s \in \{1, \dots, n\}$ be such that $(I_{j_1}, \dots, I_{j_s})$ are pairwise disjoint and satisfy

$$\lambda\left(\bigcup_{j=1}^{n} I_j\right) \le 2 \sum_{r=1}^{s} \lambda(I_{j_s}).$$

Thus we have

$$\lambda(K) \le \lambda\left(\bigcup_{j=1}^{n} I_j\right) \le 2 \sum_{r=1}^{s} \frac{1}{\sigma} \int_{I_{j_r}} |f(t)| \, dt \le \frac{2}{\lambda} \|f\|_1.$$

Now take a sequence of compact subsets $(K_j)_{j \in \mathbb{Z}_{>0}}$ of E_σ such that $\lim_{j \to \infty} \lambda(K_j) = \lambda(E_\sigma)$, cf. Theorem 2.4.19. Then

$$m_f(\sigma) = \lambda(E_\sigma) = \lim_{j \to \infty} \lambda(K_j) \le \frac{2}{\sigma} \|f\|_1,$$

as desired.

(ii) We shall employ Theorem 7.1.4 with $X = Y = \mathbb{R}$ (with Lebesgue measure), with A being the operator $f \mapsto \mathsf{M}f$, and with $\bar{p} = \infty$. Let us verify the hypotheses of Theorem 7.1.4. By definition of the maximal function, condition (i) holds. That (ii) holds was proved as part (i) of the theorem. For part (b), since

$$\frac{1}{\lambda(I)} \int_I |f(t)| \, dt \le \operatorname{ess\,sup}\{|f(t)| \mid t \in I\},$$

we conclude that $\|\mathsf{M}f\|_\infty < \|f\|_\infty$. Now an application of Theorem 7.1.4 gives the existence of $C_p \in \mathbb{R}_{>0}$ such that $\|\mathsf{M}f\|_p \le C_p\|f\|_p$, as desired.

(iii) The proof is rather like that for part (i), but we give it for completeness. For $\mu \in \mathscr{M}_{\mathrm{fin}}(\mathbb{R};\mathbb{F})$ and $\sigma \in \mathbb{R}_{>0}$, denote

$$E_\sigma = \{t \in \mathbb{R} \mid \mathsf{M}\mu(t) > \lambda\}.$$

We claim that E_σ is open. Indeed, if $t \in E_\sigma$, then there exists an open interval $I \ni t$ and such that

$$\frac{|\mu|(I)}{\lambda(I)} > \sigma.$$

Thus $I \subseteq E_\sigma$, giving openness of E_σ, as claimed. Now let $K \subseteq E_\sigma$ be compact. For each $t \in K$ there exists an open interval I_t such that

$$\lambda(I_t) < \frac{1}{\sigma}|\mu|(I_t).$$

Choose a finite subcover of the open cover $(I_t)_{t\in K}$ of K so we have a finite collection $I_1, \dots, I_n$ covering K. By Lemma 1, let $j_1, \dots, j_s \in \{1, \dots, n\}$ be such that $(I_{j_1}, \dots, I_{j_s})$ are pairwise disjoint and satisfy

$$\lambda\left(\bigcup_{j=1}^n I_j\right) \le 2\sum_{r=1}^s \lambda(I_{j_s}).$$

Thus we have

$$\lambda(K) \le \lambda\left(\bigcup_{j=1}^n I_j\right) \le 2\sum_{r=1}^s \frac{1}{\sigma}|\mu|(I_{j_r}) \le \frac{2}{\lambda}\|\mu\|_{\mathrm{TV}}.$$

Now take a sequence of compact subsets $(K_j)_{j\in\mathbb{Z}_{>0}}$ of E_σ such that $\lim_{j\to\infty}\lambda(K_j) = \lambda(E_\sigma)$, cf. Theorem 2.4.19. Then

$$m_f(\sigma) = \lambda(E_\sigma) = \lim_{j\to\infty}\lambda(K_j) \le \frac{2}{\sigma}\|\mu\|_{\mathrm{TV}},$$

as desired. ■

7.1.2 Nontangential limits

[Axler, Bourdon, and Ramey 2001].

7.1.6 Definition (Nontangential limits) Let $I \in \mathscr{I}$ and let $z_0 \in \mathbb{C}_I$. A ***cone tip*** in $\mathbb{C}_I$ at z_0 is a subset $K \subseteq \mathbb{C}$ such that there exist a convex cone $\overline{K}$ and a convex set C with the following properties:

(i) $\overline{K}$ is a closed proper convex cone (i.e., $\overline{K} \neq \mathbb{C}$) with nonempty interior;

(ii) $\overline{K}$ has its vertex at z_0;

(iii) C is a closed convex set with $z_0 \in \mathrm{int}(C)$;

(iv) $K = \overline{K} \cap C$;

(v) $K \setminus \{z_0\} \subseteq \mathbb{C}_{\mathrm{int}(I)}$.

A ***nontangential sequence*** in $\mathbb{C}_I$ at z_0 is a sequence $(z_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{C}_I$ for which there exists a cone tip K at z_0 with the following properties:

(vi) there exists $N \in \mathbb{Z}_{>0}$ such that $z_j \in K$ for $j \geq N$;

(vii) $(z_j)_{j\in\mathbb{Z}_{>0}}$ converges to z_0.

A function $F\colon \mathbb{C}_I \to \mathbb{C}$ has a ***nontangential limit*** at $z_0 \in \mathbb{C}_I$ if, for every nontangential sequence $(z_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{C}_I$ at z_0, $F(z_0) = \lim_{j\to\infty} F(z_j)$. A function $F\colon \mathbb{C}_I \to \mathbb{C}$ has ***nontangential limits*** if it has a nontangential limit at z_0 for almost every $z_0 \in \mathbb{C}_I$. •

These definitions differ enough from the usual definitions to merit a little explanation.

7.1.7 Remarks (Nontangential limits)

1. Our notion of cone tip in $\mathbb{C}_I$ at z_0 places very little in the way of restrictions on the cone if $z_0 \in \mathbb{C}_{\mathrm{int}(I)}$, as depicted in Figure 7.1. However, at boundary points

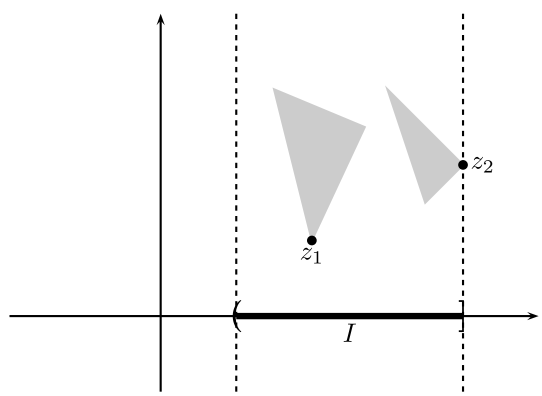


Figure 7.1 Cone tips at an interior point z_1 (left) and a boundary point z_2 (right)

the cones cannot contain vertical line segments, and so we see that cone tips as we define them capture the desired behaviour at boundary points.

2. If $F\colon \mathbb{C}_I \to \mathbb{C}$ is continuous on $\mathbb{C}_{\mathrm{int}(I)}$, then F clearly possesses a nontangential limit at every point in $\mathbb{C}_{\mathrm{int}(I)}$. We shall only be working with functions continuous on the interior of vertical strips, so the notion of nontangential limits is nonvoid only at boundary points.

3. If I has an *open* endpoint x_0, our definition does not include the possibility of having nontangential limits at points of the form $x_0 + \mathrm{i}y$ since these points are not in $\mathbb{C}_I$, and our definition explicitly allows nontangential limits to exist only at points in $\mathbb{C}_I$. The matter of extending to the boundary is one we shall consider in Theorem 7.4.4. •

In this section we introduce another sort of maximal function. Like the Hardy–Littlewood maximal function, this measures the variation of a function in a certain sense.

7.1.3 Constructions with subharmonic functions

Section 7.2

Hardy spaces of harmonic functions defined on vertical strips

In this section we consider spaces of holomorphic functions defined on vertical strips in the complex plane.

7.2.1 Definitions

We recall from Section II-3.8.1 the definition of an harmonic function of a complex variable, and some of the properties of such functions. In this section we shall be interested in harmonic functions whose domain is a vertical strip. Let us establish the notation for such domains. If $I \subseteq \mathbb{R}$, we denote by

$$\mathbb{C}_I = \{z \in \mathbb{C} \mid \operatorname{Re}(z) \in I\}$$

the vertical strip whose real part is in I; see Figure 7.2.

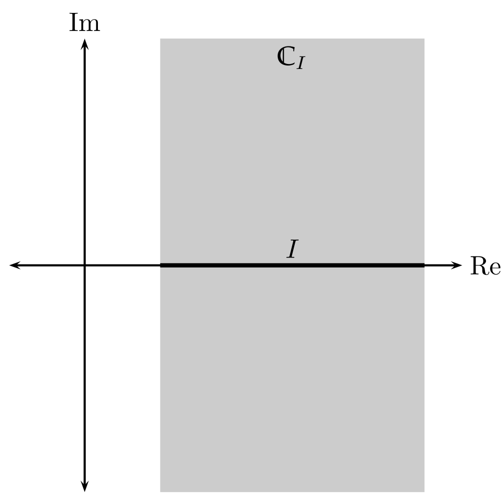


Figure 7.2 A vertical strip in the complex plane

For a function $F\colon \mathbb{C}_I \to \mathbb{F}$, $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and for $x \in I$, we denote $F_x\colon \mathbb{R} \to \mathbb{F}$ the function defined by $F_x(y) = F(x + \mathrm{i}y)$.

We now give a definition of a general class of harmonic functions.

7.2.1 Definition (Functions harmonic on a vertical strip) For an interval $I \subseteq \mathbb{R}$ and for $\mathbb{F} \in \{\mathbb{R}; \mathbb{C}\}$, we denote by $\mathsf{h}(\mathbb{C}_I; \mathbb{F})$ the mappings $F\colon \mathbb{C}_I \to \mathbb{F}$ that are harmonic on $\mathbb{C}_{\operatorname{int}(I)}$. If $\operatorname{int}(I) = \emptyset$, we take the convention that $\mathsf{h}(\mathbb{C}_I; \mathbb{F}) = \mathbb{F}^{\mathbb{C}_I}$. •

Note that, for each interval $I \subseteq \mathbb{R}$, $\mathsf{h}(\mathbb{C}_I; \mathbb{F})$ is a $\mathbb{F}$-vector space with respect to the operations of pointwise addition and scalar multiplication. Note that if $\operatorname{int}(I) \neq I$, then we place no restrictions on the values of functions in $\mathsf{h}(\mathbb{C}_I;\mathbb{C})$ when restricted to $\mathrm{bd}(\mathbb{C}_I)$. For more useful classes of functions, we will find it beneficial to assign these boundary values in a meaningful way, as in Section 7.1.2.

Now let us define the particular classes of harmonic functions in which we shall be interested. We begin by defining a large class of these functions, and then we consider a subset of these prescribed by some norm being finite.

7.2.2 Definition (Big harmonic Hardy space on a vertical strip) Let $I \subseteq \mathbb{R}$ be an interval, $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and $p \in [1,\infty]$. Denote by $\mathsf{h}^p(\mathbb{C}_I;\mathbb{F})$ the mappings $F \in \mathsf{h}(\mathbb{C}_I;\mathbb{F})$ such that

(i) $F \in \mathsf{h}(\mathbb{C}_I;\mathbb{F})$,

(ii) $F_x \in \mathsf{L}^p(\mathbb{R};\mathbb{F})$ for each $x \in I$, and

(iii) F has nontangential limits.

The space $\mathsf{h}^p(\mathbb{C}_I;\mathbb{F})$ is the ***big harmonic Hardy space***. •

Note that the condition for having nontangential limits is vacuous if I is open. However, if I does contain an endpoint or two, then the requirement of having nontangential limits at these endpoints becomes nonvoid. Indeed, the study of the existence of such nontangential limits will be something that will be of interest to us in Section 7.2.3 below.

Of particular interest is the following subset of the big harmonic Hardy space, where we ask that the L^p-norms on vertical lines be uniformly bounded.

7.2.3 Definition (Harmonic Hardy space on a vertical strip) Let $I \subseteq \mathbb{R}$ be an interval and let $p \in [1,\infty]$. For $F \in \mathsf{h}^p(\mathbb{C}_I;\mathbb{C})$ we denote

$$\|F\|_{\mathsf{h}^p,I} = \sup\{\|F_x\|_p \mid x \in I\}.$$

The subset $\{F \in \mathsf{h}^p(\mathbb{C}_I;\mathbb{C}) \mid \|F\|_{\mathsf{h}^p,I} < \infty\}$ is denoted by $\overline{\mathsf{h}}^p(\mathbb{C}_I;\mathbb{C})$ and we call these spaces the ***harmonic Hardy spaces***. •

We shall be primarily interested in the harmonic hardy spaces in this chapter. Indeed, we shall not have use in these volumes for the big harmonic Hardy spaces; we will, however, have use for their holomorphic analogues introduced in Section 7.4.1. These we have introduced here the big harmonic Hardy spaces in the interest of symmetry and aesthetics.

7.2.2 Poisson integral representations of harmonic functions on vertical strips

We shall characterise the harmonic Hardy spaces using Poisson integral representations. Let us define the appropriate Poisson kernels. We do this for three types of intervals, one for (a,∞), one for $(-\infty,b)$, and one for (a,b).

7.2.4 Definition (Poisson kernels for strips) Let $I \subset \mathbb{R}$ be an open interval. The ***Poisson kernel*** for I is the function $P_I \colon \mathbb{C}_I \to \mathbb{C}$ defined by:

(i) $I = (a, \infty)$: $P_I(x + \mathrm{i}y) = \dfrac{1}{\pi}\dfrac{x}{(x-a)^2 + y^2}$;

(ii) $I = (-\infty, b)$: $P_I(x + \mathrm{i}y) = \dfrac{1}{\pi}\dfrac{x}{(b-x)^2 + y^2}$;

(iii) $I = (a, b)$: $P_I(x + \mathrm{i}y) = \dfrac{1}{2\pi}\dfrac{\sin(\pi\frac{x-a}{b-a})}{\cosh(\pi\frac{y}{b-a}) - \cos(\pi\frac{x-a}{b-a})}$. •

Let us record some elementary properties of these Poisson kernels.

7.2.5 Lemma (Properties of Poisson kernels) *If* I ⊂ ℝ *be an open interval, the following statements hold:*

(i) $\mathrm{P_I}$ *is harmonic and so are its translates in* y*;*

(ii) $\mathrm{P_{I,x}} \in \mathsf{L}^{\mathrm{p}}$

(iii) integral of $\mathrm{P_I}$

(iv) tail estimate for $\mathrm{P_I^p}$

Proof ■

Now we use these Poisson kernels to construct functions from boundary data. We do this in the three cases. Let $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$.

1. $I = (a, \infty)$: Given $f \in \mathsf{L}^p(\mathbb{R}; \mathbb{F})$, we define

$$\begin{aligned} P_I f \colon \mathbb{C}_I &\to \mathbb{F} \\ x + \mathrm{i}y &\mapsto \frac{1}{\pi}\int_{\mathbb{R}} \frac{x}{(x-a)^2 + (y-\eta)^2} f(\eta)\,\mathrm{d}\eta. \end{aligned}$$

Given $\mu \in \mathscr{M}_{\mathrm{fin}}(\mathbb{R}; \mathbb{F})$, we define

$$\begin{aligned} P_I \mu \colon \mathbb{C}_I &\to \mathbb{F} \\ x + \mathrm{i}y &\mapsto \frac{1}{\pi}\int_{\mathbb{R}} \frac{x}{(x-a)^2 + (y-\eta)^2}\,\mathrm{d}\mu(\eta). \end{aligned}$$

2. $I = (-\infty, b)$: Given $g \in \mathsf{L}^p(\mathbb{R}; \mathbb{F})$, we define

$$\begin{aligned} P_I g \colon \mathbb{C}_I &\to \mathbb{F} \\ x + \mathrm{i}y &\mapsto \frac{1}{\pi}\int_{\mathbb{R}} \frac{x}{(b-x)^2 + (y-\eta)^2} g(\eta)\,\mathrm{d}\eta. \end{aligned}$$

Given $\nu \in \mathscr{M}_{\mathrm{fin}}(\mathbb{R}; \mathbb{F})$, we define

$$\begin{aligned} P_I \nu \colon \mathbb{C}_I &\to \mathbb{F} \\ x + \mathrm{i}y &\mapsto \frac{1}{\pi}\int_{\mathbb{R}} \frac{x}{(b-x)^2 + (y-\eta)^2}\,\mathrm{d}\nu(\eta). \end{aligned}$$

3. $I = (a, b)$: Given $f, g \in \mathsf{L}^p(\mathbb{R}; \mathbb{F})$, we define

$$\begin{aligned} P_I(f,g)\colon \mathbb{C}_I &\to \mathbb{F} \\ x + \mathrm{i}y &\mapsto \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\sin(\pi\frac{x-a}{b-a})}{\cosh(\pi\frac{y-\eta}{b-a}) - \cos(\pi\frac{x-a}{b-a})} f(\eta)\,\mathrm{d}\eta \\ &\quad + \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\sin(\pi\frac{b-x}{b-a})}{\cosh(\pi\frac{y-\eta}{b-a}) - \cos(\pi\frac{b-x}{b-a})} g(\eta)\,\mathrm{d}\eta. \end{aligned}$$

Given $\mu, \nu \in \mathscr{M}_{\mathrm{fin}}(\mathbb{R}; \mathbb{F})$, we define

$$\begin{aligned} P_I(\mu,\nu)\colon \mathbb{C}_I &\to \mathbb{F} \\ x + \mathrm{i}y &\mapsto \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\sin(\pi\frac{x-a}{b-a})}{\cosh(\pi\frac{y-\eta}{b-a}) - \cos(\pi\frac{x-a}{b-a})}\,\mathrm{d}\mu(\eta) \\ &\quad + \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\sin(\pi\frac{b-x}{b-a})}{\cosh(\pi\frac{y-\eta}{b-a}) - \cos(\pi\frac{b-x}{b-a})}\,\mathrm{d}\nu(\eta). \end{aligned}$$

At this point, we do not fuss about whether the integrals exist. We will take this into consideration as a matter of course when discussing the deeper properties of the above constructions.

The following properties of Poisson kernels explain our interest in them.

7.2.6 Theorem (Poisson integral representations of harmonic functions) *Let* $I \subset \mathbb{R}$ *be an open interval, let* $\mathbb{F} \in \{\mathbb{R}; \mathbb{C}\}$*, and let* $F\colon \mathbb{C}_I \to \mathbb{F}$*. For* $x \in I$ *define* $F_x(y) = F(x + \mathrm{i}y)$*. For* $p \in (1, \infty]$*, the following statements hold:*

(i) $I = (a, \infty)$ *or* $I = (-\infty, b)$*: the following statements are equivalent:*

  (a) there exists $f \in \mathsf{L}^p(\mathbb{R}; \mathbb{F})$ *such that* $F = P_I f$*;*

  (b) $F \in \overline{\mathsf{h}}^p(\mathbb{C}_I; \mathbb{F})$*;*

(ii) $I = (a, b)$*: the following statements are equivalent:*

  (a) there exists $f, g \in \mathsf{L}^p(\mathbb{R}; \mathbb{F})$ *such that* $F = P_I(f, g)$*;*

  (b) $F \in \overline{\mathsf{h}}^p(\mathbb{C}_I; \mathbb{F})$*;*

For $p = 1$*, the following statements hold:*

(iii) $I = (a, \infty)$ *or* $I = (-\infty, b)$*: the following statements are equivalent:*

  (a) there exists $\mu \in \mathscr{M}_{\mathrm{fin}}(\mathbb{R}; \mathbb{F})$ *such that* $F = P_I\mu$*;*

  (b) $F \in \overline{\mathsf{h}}^1(\mathbb{C}_I; \mathbb{F})$*;*

(iv) $I = (a, b)$*: the following statements are equivalent:*

  (a) there exists $\mu, \nu \in \mathscr{M}_{\mathrm{fin}}(\mathbb{R}; \mathbb{F})$ *such that* $F = P_I(\mu, \nu)$*;*

  (b) $F \in \overline{\mathsf{h}}^1(\mathbb{C}_I; \mathbb{F})$*;*

Proof We shall denote $P_{I,x}(y) = P_I(x+\mathrm{i}y)$ for $x \in I$. We also assume, without loss of generality, that $\mathbb{F} = \mathbb{R}$.

(i) We take $I = (0,\infty)$, the cases of a general unbounded interval following by translation and/or reflection.

(a) $\Longrightarrow$ (b) Let us first suppose that $f \in \mathsf{L}^p(\mathbb{R};\mathbb{R})$ and that

$$F(x+\mathrm{i}y) = \int_{\mathbb{R}} P_{I,x}(y-\eta)f(\eta)\,\mathrm{d}\eta.$$

We first show that F is harmonic. We will do this by differentiating $P_I f$ under the integral sign with respect to x and y. This requires that we verify that, for $x_0+\mathrm{i}y_0 \in \mathbb{C}_I$, the function

$$\alpha_{(x,y)}\colon \eta \mapsto \frac{x}{x^2+(y-\eta)^2}f(\eta),$$

and its first and second derivatives with respect to x and y, are bounded by an integrable function of η, uniformly for $x+\mathrm{i}y$ in some neighbourhood of $x_0+\mathrm{i}y_0$. This we argue as follows. Write $\alpha_{(x,y)}(\eta) = \beta_{(x,y)}(\eta)\frac{1}{1+\eta^2}$ for

$$\beta_{(x,y)}(\eta) = \frac{x(1+\eta^2)}{x^2+(y-\eta)^2}$$

and note that $\beta_{(x,y)}$, and all of its derivatives with respect to x and y, are bounded as functions of η. Thus, by continuity, given $x_0+\mathrm{i}y_0 \in \mathbb{C}_I$, there exists a neighbourhood $\mathcal{U}$ of $x_0+\mathrm{i}y_0$ such that $\beta_{(x,y)}$, and any finite number of its derivatives with respect to x and y, are bounded, as functions of η, uniformly in $\mathcal{U}$. Since $\eta \mapsto \frac{1}{1+\eta^2}$ is in $\mathsf{L}^{p'}(\mathbb{R};\mathbb{R})$ for every $p \in [1,\infty)$, we conclude using Hölder's inequality that, if $f \in \mathsf{L}^p(\mathbb{R};\mathbb{R})$, then $\alpha_{(x,y)}$, and any finite number of its derivatives with respect to x and y, are in $\mathsf{L}^1(\mathbb{R};\mathbb{R})$, uniformly in $\mathcal{U}$. Thus, by Theorem 2.9.16(ii), we can differentiate the expression

$$x+\mathrm{i}y \mapsto \frac{1}{\pi}\int_{\mathbb{R}} \frac{x}{x^2+(y-\eta)^2}f(\eta)\,\mathrm{d}\eta$$

under the integral sign, and verify that the expression defines an harmonic function of $x+\mathrm{i}y$ on $\mathbb{C}_I$ if $f \in \mathsf{L}^p(\mathbb{R};\mathbb{R})$, this by Lemma 7.2.5(i).

Since F_x is the convolution of $P_{I,x}$ with f for each $x \in I$, by Young's inequality (Theorem IV-4.2.8) and Lemma 7.2.5(iii), we have

$$\|F_x\|_p \le \|P_{I,x}\|_1\|f\|_p = \|f\|_p,$$

Thus $F \in \overline{\mathsf{h}}^p(\mathbb{C}_I;\mathbb{R})$.

(b) $\Longrightarrow$ (a) Let us first prove a couple of technical lemmata.

1 **Lemma** *Let* $\mathrm{I} = (0,\infty)$*, let* $\mathrm{p} \in [1,\infty]$*, let* $\mathrm{f} \in \mathsf{L}^{\mathrm{p}}(\mathbb{R};\mathbb{R})$*, let* $\mathrm{y}_0 \in \mathbb{R}$*, and suppose that* f *is continuous at* y_0*. Then*

$$\lim_{\mathrm{z}\to \mathrm{i}\mathrm{y}_0} \mathrm{P}_\mathrm{I}\mathrm{f}(\mathrm{z}) = \mathrm{f}(\mathrm{y}_0).$$

Proof Let $\epsilon \in \mathbb{R}_{>0}$. Let $\delta_2 \in \mathbb{R}_{>0}$ be sufficiently small that, if $|y - y_0| < \delta_2$ and $|\eta| < \delta_2$,

$$|f(y-\eta) - f(y_0)| < \frac{\epsilon}{3}.$$

With δ_2 so chosen, let $\delta_1 \in \mathbb{R}_{>0}$ be sufficiently small that, if $x < \delta_1$,

$$\int_{|\eta|>\delta_2} P_{I,x}(\eta)|f(y-\eta)|\,d\eta < \frac{\epsilon}{3}$$

(using Hölder's inequality and Lemma 7.2.5(iv)) and

$$|f(y_0)| \int_{|\eta|>\delta_2} P_{I,x}(\eta)\,d\eta < \frac{\epsilon}{3}$$

(using Lemma 7.2.5(iv)). By Lemma 7.2.5(iii) we have

$$P_I f(x+iy) - f(y_0) = \int_{\mathbb{R}} P_{I,x}(\eta)(f(y-\eta) - f(y_0))\,d\eta.$$

Then, for $|x| < \delta_1$ and $|y - y_0| < \delta_2$, we have

$$\begin{aligned}
|P_I f(x+iy) - f(y_0)| &\le \int_{\mathbb{R}} P_{I,x}(\eta)|f(y-\eta) - f(y_0)|\,d\eta \\
&= \int_{|\eta|\le\delta_2} P_{I,x}(\eta)|f(y-\eta) - f(y_0)|\,d\eta \\
&\quad + \int_{|\eta|>\delta_2} P_{I,x}(\eta)|f(y-\eta) - f(y_0)|\,d\eta \\
&\le \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon,
\end{aligned}$$

using Lemma 7.2.5(iii). ▼

2 Lemma *Let* $\mathrm{I} = (0, \infty)$. *If* $\mathrm{F}\colon \mathbb{C}_{\mathrm{cl(I)}} \to \mathbb{R}$ *is continuous and if* $\mathrm{F}|\mathbb{C}_{\mathrm{I}} \in \overline{\mathsf{h}}^{\infty}(\mathbb{C}_{\mathrm{I}}; \mathbb{R})$, *then*

$$\mathrm{F}(x+iy) = \int_{\mathbb{R}} \mathrm{P}_{\mathrm{I},x}(y-\eta)\mathrm{F}(i\eta)\,d\eta, \qquad x+iy \in \mathbb{C}_{\mathrm{I}}.$$

Proof Denote

$$G(x+iy) = F(x+iy) - \int_{\mathbb{R}} P_{I,x}(y-\eta)F(i\eta)\,d\eta.$$

From the proof of the implication (a) $\Longrightarrow$ (b) above in the case of $p = \infty$, the second term in the definition of G, and thus G itself, is harmonic and bounded on $\mathbb{C}_I$. By Lemma 1, G can be extended continuously to $\mathbb{C}_{\mathrm{cl}(I)}$ by taking the value 0 on $\mathrm{bd}(\mathbb{C}_I)$. Now define $H\colon \mathbb{C} \to \mathbb{R}$ by

$$H(x+iy) = \begin{cases} G(x+iy), & x \ge 0, \\ -G(-x+iy), & x < 0. \end{cases}$$

Note that H is bounded and continuous. We claim that it is also harmonic. By Theorem II-3.8.2(iv), it suffices to show that H has the mean value property at any point in $\mathbb{C}$. For points z in $\mathbb{C}_I$, we can choose a disk $\overline{\mathsf{D}}(r,z)$ small enough that $\overline{\mathsf{D}}(r,z) \subseteq \mathbb{C}_I$. Because G is harmonic, the mean value property for H will be satisfied for this disk. Next let $z = x + \mathrm{i}y \in \mathbb{C}_{-I}$. We let $r \in \mathbb{R}_{>0}$ be sufficiently small that $\overline{\mathsf{D}}(r,z) \subseteq \mathbb{C}_{-I}$. Then we have

$$\int_0^{2\pi} H(z + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta = -\int_0^{2\pi} G(-x + \mathrm{i}y + r\mathrm{e}^{-\mathrm{i}\theta})\,\mathrm{d}\theta = -G(-x + \mathrm{i}y) = H(z),$$

verifying the mean value property in this case. Finally, if $z = \mathrm{i}y$ and if $r \in \mathbb{R}_{>0}$, then

$$\begin{aligned}\int_0^{2\pi} H(\mathrm{i}y + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} H(\mathrm{i}y + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta + \int_{\frac{\pi}{2}}^{-\frac{\pi}{2}} H(\mathrm{i}y + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} G(\mathrm{i}y + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta - \int_{\frac{\pi}{2}}^{-\frac{\pi}{2}} G(\mathrm{i}y + r\mathrm{e}^{-\mathrm{i}\theta})\,\mathrm{d}\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} G(\mathrm{i}y + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} G(\mathrm{i}y + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta = 0,\end{aligned}$$

showing that the mean value property is satisfied on the imaginary axis, and so everywhere in $\mathbb{C}$. Thus H is a bounded harmonic function on $\mathbb{C}$, and so is constant . ref
Since $H(0) = 0$, the lemma follows. ▼

3 Lemma *Let* $\mathrm{I} = (0, \infty)$, *let* $\mathrm{p} \in [1, \infty)$, *and let and let* $\mathrm{F} \in \overline{\mathsf{h}}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}}; \mathbb{R})$. *Then, for* $\mathrm{z} = \mathrm{x} + \mathrm{iy} \in \mathbb{C}_{\mathrm{I}}$, *we have*

$$|\mathrm{F}(\mathrm{z})| \le \left(\frac{1}{\pi \mathrm{x}^2}\right)^{1/\mathrm{p}} \|\mathrm{F}\|_{\mathsf{h}^{\mathrm{p}},\mathrm{I}}.$$

Proof By the mean value property of harmonic functions (Theorem II-3.8.2(iii)), we have

$$F(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(z + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta$$

for every $r \in (0, \rho]$ for which $\overline{\mathsf{D}}(r,z) \subseteq \mathbb{C}_I$. Thus

$$\begin{aligned}\frac{1}{\mathrm{vol}(\overline{\mathsf{D}}(\rho,z))} \int_{\overline{\mathsf{D}}(\rho,z)} F(\zeta)\,\mathrm{d}\zeta &= \frac{1}{\pi\rho^2} \int_0^{\rho} \int_{-\pi}^{\pi} F(z + r\mathrm{e}^{\mathrm{i}\theta}) r\,\mathrm{d}\theta\,\mathrm{d}r \\ &= \frac{2}{\rho^2} \int_0^{\rho} F(z) r\,\mathrm{d}r = F(z).\end{aligned}$$

Therefore, by Hölder's Inequality, if $z = x + \mathrm{i}y$, we have

$$\begin{aligned}
|F(z)| &\le \frac{1}{\operatorname{vol}(\overline{\mathsf{D}}(x,z))}\int_{\overline{\mathsf{D}}(x,z)} |F(\zeta)|\,\mathrm{d}\zeta \\
&\le \frac{1}{\pi x^2}\left(\int_{\overline{\mathsf{D}}(x,z)} |F(\zeta)|^p\,\mathrm{d}\zeta\right)^{1/p} \operatorname{vol}(\overline{\mathsf{D}}(x,z))^{1/p'} \\
&\le \left(\frac{1}{\pi x^2}\int_0^{2x}\int_{\mathbb{R}} |F(\xi+\mathrm{i}\eta)|^p\,\mathrm{d}\eta\,\mathrm{d}\xi\right)^{1/p} \\
&\le \left(\frac{2}{\pi x}\right)^{1/p} \|F\|_{\mathsf{h}^p,I},
\end{aligned}$$

as claimed. ▼

With these technicalities at hand, we can proceed with the proof. We let $F \in \overline{\mathsf{h}}^p(\mathbb{C}_I;\mathbb{R})$. Let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence converging to 0 from above and define $f_j \in \mathsf{L}^p(\mathbb{R};\mathbb{R})$ by $f_j(\eta) = F(x_j + \mathrm{i}\eta)$. Note that, since $F \in \overline{\mathsf{h}}^p(\mathbb{C}_I;\mathbb{R})$, $(f_j)_{j\in\mathbb{Z}_{>0}}$ is a bounded sequence in $\mathsf{L}^p(\mathbb{R};\mathbb{R}) \simeq (\mathsf{L}^{p'}(\mathbb{R};\mathbb{R}))^*$ (this characterisation of the dual space is Theorem 3.10.1). It follows by the Banach–Alaoglu Theorem that there is a subsequence $(f_{j_k})_{k\in\mathbb{Z}_{>0}}$ and $f \in \mathsf{L}^p(\mathbb{R};\mathbb{R})$ such that $(f_{j_k})_{k\in\mathbb{Z}_{>0}}$ converges weak-* to f. Now define

ref

$$\begin{aligned}
F_j\colon \mathbb{C}_I &\to \mathbb{R} \\
z &\mapsto F(x_j + z).
\end{aligned}$$

By Lemma 3, for $p \in (1,\infty)$, F_j is bounded and continuous $\mathbb{C}_{\mathrm{cl}(I)}$ and harmonic in $\mathbb{C}_I$. In case $p = \infty$, this holds by hypothesis. Therefore, by Lemma 2 we have, for $z = x + \mathrm{i}y \in \mathbb{C}_I$,

$$F_j(z) = \int_{\mathbb{R}} P_{I,x}(y-\eta)F_j(\mathrm{i}\eta)\,\mathrm{d}\eta = \int_{\mathbb{R}} P_{I,x}(y-\eta)f_j(\eta)\,\mathrm{d}\eta, \qquad j \in \mathbb{Z}_{>0}.$$

Since $P_{I,x} \in \mathsf{L}^{p'}(\mathbb{R};\mathbb{R})$ by Lemma 7.2.5(ii), we have

$$\begin{aligned}
F(z) &= \lim_{k\to\infty} F(x_{j_k} + z) = \lim_{k\to\infty} F_{j_k}(z) \\
&= \lim_{k\to\infty}\int_{\mathbb{R}} P_{I,x}(y-\eta)f_{j_k}(\eta)\,\mathrm{d}\eta \\
&= \int_{\mathbb{R}} P_{I,x}(y-\eta)f(\eta)\,\mathrm{d}\eta,
\end{aligned}$$

by definition of weak-* convergence and the pairing of $\mathsf{L}^{p'}(\mathbb{R};\mathbb{R})$ with its dual. This concludes this part of the proof.

(ii) We assume that $I = (0,\pi)$, the other cases following by change of variables.

(a) $\Longrightarrow$ (b) We suppose that

$$F(x+\mathrm{i}y) = \frac{1}{2\pi}\int_{\mathbb{R}} P_{I,x}(y-\eta)f(\eta)\,\mathrm{d}\eta + \frac{1}{2\pi}\int_{\mathbb{R}} P_{I,\pi-x}(y-\eta)g(\eta)\,\mathrm{d}\eta$$

for $f, g \in \mathsf{L}^p(\mathbb{R};\mathbb{R})$. Let us first show that F is harmonic. To do this, assume that $g = 0$. A similar argument can be used when $f = 0$, and the general case follows since the sum of harmonic functions is harmonic. We will do this, as we did in the previous part of the proof, by showing that we can differentiate under the integral sign. We define

$$\alpha_{(x,y)} \colon \eta \mapsto \frac{\sin(x)}{\cosh(y-\eta)-\cos(x)} f(\eta)$$

and show that it, and its first two derivatives with respect to x and y, are bounded in a neighbourhood of any point $x_0 + \mathrm{i}y_0 \in \mathbb{C}_I$. We do this by writing $\alpha_{(x,y)}(\eta) = \beta_{(x,y)} \frac{1}{\cosh(\eta)}$, where

$$\beta_{(x,y)}(\eta) = \frac{\sin(x)\cosh(\eta)}{\cosh(y-\eta)-\cos(x)}.$$

Then $\beta_{(e,y)}$, and any of its derivatives with respect to x and y, are bounded as functions of η. We can argue just as we did in the first part of the proof that $\alpha_{(x,y)}$, and any finite number of its derivatives with respect to x and y, are in $\mathsf{L}^1(\mathbb{R};\mathbb{R})$, uniformly in $\mathcal{U}$. Thus we can differentiate the expression

$$x + \mathrm{i}y \mapsto \int_{\mathbb{R}} P_{I,x}(y-\eta) f(\eta)\,\mathrm{d}\eta$$

with respect to x and y under the integral sign, and so conclude by Lemma 7.2.5(i) that the expression is harmonic in $\mathbb{C}_I$.

We still assume that $g = 0$. Since F_x is the convolution of $P_{I,x}$ with f for each $x \in I$, by Young's inequality (Theorem IV-4.2.8) and Lemma 7.2.5(iii), we have

$$\|F_x\|_p \le \|P_{I,x}\|_1 \|f\|_p \le \|f\|_p,$$

Thus $F \in \overline{\mathsf{h}}^p(\mathbb{C}_I;\mathbb{R})$.

(b) $\implies$ (a) We first prove a few technical lemmata which serve rather the same purpose as those above for $I = (0,\infty)$.

4 Lemma *Let* $\mathrm{I} = (0,\pi)$*, let* $\mathrm{p} \in [1,\infty]$*, and let* $\mathrm{f} \in \mathsf{L}^{\mathrm{p}}(\mathbb{R};\mathbb{R})$*. Then*

$$\lim_{\mathrm{z}\to\pi+\mathrm{iy}} \mathrm{P_I}(\mathrm{f},0)(\mathrm{z}) = 0, \qquad \mathrm{y} \in \mathbb{R}.$$

Proof Let $y_0 \in \mathbb{R}$. Choose x_0 sufficiently close to π that $\cos(x) \in (-1, -\frac{1}{2}]$ for $x \in [x_0,\pi)$. For $x \in [x_0,\pi)$ we thus have $-\cos(x) \ge \frac{1}{2}$ and so

$$P_{I,x}(y-\eta) \le \frac{\cosh(\eta)}{\cosh(y-\eta)+\frac{1}{2}} \frac{1}{\cosh(\eta)}.$$

Since

$$\frac{\cosh(\eta)}{\cosh(y-\eta)+\frac{1}{2}}$$

is bounded as a function of η for fixed y, for a given $y_0 \in \mathbb{R}$, there exists a neighbourhood $\mathcal{U} \subseteq \mathbb{R}$ of y_0 and $C \in \mathbb{R}_{>0}$ such that

$$\frac{\cosh(\eta)}{\cosh(y-\eta)+\frac{1}{2}} < C$$

for $y \in \mathcal{U}$. Since $\eta \mapsto \frac{1}{\cosh(\eta)}$ is in $\mathsf{L}^{p'}(\mathbb{R};\mathbb{R})$ for every $p' \in [1,\infty)$, we conclude by Hölder's inequality that

$$\eta \mapsto P_{I,x}(y-\eta)f(\eta)$$

is in $\mathsf{L}^1(\mathbb{R};\mathbb{R})$, uniformly in $[x_0,\pi) \times \mathcal{U} \subseteq \mathbb{C}_I$. Thus, by the Dominated Convergence Theorem,

$$\lim_{x+\mathrm{i}y\mapsto\pi+\mathrm{i}y_0} \int_{\mathbb{R}} P_{I,x}(y-\eta)f(\eta)\,\mathrm{d}\eta = \int_{\mathbb{R}} \lim_{x+\mathrm{i}y\to\pi+\mathrm{i}y_0} \frac{\sin(x)}{\cosh(y-\eta)-\cos(x)} f(\eta)\,\mathrm{d}\eta = 0,$$

as desired. ▼

In similar manner, of course, if $g \in \mathsf{L}^p(\mathbb{R};\mathbb{R})$, we have

$$\lim_{z\to\mathrm{i}y} P_I(0,g)(z) = 0, \qquad y \in \mathbb{R}.$$

We will use this fact subsequently.

5 Lemma *Let* $\mathrm{I} = (0,\pi)$*, let* $\mathrm{p} \in [1,\infty]$*, let* $\mathrm{f} \in \mathsf{L}^{\mathrm{p}}(\mathbb{R};\mathbb{R})$*, let* $\mathrm{y}_0 \in \mathbb{R}$*, and suppose that* f *is continuous at* y_0*. Then*

$$\lim_{\mathrm{z}\to\mathrm{i}\mathrm{y}_0} \mathrm{P}_\mathrm{I}(\mathrm{f},0)(\mathrm{z}) = \mathrm{f}(\mathrm{y}_0).$$

Proof Abbreviate $F = P_I(f,0)$. Let $\epsilon \in \mathbb{R}_{>0}$ and choose $\delta \in \mathbb{R}_{>0}$ sufficiently small that, if $|\eta| < \delta_2$ and if $|y-y_0| < \delta_2$, then

$$|f(y-\eta) - f(y_0)| < \frac{\epsilon}{4}.$$

For $x \in (0,\pi)$ we have

$$|F(x+\mathrm{i}y) - f(y_0)| \le \left|F(x+\mathrm{i}y) - \frac{\pi-x}{\pi}f(y_0)\right| + \left|f(y_0) - \frac{\pi-x}{\pi}f(y_0)\right|.$$

Choose $\delta_1 \in \mathbb{R}_{>0}$ sufficiently small that

$$\left|f(y_0) - \frac{\pi-x}{\pi}f(y_0)\right| < \frac{\epsilon}{4}$$

if $x < \delta_1$. With δ_2 as chosen above, we can suppose that $\delta_1 \in \mathbb{R}_{>0}$ is sufficiently small that, if $x < \delta_1$,

$$\int_{|\eta|>\delta_2} P_{I,x}(\eta)|f(y-\eta)|\,\mathrm{d}\eta < \frac{\epsilon}{4}$$

(using Hölder's inequality and Lemma 7.2.5(iv)) and

$$|f(y_0)| \int_{|\eta|>\delta_2} P_{I,x}(\eta)\,\mathrm{d}\eta < \frac{\epsilon}{4}$$

(using Lemma 7.2.5(iv)). Now we compute

$$\begin{aligned}\left|F(x+\mathrm{i}y)-\frac{\pi-x}{\pi}f(y_0)\right| &= \left|\int_{\mathbb{R}} P_{I,x}(\eta)f(y-\eta)\,\mathrm{d}\eta - f(y_0)\int_{\mathbb{R}} P_{I,x}(\eta)\,\mathrm{d}\eta\right| \\ &\le \int_{\mathbb{R}} P_{I,x}(\eta)|f(y-\eta)-f(y_0)|\,\mathrm{d}\eta \\ &\le \int_{|\eta|\le\delta_2} P_{I,x}(\eta)|f(y-\eta)-f(y_0)|\,\mathrm{d}\eta \\ &\quad + \int_{|\eta|>\delta_2} P_{I,x}(\eta)|f(y-\eta)-f(y_0)|\,\mathrm{d}\eta \\ &\le \frac{\epsilon}{4}+\frac{\epsilon}{4}+\frac{\epsilon}{4}=\frac{3\epsilon}{4},\end{aligned}$$

using Lemma 7.2.5(iii). This then gives

$$|F(x+\mathrm{i}y)-f(y_0)|<\epsilon,$$

as desired. ▼

In a similar manner, of course, if $g \in \mathsf{L}^p(\mathbb{R};\mathbb{R})$ is continuous at $y_0 \in \mathbb{R}$, we have

$$\lim_{z\to\pi+\mathrm{i}y_0} P_I(0,g)(z) = g(y_0),$$

a fact which we shall use subsequently.

6 Lemma *Let* $\mathrm{I}=(0,\pi)$*. If* $\mathrm{F}\colon \mathbb{C}_{\mathrm{cl(I)}} \to \mathbb{R}$ *is continuous and if* $\mathrm{F}|\mathbb{C}_{\mathrm{I}} \in \overline{\mathsf{h}}^\infty(\mathbb{C}_{\mathrm{I}};\mathbb{R})$*, then*

$$\mathrm{F}(\mathrm{x}+\mathrm{i}\mathrm{y}) = \int_{\mathbb{R}} \mathrm{P}_{\mathrm{I,x}}(\mathrm{y}-\eta)\mathrm{F}(\mathrm{i}\eta)\,\mathrm{d}\eta + \int_{\mathbb{R}} \mathrm{P}_{\mathrm{I},\pi-\mathrm{x}}(\mathrm{y}-\eta)\mathrm{F}(\pi+\mathrm{i}\eta)\,\mathrm{d}\eta, \qquad \mathrm{x}+\mathrm{i}\mathrm{y}\in\mathbb{C}_{\mathrm{I}}.$$

Proof Define $G\colon \mathbb{C}_I \to \mathbb{R}$ by

$$G(z) = F(z) - \int_{\mathbb{R}} P_{I,x}(y-\eta)F(\mathrm{i}\eta)\,\mathrm{d}\eta + \int_{\mathbb{R}} P_{I,\pi-x}(y-\eta)F(\pi+\mathrm{i}\eta)\,\mathrm{d}\eta.$$

Note that, for $y_0 \in \mathbb{R}$,

$$\lim_{z\to\mathrm{i}y} G(z) = F(\mathrm{i}y) - \lim_{z\to\mathrm{i}y}\int_{\mathbb{R}} P_{I,x}(y-\eta)F(\mathrm{i}\eta)\,\mathrm{d}\eta - \lim_{z\to\mathrm{i}y}\int_{\mathbb{R}} P_{I,\pi-x}(y-\eta)F(\pi+\mathrm{i}\eta)\,\mathrm{d}\eta = 0,$$

using Lemmata 4 and 5. In like manner,

$$\lim_{z\to\pi+\mathrm{i}y} G(z) = 0.$$

Thus G can be extended to a continuous function on $\mathbb{C}_{\mathrm{cl}(I)}$ by assigning to G the value 0 on $\mathrm{bd}(\mathbb{C}_I)$. Now we define $H\colon \mathbb{C}\to\mathbb{R}$ by "periodically extending" it in x in a particular manner. We first define $H|\mathbb{C}_{[0,2\pi]}$ by

$$H(z) = \begin{cases} G(z), & z \in \mathbb{C}_{[0,\pi]}, \\ -G(2\pi - z), & z\in\mathbb{C}_{[\pi,2\pi]}. \end{cases}$$

Then we require that, for $k \in \mathbb{Z}$ and for $z \in \mathbb{C}_{[2k\pi,2(k+1)\pi]}$, $H(z) = G(z-2k\pi)$. The resulting function will be harmonic on $\mathbb{C}$, which can be shown be demonstrating that it has the mean value property, following the proof of Lemma 2. And, just as in the proof of that lemma, we conclude that H is constant by Liouville's Theorem. This gives the lemma. ▼

7 Lemma *Let* $\mathrm{I} = (0, \pi)$, *let* $\mathrm{p} \in (1, \infty)$, *and let* $\mathrm{F} \in \overline{\mathsf{h}}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}}; \mathbb{R})$. *Then, for* $\mathrm{z} = \mathrm{x} + \mathrm{iy} \in \mathbb{C}_{\mathrm{I}}$,

$$|\mathrm{F}(\mathrm{z})| \leq \left(\frac{1}{\pi \min\{\mathrm{x}, \pi - \mathrm{x}\}^2}\right)^{1/\mathrm{p}} \|\mathrm{F}\|_{\mathsf{h}^{\mathrm{p}},\mathrm{I}}.$$

Proof The idea of the proof is like that we used for Lemma 3. In the present case, we let $z = x + \mathrm{i}y \in \mathbb{C}_I$, let $r = \min\{x, \pi - x\}$, and compute

$$\begin{aligned}
|F(z)| &\leq \frac{1}{\mathrm{vol}(\overline{\mathsf{D}}(r,z))} \int_{\overline{\mathsf{D}}(r,z)} |F(\zeta)| \,\mathrm{d}\zeta \\
&\leq \frac{1}{\pi r^2} \left(\int_{\overline{\mathsf{D}}(r,z)} |F(\zeta)|^p \,\mathrm{d}\zeta\right)^{1/p} \mathrm{vol}(\overline{\mathsf{D}}(r,z))^{1/p'} \\
&\leq \left(\frac{1}{\pi r^2} \int_0^{2r} \int_{\mathbb{R}} |F(x + \mathrm{i}y)|^p \,\mathrm{d}y \,\mathrm{d}x\right)^{1/p} \\
&\leq \left(\frac{2}{\pi r}\right)^{1/p} \|F\|_{\mathsf{h}^p,I},
\end{aligned}$$

as claimed. ▼

Now we proceed with the proof. Let $F \in \overline{\mathsf{h}}^p(\mathbb{C}_I; \mathbb{R})$. Let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $(0, \frac{\pi}{2})$ converging to 0 from above. Define $(f_j)_{j\in\mathbb{Z}_{>0}}$ and $(g_j)_{j\in\mathbb{Z}_{>0}}$ by

$$f_j(\eta) = F(x_j + \mathrm{i}\eta), \quad g_j(\eta) = F(\pi - x_j + \mathrm{i}\eta).$$

We think of these as together defining a sequence of functions that we denote by $((f_j, g_j))_{j\in\mathbb{Z}_{>0}}$ on $\mathrm{bd}(\mathbb{C}_I)$ in the obvious way:

$$(f_j, g_j)(z) = \begin{cases} f_j(\mathrm{Im}(z)), & \mathrm{Re}(z) = 0, \\ g_j(\mathrm{Im}(z)), & \mathrm{Re}(z) = \pi. \end{cases}$$

Since $F \in \overline{\mathsf{H}}^p(\mathbb{C}_I; \mathbb{R})$, these are bounded sequences in $\mathsf{L}^p(\mathrm{bd}(\mathbb{C}_I); \mathbb{R})$ and so, by the same arguments as above for $I = (0, \infty)$, there exists a subsequence $((f_{j_k}, g_{j_k}))_{k\in\mathbb{Z}_{>0}}$ that converges weak-$*$ to a limit $(f, g) \in \mathsf{L}^p(\mathrm{bd}(\mathbb{C}_I); \mathbb{R})$. Now define

$$\begin{aligned}
F_j \colon \mathbb{C}_I &\to \mathbb{R} \\
z &\mapsto F(x_j + \frac{\pi - 2x_j}{\pi} z).
\end{aligned}$$

One directly verifies that F_j is harmonic. By Lemma 3, for $p \in (1, \infty)$, F_j is bounded and continuous in $\mathbb{C}_{\mathrm{cl}(I)}$ and harmonic in $\mathbb{C}_I$. In case $p = \infty$, this holds by hypothesis.

Therefore, by Lemma 2, we have, for $z = x + \mathrm{i}y \in \mathbb{C}_I$,

$$\begin{aligned} F_j(z) &= \int_{\mathbb{R}} P_{I,x}(y-\eta)F_j(\mathrm{i}\eta)\,\mathrm{d}\eta + \int_{\mathbb{R}} P_{I,\pi-x}(y-\eta)F_j(\pi+\mathrm{i}\eta)\,\mathrm{d}\eta \\ &= \int_{\mathbb{R}} P_{I,x}(y-\eta)f_j(\eta)\,\mathrm{d}\eta + \int_{\mathbb{R}} P_{I,\pi-x}(y-\eta)g_j(\eta)\,\mathrm{d}\eta \end{aligned}$$

for $j \in \mathbb{Z}_{>0}$. Since $P_{I,x}, P_{I,\pi-x} \in \mathsf{L}^{p'}(\mathbb{R};\mathbb{R})$ by Lemma 7.2.5(ii), we can think of $(P_{I,x}, P_{I,\pi-x})$ as an element of $\mathsf{L}^{p'}(\mathrm{bd}(\mathbb{C}_I);\mathbb{R})$ in the obvious way:

$$(P_{I,x}, P_{I,\pi-x})(z) = \begin{cases} P_{I,x}(\mathrm{Im}(z)), & \mathrm{Re}(z) = 0, \\ P_{I,\pi-x}(\mathrm{Im}(z)), & \mathrm{Re}(z) = \pi. \end{cases}$$

We then have

$$\begin{aligned} F(z) &= \lim_{k\to\infty} F(x_{j_k} + \frac{\pi - 2x_{j_k}}{\pi} z) = \lim_{k\to\infty} F_{j_k}(z) \\ &= \lim_{k\to\infty} \left(\int_{\mathbb{R}} P_{I,x}(y-\eta) f_{j_k}(\eta)\,\mathrm{d}\eta + \int_{\mathbb{R}} P_{I,\pi-x}(y-\eta) g_{j_k}(\eta)\,\mathrm{d}\eta \right) \\ &= \int_{\mathbb{R}} P_{I,x}(y-\eta) f(\eta)\,\mathrm{d}\eta + \int_{\mathbb{R}} P_{I,\pi-x}(y-\eta) g(\eta)\,\mathrm{d}\eta, \end{aligned}$$

by definition of weak-* convergence and the pairing of $\mathsf{L}^{p'}(\mathrm{bd}(\mathbb{C}_I);\mathbb{R})$ with its dual, $\mathsf{L}^{p}(\mathrm{bd}(\mathbb{C}_I);\mathbb{R})$ Theorem 3.10.1.

(iii) We again take $I = (0,\infty)$.

(a) $\implies$ (b) We suppose that

$$F(x+\mathrm{i}y) = \frac{1}{\pi} \int_{\mathbb{R}} \frac{x}{x^2 + (y-\eta)^2}\,\mathrm{d}\mu(\eta)$$

for a finite measure μ. For $z = x + \mathrm{i}y \in \mathbb{C}_I$, let $r \in \mathbb{R}_{>0}$ be such that $\overline{\mathsf{D}}(r,z) \subseteq \mathbb{C}_I$. We then compute, by Fubini's Theorem,

$$\begin{aligned} \int_{\mathrm{bd}(\overline{\mathsf{D}}(r,z))} F(z + r\mathrm{e}^{\mathrm{i}\theta}) &= \int_0^{2\pi} \left(\frac{1}{\pi} \int_{\mathbb{R}} \frac{x + r\cos\theta}{(x + r\cos\theta)^2 + (y + r\sin\theta - \eta)^2}\,\mathrm{d}\mu(\eta) \right) \mathrm{d}\theta \\ &= \frac{1}{\pi} \int_{\mathbb{R}} \left(\int_0^{2\pi} \frac{x + r\cos\theta}{(x + r\cos\theta)^2 + (y + r\sin\theta - \eta)^2}\,\mathrm{d}\theta \right) \mathrm{d}\mu(\eta) \\ &= \frac{1}{\pi} \int_{\mathbb{R}} \frac{x}{x^2 + (y-\eta)^2}\,\mathrm{d}\mu(\eta) = F(z), \end{aligned}$$

using the mean value property of P_I, since P_I is harmonic (Theorem II-3.8.2(iii)). By Theorem II-3.8.2(iv), we conclude that F is harmonic on $\mathbb{C}_I$.

Now let us note that

$$F(x+\mathrm{i}y) = \int_{\mathbb{R}} P_{I,x}(y-\eta)\,\mathrm{d}\mu(\eta),$$

i.e., F_x is the convolution of $P_{I,x} \in \mathsf{L}^1(\mathbb{R};\mathbb{R})$ with μ. Thus $F_x \in \mathsf{L}^1(\mathbb{R};\mathbb{R})$ and

$$\|F_x\|_1 \leq \|P_{I,x}\|_1\|\mu\|_{\mathrm{TV}} = \|\mu\|_{\mathrm{TV}},$$

using Lemma 7.2.5(iii), $\|\cdot\|_{\mathrm{TV}}$ being the total variation norm as in Theorem 3.8.63. We conclude that $F \in \overline{\mathsf{h}}^1(\mathbb{C}_I;\mathbb{R})$.

(b) $\Longrightarrow$ (a) We let $F \in \overline{\mathsf{h}}^1(\mathbb{C}_I;\mathbb{R})$ and let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence converging to 0 from above. Define $f_j \in \mathsf{L}^1(\mathbb{R};\mathbb{R})$ by $f_j(\eta) = F(x_j + \mathrm{i}\eta)$. Note that, since $F \in \overline{\mathsf{h}}^1(\mathbb{C}_I;\mathbb{R})$, $(f_j)_{j\in\mathbb{Z}_{>0}}$ is a bounded sequence in $\mathsf{L}^1(\mathbb{R};\mathbb{R}) \subseteq \mathscr{M}_{\mathrm{fin}}(\mathbb{R};\mathbb{R}) = (\mathsf{C}^0_0(\mathbb{R};\mathbb{R}))^*$ (this characterisation of the dual space is). It follows by the Banach–Alaoglu Theorem that there is a subsequence $(f_{j_k})_{k\in\mathbb{Z}_{>0}}$ and $\mu \in \mathscr{M}_{\mathrm{fin}}(\mathbb{R};\mathbb{R})$ such that $(f_{j_k})_{k\in\mathbb{Z}_{>0}}$ converges weak-* to μ. Now define

Theorem 7.3.6, DLC:13 ref

$$\begin{aligned} F_j\colon \mathbb{C}_I &\to \mathbb{R} \\ z &\mapsto F(x_j + z). \end{aligned}$$

By Lemma 3, F_j is bounded and continuous on $\mathbb{C}_{\mathrm{cl}(I)}$ and harmonic in $\mathbb{C}_I$. By Lemma 2 we have, for $z = x + \mathrm{i}y \in \mathbb{C}_I$,

$$F_j(z) = \int_{\mathbb{R}} P_{I,x}(y-\eta)F_j(\mathrm{i}\eta)\,\mathrm{d}\eta = \int_{\mathbb{R}} P_{I,x}(y-\eta)f_j(\eta), \qquad j \in \mathbb{Z}_{>0}.$$

Since $P_{I,x} \in \mathsf{C}^0_0(\mathbb{R};\mathbb{R})$, we have

$$\begin{aligned} F(z) &= \lim_{k\to\infty} F(x_j + z) = \lim_{k\to\infty} F_{j_k}(z) \\ &= \lim_{k\to\infty} \int_{\mathbb{R}} P_{I,x}(y-\eta)f_{j_k}(\eta)\,\mathrm{d}\eta \\ &= \int_{\mathbb{R}} P_{I,x}(y-\eta)\,\mathrm{d}\mu(\eta), \end{aligned}$$

by definition of weak-* convergence and the pairing of $\mathsf{C}^0_0(\mathbb{R};\mathbb{R})$ with its dual. This concludes this part of the proof.

(iv) As in part (ii) we take $(0,\pi)$.

(a) $\Longrightarrow$ (b) We suppose that

$$F(x+\mathrm{i}y) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\sin(x)}{\cosh(y-\eta)-\cos(x)}\,\mathrm{d}\mu(\eta) + \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\sin(\pi-x)}{\cosh(y-\eta)-\cos(\pi-x)}\,\mathrm{d}\nu(\eta).$$

As in the proof of (a) $\Longrightarrow$ (b), we can verify the mean value property of F using Fubini's Theorem. This shows that F is harmonic. Moreover, since

$$F(x+\mathrm{i}y) = \int_{\mathbb{R}} P_{I,x}(y-\eta)\,\mathrm{d}\mu(\eta) + \int_{\mathbb{R}} P_{I,\pi-x}(y-\eta)\,\mathrm{d}\nu(\eta),$$

Proposition 3.9.9, VIB:07a

i.e., F is defined by a sum of convolutions, we can use Lemma 7.2.5(iii) and to give

$$\|F_x\|_1 \leq \|P_{I,x}\|_1\|\mu\|_{\mathrm{TV}} + \|P_{I,\pi-x}\|_1\|\nu\|_{\mathrm{TV}},$$

which gives $F \in \mathsf{h}^1(\mathbb{C}_I;\mathbb{R})$.

(b) $\Longrightarrow$ (a) Let $F \in \overline{\mathsf{h}}^1(\mathbb{C}_I;\mathbb{R})$. Let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $(0,\frac{\pi}{2})$ converging to 0 from above. Define $(f_j)_{j\in\mathbb{Z}_{>0}}$ and $(g_j)_{j\in\mathbb{Z}_{>0}}$ by

$$f_j(\eta) = F(x_j + \mathrm{i}\eta), \quad g_j(\eta) = F(\pi - x_j + \mathrm{i}\eta).$$

We think of these as together defining a sequence of measures that we denote by $((f_j, g_j))_{j\in\mathbb{Z}_{>0}}$ on $\mathrm{bd}(\mathbb{C}_I)$ in the obvious way:

$$(f_j, g_j)(A) = \int_{A\cap\mathrm{i}\mathbb{R}} f_j(\eta)\,\mathrm{d}\eta + \int_{A\cap(\pi+\mathrm{i}\mathbb{R})} g_j(\eta)\,\mathrm{d}\eta.$$

Since $F \in \overline{\mathsf{h}}^1(\mathbb{C}_I;\mathbb{R})$, these are bounded sequences in $\mathsf{L}^1(\mathrm{bd}(\mathbb{C}_I);\mathbb{R}) \subseteq \mathscr{M}_{\mathrm{fin}}(\mathrm{bd}(\mathbb{C}_I);\mathbb{R}) \simeq (\mathsf{C}^0_0(\mathrm{bd}(\mathbb{C}_I);\mathbb{R}))^*$ and so, by the same arguments as above for $I = (0,\infty)$, there exists a subsequence $((f_{j_k}, g_{j_k}))_{k\in\mathbb{Z}_{>0}}$ that converges weak-$*$ to a limit $(\mu,\nu) \in \mathscr{M}_{\mathrm{fin}}(\mathrm{bd}(\mathbb{C}_I);\mathbb{R})$. Now define

$$\begin{aligned} F_j \colon \mathbb{C}_I &\to \mathbb{R} \\ z &\mapsto F(x_j + \frac{\pi - 2x_j}{\pi} z). \end{aligned}$$

One directly verifies that F_j is harmonic. By Lemma 3, F_j is bounded and continuous in $\mathbb{C}_{\mathrm{cl}(I)}$ and harmonic in $\mathbb{C}_I$. Therefore, by Lemma 2, we have, for $z = x + \mathrm{i}y \in \mathbb{C}_I$,

$$\begin{aligned} F_j(z) &= \int_{\mathbb{R}} P_{I,x}(y-\eta)F_j(\mathrm{i}\eta)\,\mathrm{d}\eta + \int_{\mathbb{R}} P_{I,\pi-x}(y-\eta)F_j(\pi+\mathrm{i}\eta)\,\mathrm{d}\eta \\ &= \int_{\mathbb{R}} P_{I,x}(y-\eta)f_j(\eta)\,\mathrm{d}\eta + \int_{\mathbb{R}} P_{I,\pi-x}(y-\eta)g_j(\eta)\,\mathrm{d}\eta \end{aligned}$$

for $j \in \mathbb{Z}_{>0}$. Since $P_{I,x}, P_{I,\pi-x} \in \mathsf{C}^0_0(\mathbb{R};\mathbb{R})$, we can think of $(P_{I,x}, P_{I,\pi-x})$ as an element of $\mathsf{C}^0_0(\mathrm{bd}(\mathbb{C}_I);\mathbb{R})$ in the obvious way:

$$(P_{I,x}, P_{I,\pi-x})(z) = \begin{cases} P_{I,x}(\mathrm{Im}(z)), & \mathrm{Re}(z) = 0, \\ P_{I,\pi-x}(\mathrm{Im}(z)), & \mathrm{Re}(z) = \pi. \end{cases}$$

We then have

$$\begin{aligned} F(z) &= \lim_{k\to\infty} F(x_{j_k} + \frac{\pi - 2x_{j_k}}{\pi} z) = \lim_{k\to\infty} F_{j_k}(z) \\ &= \lim_{k\to\infty}\left(\int_{\mathbb{R}} P_{I,x}(y-\eta)f_{j_k}(\eta)\,\mathrm{d}\eta + \int_{\mathbb{R}} P_{I,\pi-x}(y-\eta)g_{j_k}(\eta)\,\mathrm{d}\eta\right) \\ &= \int_{\mathbb{R}} P_{I,x}(y-\eta)\,\mathrm{d}\mu(\eta) + \int_{\mathbb{R}} P_{I,\pi-x}(y-\eta)\,\mathrm{d}\nu(\eta), \end{aligned}$$

by definition of weak-$*$ convergence and the pairing of $\mathsf{C}^0_0(\mathrm{bd}(\mathbb{C}_I);\mathbb{R})$ with its dual, $\mathscr{M}_{\mathrm{fin}}(\mathrm{bd}(\mathbb{C}_I);\mathbb{R})$. ■ Theorem 7.3.6, DLC:13

While the results of the preceding theorem are satisfactory in that they give a useful integral representation for harmonic functions. There are, however, some unsatisfactory elements of the development.

1. If $I \subseteq \mathbb{R}$ is an open interval and if $F \in \mathsf{h}^p(\mathbb{C};\mathbb{F})$, then it is not clear what is the relationship between the values of F as we approach the boundary of $\mathbb{C}_I$ and the values of the boundary functions (or measures) that give rise to the Poisson integral representation for F.
2. One might hope that, in some interesting cases, in the case of $p = 1$ the Poisson integral representation involves boundary functions and not boundary measures.

We shall address these questions by showing that (1) the boundary functions (or measures) can be chosen in such a way that they are the nontangential limit of F as it approaches the boundary () and (2) if F is holomorphic, then its boundary measure is absolutely continuous with respect to the Lebesgue measure, and so is determined by an integrable function. The complete development of these results requires some substantial and seemingly unmotivated diversions. This is the more complicated for us here, as contrasted to the standard presentations, because we work with harmonic functions on arbitrary vertical strips, and so the results must be flexible enough to handle all possibilities.

what

where

The programme that will occupy us for the next few sections is outlined as follows.

1. We first study the Hardy–Littlewood maximal function. This function has two attributes that will be of interest. One is the exact character of the maximal function when constructed for functions in $\mathsf{L}^p(\mathbb{R};\mathbb{R})$. We shall see here that, just as with Poisson integral representations, the case of $p = 1$ is distinguished in character. The other attribute of the maximal function is that it provides a useful bound for other sorts of constructions involving Poisson integrals that will be useful for ensuring pointwise limits.
2. Next we make use of the latter attribute of the maximal function to prove the existence of nontangential boundary limits for functions in the harmonic Hardy spaces. These nontangential boundary limits exist even in the case of $p = 1$ when the Poisson integrals involve boundary measures. In this case, one sees that the difference between the boundary measure and the boundary function is a singular measure.
3. We close the section by considering some constructions with subharmonic functions and majorants of subharmonic functions by harmonic functions. We shall see that the harmonic majorant can be chosen to be a Poisson integral. We shall use this fact in to show that the Poisson integral representation of holomorphic functions can be made using boundary functions, even though boundary measures are required for general harmonic functions in the case $p = 1$.

where

7.2.3 Nontangential limits for the harmonic Hardy spaces on vertical strips

8 Lemma *Let* $p \in [1,\infty)$*, let* $I \in \mathscr{I}$ *with* $\mathrm{cl}(I) \in \{[a,\infty),(-\infty,b]\}$*, and let* $F \in \overline{\mathsf{H}}^p(\mathbb{C}_I;\mathbb{C})$.

Then, for $\mathrm{z} \in \mathrm{int}(\mathrm{I})$,

$$|\mathrm{F}(\mathrm{z})| \leq \begin{cases} \left(\dfrac{2}{\pi(\mathrm{Re}(\mathrm{z})-\mathrm{a})}\right)^{1/\mathrm{p}} \|\mathrm{I}\|_{\mathsf{H}^{\mathrm{F}},\mathrm{p}}, & \mathrm{cl}(\mathrm{I}) = [\mathrm{a}, \infty), \\ \left(\dfrac{2}{\pi(\mathrm{b}-\mathrm{Re}(\mathrm{z}))}\right)^{1/\mathrm{p}} \|\mathrm{I}\|_{\mathsf{H}^{\mathrm{F}},\mathrm{p}}, & \mathrm{cl}(\mathrm{I}) = (-\infty, \mathrm{b}]. \end{cases}$$

Proof We will work out the case of $\mathrm{cl}(I) = [a, \infty)$. The other case follows similarly.

Let $z \in [a, \infty)$ and let $\rho = \mathrm{Re}(z) - a$. By Cauchy's Integral Formula we have

$$F(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(z + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta$$

for every $r \in (0, \rho]$. Thus

$$\begin{aligned} \frac{1}{\mathrm{vol}(\overline{\mathsf{D}}(\rho, z))} \int_{\overline{\mathsf{D}}(\rho,z)} F(\zeta)\,\mathrm{d}\zeta &= \frac{1}{\pi\rho^2} \int_0^{\rho} \int_{-\pi}^{\pi} F(z + r\mathrm{e}^{\mathrm{i}\theta}) r \mathrm{d}\theta \mathrm{d}r \\ &= \frac{2}{\rho^2} \int_0^{\rho} F(z) r\,\mathrm{d}r = F(z). \end{aligned}$$

We then use Hölder's inequality to compute

$$\begin{aligned} |F(z)| &\leq \frac{1}{\mathrm{vol}(\overline{\mathsf{D}}(\rho, z))} \int_{\overline{\mathsf{D}}(\rho,z)} |F(\zeta)|\,\mathrm{d}\zeta \\ &\leq \frac{1}{\pi\rho^2} \Big(|F(\zeta)|^p\Big)^{1/p} \mathrm{vol}(\overline{\mathsf{D}}(\rho, z))^{1/p'} \\ &\leq \Big(\frac{1}{\pi\rho^2} \int_{\mathbb{C}_{[a,a+2\rho]}} |F(\zeta)|^p\,\mathrm{d}\zeta\Big)^{1/p} \\ &\leq \Big(\frac{1}{\pi\rho^2} \int_0^{2\rho} |F(a + x + \mathrm{i}y)|^p\,\mathrm{d}y\mathrm{d}x\Big)^{1/p} \\ &\leq \Big(\frac{2}{\pi\rho}\Big)^{1/p} \|I\|_{\mathsf{H}^F,p}, \end{aligned}$$

which is the desired result. ▼

Now let $(F_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\overline{\mathsf{H}}^p(\mathbb{C}_I; \mathbb{C})$. We first consider the cases $\mathrm{cl}(I) \in \{[a, b], [a, \infty), (-\infty, b]\}$. If $K \subseteq \mathbb{C}_{\mathrm{int}(I)}$ is compact, then there exists $J \subseteq \mathrm{int}(I)$ compact such that $K \subseteq \mathbb{C}_J$. By the lemmata above, if $p \in [1, \infty)$ we have that $(F_j|K)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in the ∞-norm. Of course, this is also true when $p = \infty$, and so we conclude that (F_j) converges in the compact-open topology of $\mathsf{H}(\mathbb{C}_{\mathrm{int}(I)}; \mathbb{C})$ to an holomorphic limit F. Next let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that $\|I\|_{\mathsf{H}^{F_j - F_k},p} < \epsilon$ for $j, k \geq N$. For $y \in \mathrm{int}(I)$ and for $j \geq N$ we have ref

$$\|F_j - F\|_{\{y\},p} = \lim_{k\to\infty} \|F_j - F_k\|_{\{y\},p} \leq \lim_{k\to\infty} \|I\|_{\mathsf{H}^{F_j - F_k},p} < \epsilon.$$

As this holds for every $y \in \operatorname{int}(I)$ we conclude that $\|I\|_{\mathsf{H}^{F_j - F},p} < \epsilon$ for $j \geq N$, giving the desired convergence.

The Poisson kernel for the strip $\mathcal{S}_{a,b} = (a, b) \times \mathbb{R} \subseteq \mathbb{R}^2 \simeq \mathbb{C}$ is

$$P_{a,b}(x, y) = \frac{1}{2(b-a)} \frac{\sin(\pi \frac{x-a}{b-a})}{\cosh(\pi \frac{y}{b-a}) - \cos(\pi \frac{x-a}{b-a})}.$$

We directly compute that the Laplacian of $P_{a,b}$ is zero, and so $P_{a,b}$ is harmonic. For $f_a, f_b \in \mathsf{L}^p(\mathbb{R}; \mathbb{C})$ we denote $P_{a,b} * (f_a, f_b) \colon \mathcal{S}_{a,b} \to \mathbb{C}$ by

$$P_{a,b} * (f_a, f_b)(x + \mathrm{i}y) = \int_{\mathbb{R}} f_a(t) P_{a,b}(x, t - y)\, \mathrm{d}t + \int_{\mathbb{R}} f_b(t) P_{a,b}(a + b - x, t - y)\, \mathrm{d}t.$$

Similarly, if $\mu_a, \mu_b \in \mathscr{M}_{\mathrm{fin}}(\mathbb{R}; \mathbb{C})$, then we define

$$P_{a,b} * (\mu_a, \mu_b)(x + \mathrm{i}y) = \int_{\mathbb{R}} P_{a,b}(x, t - y)\, \mathrm{d}\mu_a(t) + \int_{\mathbb{R}} P_{a,b}(a + b - x, t - y)\, \mathrm{d}\mu_b(t). \quad (7.1)$$

If $r \in \mathbb{R}_{>0}$ and $z_0 \in \mathcal{S}_{a,b}$ is such that $\overline{\mathsf{D}}(r, z_0) \subseteq \mathcal{S}_{a,b}$, then we may use the Dominated Convergence Theorem to directly compute

$$\int_{\mathrm{bd}(\overline{\mathsf{D}}(r,z_0))} P_{a,b} * (f_a, f_b)(z)\, \mathrm{d}z = 0,$$

giving that $P_{a,b} * (f_a, f_b)$ is harmonic on $\mathcal{S}_{a,b}$, and similarly for finite Radon measures.
DVW:61 One verifies that $y \mapsto P_{a,b}(x + \mathrm{i}y)$ is an approximate identity as $x \downarrow a$, and so $y \mapsto P_{a,b}((a + b - x) + \mathrm{i}y)$ is also an approximate identity as $x \uparrow b$. The usual
Theorem I.3.1, JBG:07 arguments then yield that we have the following flavours of convergence of the limits

$$\lim_{x \downarrow a} \int_{\mathbb{R}} P_{a,b} * (f_a, f_b)(x + \mathrm{i}y)\, \mathrm{d}y = f_a, \qquad \lim_{x \uparrow b} \int_{\mathbb{R}} P_{a,b} * (f_a, f_b)(x + \mathrm{i}y)\, \mathrm{d}y = f_b :$$

1. convergence in L^p for $f_a, f_b \in \mathsf{L}^p(\mathbb{R}; \mathbb{C})$, $p \in [1, \infty)$;
2. weak-$*$ convergence for $f_a, f_b \in \mathsf{L}^\infty(\mathbb{R}; \mathbb{C})$ (keeping in mind that $\mathsf{L}^1(\mathbb{R}; \mathbb{C})$ is the continuous dual of $\mathsf{L}^\infty(\mathbb{R}; \mathbb{C})$);
3. uniform convergence if f_a and f_b are uniformly continuous and bounded;
4. weak-$*$ convergence if $f_a = \mu_a$ and $f_b = \mu_b$ are finite Radon measures (keeping in mind that the space of Radon measures is the continuous dual to $\mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R}; \mathbb{C})$).

Let $\mathsf{h}^p(\mathbb{C}_I; \mathbb{R})$ be the set of $\mathbb{R}$-valued functions F harmonic on $\mathbb{C}_{\mathrm{int}(I)}$ for which $x \mapsto$
Theorem I.3.5, JBG:07 $\|F_x\|_p$ is bounded on I, where $F_x(y) = F(x + \mathrm{i}y)$. Standard arguments using the
ref Banach–Alaoglu Theorem show that

1. for $p \in (1, \infty]$, the mapping $(f_a, f_b) \mapsto P_{a,b} * (f_a, f_b)$ is an isometry of $\mathsf{L}^p(\mathbb{R}; \mathbb{R})^2$ onto $\mathsf{h}^p(\mathbb{C}_I; \mathbb{R})$, and

2. for $p = 1$, the mapping (μ_a, μ_b) is an isometry of $\mathscr{M}_{\mathrm{fin}}(\mathbb{R};\mathbb{R})^2$ onto $\mathsf{h}^1(\mathbb{C}_I;\mathbb{R})$.

Let us now sketch the argument one may use to prove the existence of nontangential limits for functions in $\mathsf{h}^p(\mathbb{C}_I;\mathbb{R})$, from which the result for functions in $\overline{\mathsf{H}}^p(\mathbb{C}_I;\mathbb{C})$ follows. We take $F \in \mathsf{h}^1(\mathbb{C}_I;\mathbb{C})$ and write $F = P_{a,b}*(\mu_a, \mu_b)$ for some $\mu_a, \mu_b \in \mathscr{M}_{\mathrm{fin}}(\mathbb{R};\mathbb{C})$. We shall consider the boundary $\mathbb{C}_{\{a\}}$, the case for the other boundary following in a similar manner. For $y \in \mathbb{R}$, $\alpha \in \mathbb{R}_{>0}$, and $\ell \in (0, \frac{1}{2}(a+b))$, denote

$$\Gamma^+_{a,\alpha,\ell}(y) = \{(a + \mathrm{i}y) + (\xi + \mathrm{i}\eta) \in \mathbb{C} \mid \ |\eta| < \alpha\xi,\ |\xi| \le \ell\}$$

be a truncated cone with vertex at $a + \mathrm{i}y$ in $\mathbb{C}$ pointing into $\mathbb{C}_{\mathrm{int}(I)}$; see Figure 7.3. (The subscript "+" connotes the fact that there will also be a cone at the right

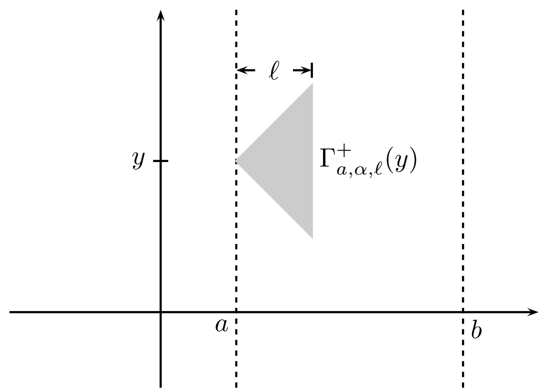


Figure 7.3 The truncated cone $\Gamma^+_{a,\alpha,\ell}(y)$

boundary that will point in the "negative direction.") First let us suppose that μ_a is absolutely continuous with respect to the Lebesgue measure, i.e., $\mu_a = f_a\lambda$ for some $f_a \in \mathsf{L}^1(\mathbb{R};\mathbb{R})$. Denote

$$\begin{aligned} T^+_{a,\alpha,\ell}F\colon \mathbb{R} &\to \overline{\mathbb{R}}_{\ge 0} \\ y &\mapsto \sup\{|F(z)| \mid \ z \in \Gamma^+_{a,\alpha,\ell}(y)\}. \end{aligned}$$

For $y \in \mathbb{R}$ define

$$\Omega^+_{a,\alpha,\ell}F(y) = \limsup_{\substack{z\to a+\mathrm{i}y\\ z\in\Gamma^+_{a,\alpha}(y)}} F(z) - \liminf_{\substack{z\to a+\mathrm{i}y\\ z\in\Gamma^+_{a,\alpha}(y)}} F(z).$$

Clearly, $\Omega^+_{a,\alpha,\ell}F(y) = 0$ for some $\alpha \in \mathbb{R}_{>0}$ and some $\ell \in (0, \frac{1}{2}(a+b))$ if and only if F has a nontangential limit at $a + \mathrm{i}y$. We shall show that, for a given $\alpha \in \mathbb{R}_{>0}$ and $\ell \in (0, \frac{1}{2}(a+b))$, $\Omega^+_{a,\alpha,\ell}F(y) = 0$ for almost every $y \in \mathbb{R}$. The following lemma allows us to make the crucial step in the proof, and is the place where one has to keep track of the specific form of the Poisson kernel for the strip.

9 Lemma *For* $\mathrm{I} \in \mathscr{I}$ *satisfying* $\mathrm{cl(I)} = [\mathrm{a}, \mathrm{b}]$, *and for* $\mathrm{f_a}, \mathrm{f_b} \in \mathsf{L}^1(\mathbb{R}; \mathbb{R})$, *let* $\mathrm{F} = \mathrm{P_{a,b}} * (\mathrm{f_a}, \mathrm{f_b}) \in \mathsf{h}^1(\mathbb{C}_\mathrm{I}; \mathbb{R})$. *Then, for each* $\alpha \in \mathbb{R}_{>0}$ *and* $\ell \in (0, \frac{1}{2}(\mathrm{a} + \mathrm{b}))$, *there exists* $\mathrm{A}_{\mathrm{a},\alpha,\ell} \in \mathbb{R}_{>0}$ *such that*

$$\mathrm{T}^+_{\mathrm{a},\alpha,\ell}\mathrm{F}(\mathrm{y}) \leq \mathrm{A}_{\mathrm{a},\alpha,\ell}(\mathrm{Mf_a}(\mathrm{y}) + \mathrm{Mf_b}(\mathrm{y})), \qquad \mathrm{y} \in \mathbb{R},$$

page 24, JBG:07 *where* $\mathrm{Mf_a}$ *and* $\mathrm{Mf_b}$ *are the Hardy–Littlewood maximal functions for* $\mathrm{f_a}$ *and* $\mathrm{f_b}$.

Proof Given the form (7.1) for $P_{a,b} * (f_a, f_b)$ we may without loss of generality take $y = 0$.

For $t \in \mathbb{R}$ let

$$\psi_{a,x,y}(t) = \sup\{P_{a,b}(x, s - y) \mid |s| > t\}, \qquad \psi_{b,x,y}(t) = \sup\{P_{a,b}(a + b - x, s - y) \mid |s| > t\}.$$

The functions $\psi_{a,x,y}, \psi_{b,x,y} \colon \mathbb{R} \to \mathbb{R}_{>0}$ have the following features:

1. they are even;
2. they are strictly decreasing on $\mathbb{R}_{>0}$ (and so strictly increasing on $\mathbb{R}_{<0}$);
3. for each $t \in \mathbb{R}$, $P_{a,b}(x, t - y) \leq \psi_{a,x,y}(t)$ and $P_{a,b}(b - a, t - y) \leq \psi_{b,x,y}(t)$.

For the next little while let us fix $x + \mathrm{i}y \in \Gamma^+_{a,\alpha,\ell}(0)$. We then compute (after a moments reflection about what the graphs of the functions $\psi_{a,x,y}$ and $\psi_{b,x,y}$ look like as a consequence of properties 1 and 2),

$$\begin{aligned}\int_\mathbb{R} \psi_{a,x,y}(t)\,\mathrm{d}t &= \int_{t<-|y|} P_{a,b}(x,t)\,\mathrm{d}t + \int_{-|y|}^{|y|} P_{a,b}(x,0)\,\mathrm{d}t + \int_{t>|y|} P_{a,b}(x,t)\,\mathrm{d}t \\ &= \int_\mathbb{R} P_{a,b}(x,t)\,\mathrm{d}t + \frac{|y|\sin(\pi\frac{x-a}{b-a})}{(b-a)(1-\cos(\pi\frac{x-a}{b-a}))} \\ &\leq \frac{x-a}{b-a}\Big(1 + \frac{\alpha\sin(\pi\frac{x-a}{b-a})}{(1-\cos(\pi\frac{x-a}{b-a}))}\Big) \leq 1 + 2\alpha,\end{aligned}$$

DVW:61 where we use the computations of and the fact that the function

$$\theta \mapsto \frac{\theta\sin\theta}{1-\cos\theta}$$

is bounded by 2 on $[0, \pi]$. A similarly style computation gives

$$\int_\mathbb{R} \psi_{b,x,y}(t)\,\mathrm{d}t \leq 1 + 2\alpha.$$

Now, by properties 1 and 2 we can arbitrarily well approximate (in the L^1-sense)
Figure I.4, JBG:07 $\psi_{a,x,y}$ and $\psi_{b,x,y}$ from below by step functions with the same properties (see). If $\sigma_{a,x,y}$ and $\sigma_{b,x,y}$ are such step functions we can write them as

$$\sigma_{a,x,y}(t) = \sum_{j=1}^N \alpha_j\chi(-y_j, y_j), \qquad \sigma_{b,x,y}(t) = \sum_{j=1}^N \beta_j\chi(-y_j, y_j),$$

for some $0 < y_1 < \cdots < y_N$ and some $a_1, \dots, a_N \in \mathbb{R}_{>0}$, and where, by our estimates just preceding,

$$\sum_{j=1}^{N} 2y_j\alpha_j \leq 1 + 2\alpha, \qquad \sum_{j=1}^{N} 2y_j\beta_j \leq 1 + 2\alpha.$$

For such step functions we compute

$$\begin{aligned}\left|\int_{\mathbb{R}} \sigma_{a,x,y}(t) f_a(t)\, \mathrm{d}t\right| &\leq \int_{\mathbb{R}} \sigma_{a,x,y}(t)|f_a(t)|\, \mathrm{d}t \\ &\leq \sum_{j=1}^{N} 2y_j\alpha_j \frac{1}{2y_j} \int_{-y_j}^{y_j} |f_a(t)|\, \mathrm{d}t \\ &\leq (1 + 2\alpha)Mf_a(0).\end{aligned}$$

By the Monotone Convergence Theorem, taking the supremum over all such step functions, we arrive at the conclusion

$$\int_{\mathbb{R}} |f_a(t)|\psi_{a,x,y}(t)\, \mathrm{d}t \leq (1 + 2\alpha)Mf_a(0)$$

and, similarly

$$\int_{\mathbb{R}} |f_b(t)|\psi_{b,x,y}(t)\, \mathrm{d}t \leq (1 + 2\alpha)Mf_b(0).$$

Thus we finally compute

$$\begin{aligned}|F(x + \mathrm{i}y)| &\leq \int_{\mathbb{R}} |f_a(t)|P_{a,b}(x, t - y)\, \mathrm{d}t + \int_{\mathbb{R}} |f_b(t)|P_{a,b}(a + b - x, t - y)\, \mathrm{d}t \\ &\leq \int_{\mathbb{R}} |f_a(t)|\psi_{a,x,y}(t)\, \mathrm{d}t + \int_{\mathbb{R}} |f_b(t)|\psi_{b,x,y}(t)\, \mathrm{d}t \\ &\leq (1 + 2\alpha)(Mf_a(0) + Mf_b(0)),\end{aligned}$$

as desired, noting that this holds for any $x + \mathrm{i}y \in \Gamma^+_{a,\alpha,\ell}(0)$. ▼

By definitions we have $\Omega^+_{a,\alpha,\ell}F(y) \leq 2T^+_{a,\alpha,\ell}F(y)$. By the Hardy–Littlewood Maximal Theorem, for $\beta \in \mathbb{R}_{>0}$ we have

$$\lambda(\{y \in \mathbb{R} \mid\ Mf_a(y) > \beta\}) \leq \frac{2}{\beta}\|f_a\|_1, \quad \lambda(\{y \in \mathbb{R} \mid\ Mf_b(y) > \beta\}) \leq \frac{2}{\beta}\|f_b\|_1.$$

Now, given the lemma, the estimates just preceding, and the fact that for $\sigma, \tau \in \mathbb{R}_{\geq 0}$

we have $\sigma + \tau \leq 2\max\{\sigma, \tau\}$, we have

$$\begin{aligned}\lambda(\{y \in \mathbb{R} \mid \Omega_{a,\alpha,\ell}F(y) > \epsilon\}) &\leq \lambda(\{y \in \mathbb{R} \mid 2T^{+}_{a,\alpha,\ell}F(y) > \epsilon\}) \\ &\leq \lambda(\{y \in \mathbb{R} \mid 2A_{a,\alpha,\ell}(Mf_a(y) + Mf_b(y)) > \epsilon\}) \\ &\leq \lambda\Big(\Big\{y \in \mathbb{R} \Bigm| Mf_a(y) > \frac{\epsilon}{4A_{a,\alpha,\ell}}\Big\}\Big) \\ &\quad + \lambda\Big(\Big\{y \in \mathbb{R} \Bigm| Mf_b(y) > \frac{\epsilon}{4A_{a,\alpha,\ell}}\Big\}\Big) \\ &\leq \frac{8A_{a,\alpha,\ell}}{\epsilon}(\|f_a\|_1 + \|f_b\|_1).\end{aligned}$$

Now let $g_a, g_b \in \mathsf{L}^1(\mathbb{R};\mathbb{R}) \cap \mathsf{C}^0_{\mathrm{cpt}}(\mathbb{R};\mathbb{R})$ be such that

$$\|f_a + g_a\|_1, \|f_b + g_b\|_1 < \frac{\epsilon^2}{16A_{a,\alpha,\ell}}.$$

If $G = P_{a,b} * (g_a, g_b)$ then $G \in \mathsf{C}^0(\mathbb{C}_{[a,b]};\mathbb{R}) \cap \mathsf{h}^1(\mathbb{C}_I;\mathbb{R})$. Therefore, since G has nontangential limits at every point $a + \mathrm{i}y$, $y \in \mathbb{R}$, by the estimate

$$\lambda(\{y \in \mathbb{R} \mid \Omega_{a,\alpha,\ell}(F + G)(y) > \epsilon\}) \leq \frac{8A_{a,\alpha,\ell}}{\epsilon}(\|f_a + g_a\|_1 + \|f_b + g_b\|_1),$$

we have

$$\begin{aligned}\lambda(\{y \in \mathbb{R} \mid \Omega_{a,\alpha,\ell}F(y) > \epsilon\}) &= \lambda(\{y \in \mathbb{R} \mid \Omega_{a,\alpha,\ell}(F + G)(y) > \epsilon\}) \\ &\leq \frac{8A_{a,\alpha,\ell}}{\epsilon}\Big(\frac{\epsilon^2}{16A_{a,\alpha,\ell}} + \frac{\epsilon^2}{16A_{a,\alpha,\ell}}\Big) = \epsilon.\end{aligned}$$

Thus, using Proposition 2.3.3(ii),

$$\lambda(\{y \in \mathbb{R} \mid \Omega_{a,\alpha,\ell}F(y) > 0\}) = \lambda\Big(\cap_{j\in\mathbb{Z}_{>0}}\{y \in \mathbb{R} \mid \Omega_{a,\alpha,\ell}F(y) > \tfrac{1}{j}\}\Big) \leq \lim_{j\to\infty} \frac{1}{j} = 0,$$

which gives the existence of nontangential limits for $p = 1$ and absolutely continuous boundary measures. The same conclusion for $p = 1$ and measures with nonsingular part follows from the argument made of . Lemma I.5.4, JBG:07

For $p = \infty$, for each $\beta \in \mathbb{R}_{>0}$ we write $g_{a,\beta} = \chi[-\beta,\beta]f_a$, $g_{b,\beta} = \chi[-\beta,\beta]f_b$, and $h_a = f_a - g_a$ and $h_b = f_b - g_b$. We also denote $G_\beta = P_{a,b} * (g_a, g_b)$ and $H_\beta = P_{a,b} * (h_a, h_b)$. Since $g_a, g_b \in \mathsf{L}^1(\mathbb{R};\mathbb{R})$, the function G_β possesses nontangential limits at $a + \mathrm{i}y$ for almost every $y \in \mathbb{R}$. By we have that H_β has nontangential limit 0 at $a + \mathrm{i}y$ for every $y \in (-\beta, \beta)$. Thus $F = G_\beta + H_\beta$ has nontangential limits at $a + \mathrm{i}y$ for almost every $y \in (-\beta, \beta)$. Letting $\beta \to \infty$ gives the existence of nontangential limits for $p = \infty$. Theorem 1, DVW:61

For $p \in (1, \infty)$ we define $g_a, g_b \in \mathsf{L}^1(\mathbb{R};\mathbb{R})$ by

$$g_a(y) = \begin{cases} f_a(y), & |f_a(y)| \geq 1, \\ 0, & \text{otherwise.}\end{cases}$$

In like manner, we define $h_a, h_b \in \mathsf{L}^\infty(\mathbb{R};\mathbb{R})$ by $h_a = f_a - g_a$ and $h_b = f_b - g_b$. The existence of nontangential limits then follows by the $p = 1$ and $p = \infty$ cases above.

It remains to show that the boundary functions defined by nontangential limits for $F \in \overline{\mathsf{H}}^p(\mathbb{C}_I;\mathbb{C})$ are in $\mathsf{L}^p(\mathbb{R};\mathbb{C})$.

Exercises

7.2.1

Section 7.3

Hardy spaces of harmonic functions defined on annuli

In this section we carry out the programme of Section 7.2 for harmonic functions whose domain is an annulus in the complex plane. We will be able to adapt some of our results for functions on vertical strips from Section 7.2, but some of the development will be special to the case of annuli.

Do I need to read this section? This section contains largely technical material that can maybe be skipped at a first read, and then read subsequently when a deeper understanding is required of the holomorphic Hardy spaces on annuli described in Section 7.5. •

7.3.1 Definitions

First we describe the domains on which the harmonic functions we describe in this section will be defined. For an interval $I \subseteq \mathbb{R}_{\geq 0}$, let us denote

$$\mathbb{A}_I = \{z \in \mathbb{C} \mid |z| \in I\}.$$

Note that, if $0 \in I$, then $\mathbb{A}_I$ is a disk. When $\sup I = \infty$, then $\mathbb{A}_I$ is the complement of a disk. If we have both $0 \in I$ and $\sup I = \infty$, then $\mathbb{A}_I = \mathbb{C}$. We depict the generic situation in Figure 7.4. One of the differences between the theory of Hardy

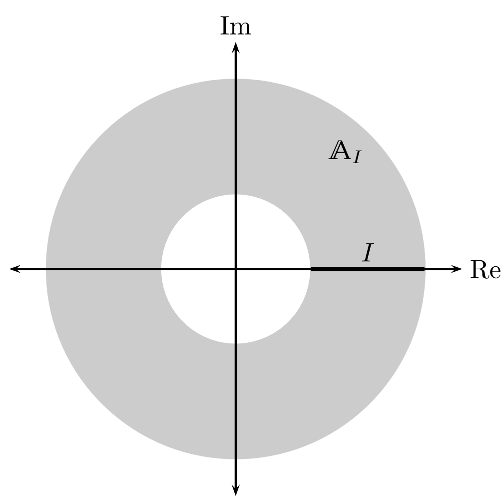


Figure 7.4 An annulus in the complex plane

spaces for vertical strips and the theory for annuli is the reduction of the analysis to a "generic" interval. For example, in the proof of Theorem 7.2.6 we reduced consideration of bounded vertical strips to $\mathbb{C}_{(0,\pi)}$. This was acceptable since all bounded vertical strips admits a standard biholomorphic mapping between them. By Theorem II-3.2.8, this is not true for annuli of the form $\mathbb{A}_{(r,R)}$ for $r, R \in \mathbb{R}_{>0}$ satisfying $r < R$. Indeed, Theorem II-3.2.8 tells us that there is a biholomorphic bijection from $\mathbb{A}_{(r_1,R_1)}$ to $\mathbb{A}_{(r_2,R_2)}$ if and only if $\frac{r_1}{R_1} = \frac{r_2}{R_2}$. Thus the best we can do to make a "canonical" choice of annulus is to fix one of the radii to, say, 1.

For harmonic and holomorphic functions on vertical strips, we used their values on vertical lines to characterise their properties. For annuli, we use the values of functions on circles. A bit of notation we use for this is

$$\mathbb{S} = \{z \in \mathbb{C} \mid |z| = 1\},$$

i.e., $\mathbb{S}$ is the unit circle. A point on $\mathbb{S}$ we denote by $\mathrm{e}^{\mathrm{i}\theta}$. Note that, if $f \colon \mathbb{S} \to \mathbb{F}$, we can identify it with a 2π-periodic function $\hat{f}\mathbb{R}\mathbb{F}$ by the formula $f(\mathrm{e}^{\mathrm{i}\theta}) = \hat{f}(\theta)$. Thus we can define spaces of functions on $\mathbb{S}$ by identifying them with their 2π-periodic analogues. For example, we define $\mathsf{L}^p(\mathbb{S};\mathbb{F})$ by

$$\mathsf{L}^p(\mathbb{S};\mathbb{F}) \simeq \mathsf{L}^p_{\mathrm{per},2\pi}(\mathbb{R};\mathbb{F})$$

with the norm

$$\|f\|_p = \int_0^{2\pi} f(\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta.$$

For a function $F \colon \mathbb{A}_I \to \mathbb{F}$, $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and for $r \in I$, we denote $F_r \colon \mathbb{S} \to \mathbb{F}$ the function defined by $F_r(\mathrm{e}^{\mathrm{i}\theta}) = F(r\mathrm{e}^{\mathrm{i}\theta})$.

We now give a definition of a general class of harmonic functions.

7.3.1 Definition (Functions harmonic on an annulus) For an interval $I \subseteq \mathbb{R}_{\geq 0}$ and for $\mathbb{F} \in \{\mathbb{R};\mathbb{C}\}$, we denote by $\mathsf{h}(\mathbb{A}_I;\mathbb{F})$ the mappings $F \colon \mathbb{A}_I \to \mathbb{F}$ that are harmonic on $\mathrm{int}(\mathbb{A}_I)$. If $\mathrm{int}(\mathbb{A}_I) = \emptyset$, we take the convention that $\mathsf{h}(\mathbb{A}_I;\mathbb{F}) = \mathbb{F}^{\mathbb{A}_I}$. •

Note that, for each interval $I \subseteq \mathbb{R}$, $\mathsf{h}(\mathbb{A}_I;\mathbb{F})$ is a $\mathbb{F}$-vector space with respect to the operations of pointwise addition and scalar multiplication. Note that if $\mathbb{A}_I$ is not open, then we place no restrictions on the values of functions in $\mathsf{h}(\mathbb{A}_I;\mathbb{C})$ when restricted to $\mathrm{bd}(\mathbb{A}_I)$. For more useful classes of functions, we will find it beneficial to assign these boundary values in a meaningful way, as in Section 7.1.2.

Now let us define the particular classes of harmonic functions in which we shall be interested. We begin by defining a large class of these functions, and then we consider a subset of these prescribed by some norm being finite.

7.3.2 Definition (Big harmonic Hardy space on an annulus) Let $I \subseteq \mathbb{R}_{\geq 0}$ be an interval, $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$, and $p \in [1, \infty]$. Denote by $\mathsf{h}^p(\mathbb{A}_I;\mathbb{F})$ the mappings $F \in \mathsf{h}(\mathbb{A}_I;\mathbb{F})$ such that

(i) $F \in \mathsf{h}(\mathbb{A}_I;\mathbb{F})$,

(ii) $F_r \in \mathsf{L}^p(\mathbb{R};\mathbb{F})$ for each $r \in I$, and

(iii) F has nontangential limits.

The space $\mathsf{h}^p(\mathbb{A}_I;\mathbb{F})$ is the ***big harmonic Hardy space***. •

Note that the condition for having nontangential limits is vacuous if I is open. However, if I does contain an endpoint or two, then the requirement of having nontangential limits at these endpoints becomes nonvoid. Indeed, the study of the existence of such nontangential limits will be something that will be of interest to us in Section 7.3.3 below.

Of particular interest is the following subset of the big harmonic Hardy space, where we ask that the L^p-norms on vertical lines be uniformly bounded.

7.3.3 Definition (Harmonic Hardy space on an annulus) Let $I \subseteq \mathbb{R}_{\geq 0}$ be an interval and let $p \in [1,\infty]$. For $F \in \mathsf{h}^p(\mathbb{A}_I;\mathbb{C})$ we denote

$$\|F\|_{\mathsf{h}^p,I} = \sup\{\|F_r\|_p \mid r \in I\}.$$

The subset $\{F \in \mathsf{h}^p(\mathbb{A}_I;\mathbb{C}) \mid \|F\|_{\mathsf{h}^p,I} < \infty\}$ is denoted by $\overline{\mathsf{h}}^p(\mathbb{A}_I;\mathbb{C})$ and we call these spaces the ***harmonic Hardy spaces***. •

We shall be primarily interested in the harmonic Hardy spaces in this chapter. Indeed, we shall not have use in these volumes for the big harmonic Hardy spaces; we will, however, have use for their holomorphic analogues introduced in Section 7.5.1. These we have introduced here the big harmonic Hardy spaces in the interest of symmetry and aesthetics.

7.3.2 Poisson integral representations of harmonic functions on annuli

7.3.3 Nontangential limits for the harmonic Hardy spaces on annuli

Section 7.4

Hardy spaces of holomorphic functions defined on vertical strips

In this section we shall extend our discussion of spaces of harmonic functions defined on vertical strips to holomorphic functions. We shall see that there are some similarities between the harmonic and holomorphic theories, but there are also some important differences, particularly in the case of the $p = 1$.

Throughout the section, we shall make use of basic properties of holomorphic functions discussed in Chapter II-3.

Do I need to read this section? The definitions in this section will be used in Chapter IV-9 and Section V-7.1. So these can certainly be read when they are needed. However, it is not necessary to understand a lot of the technical material to understand the definitions. •

7.4.1 Definitions

We recall from the notion of an holomorphic function of a complex variable. With this in mind, we have the following definition. what?

7.4.1 Definition (Functions holomorphic on a vertical strip) For an interval $I \subseteq \mathbb{R}$, we denote by $\mathsf{H}(\mathbb{C}_I;\mathbb{C})$ the mappings $F\colon \mathbb{C}_I \to \mathbb{C}$ that are holomorphic on $\mathbb{C}_{\mathrm{int}(I)}$. If $\mathrm{int}(I) = \emptyset$, we denote take the convention that $\mathsf{H}(\mathbb{C}_I;\mathbb{C}) = \mathbb{C}^{\mathbb{C}_I}$. •

Note that, for each interval $I \subseteq \mathbb{R}$, $\mathsf{H}(\mathbb{C}_I;\mathbb{C})$ is a $\mathbb{C}$-vector space with respect to the operations of pointwise addition and scalar multiplication. Note that if $\mathrm{int}(I) \neq I$, then we place no restrictions on the values of functions in $\mathsf{H}(\mathbb{C}_I;\mathbb{C})$ when restricted to $\mathrm{bd}(\mathbb{C}_I)$. Indeed, as with our investigation of harmonic functions in Section 7.2, the matter of boundary values is a subject of considerable independent interest.

With the preceding definitions and discussion, we make the following definition.

7.4.2 Definition (Big holomorphic Hardy space on a vertical strip) Let $I \subseteq \mathbb{R}$ be an interval and let $p \in [1, \infty]$. For $F\colon \mathbb{C}_I \to \mathbb{C}$ and $x \in I$ denote $F_x\colon \mathbb{R} \to \mathbb{C}$ by $F_x(y) = F(x + \mathrm{i}y)$. Denote by $\mathsf{H}^p(\mathbb{C}_I;\mathbb{C})$ the mappings $F\colon \mathbb{C}_I \to \mathbb{C}$ such that

(i) $F \in \mathsf{H}(\mathbb{C}_I;\mathbb{C})$,

(ii) $F_x \in \mathsf{L}^p(\mathbb{R};\mathbb{C})$ for each $x \in I$, and

(iii) F has nontangential limits.

The space $\mathsf{H}^p(\mathbb{C}_I;\mathbb{C})$ is the ***big holomorphic Hardy space***. •

Corresponding to Remark 7.1.7–3, we *do not* have that the big holomorphic Hardy spaces are the same as the classical Hardy spaces, e.g., $\mathsf{H}^p(\mathbb{C}_{(0,\infty)};\mathbb{C})$ is not the same as the Hardy space of the right half-plane as there is no requirement that the norms $\|F_x\|_p$ be uniformly bounded in x.

7.4.2 Poisson integral representations of holomorphic functions on vertical strips

As we saw in Section 7.2.2, we can use the boundary values for an harmonic function to prescribe the interior values. In this section we will adapt this idea to holomorphic functions.

7.4.3 Banach spaces of holomorphic functions on vertical strips

Now we study topological vector spaces of holomorphic functions whose domain is a vertical strip. We begin, in this section, by consider Banach spaces of such functions. For such spaces we shall see that there is a useful theory of nontangential limits, such as we presented in Section 7.2.3 for harmonic functions. As we shall see, we can do better for these nontangential limits in the holomorphic case than can be done in the harmonic case.

First we define the spaces we will investigate.

7.4.3 Definition (Holomorphic Hardy spaces on a vertical strip) Let $I \subseteq \mathbb{R}$ be an interval and let $p \in [1,\infty]$. For $F \in \mathsf{H}^p(\mathbb{C}_I;\mathbb{C})$ we denote

$$\|F\|_{\mathsf{H}^p,I} = \sup\{\|F_x\|_p \mid x \in I\}.$$

The subset $\{F \in \mathsf{H}(\mathbb{C}_I;\mathbb{C}) \mid \|F\|_{\mathsf{H}^p,I} < \infty\}$ is denoted by $\overline{\mathsf{H}}^p(\mathbb{C}_I;\mathbb{C})$ and we call these spaces the ***holomorphic Hardy spaces***. •

Let us prove the essential properties of the holomorphic Hardy spaces. If $\mathrm{int}(I) \neq \emptyset$, let us say that $F\colon \mathbb{C}_I \to \mathbb{C}$ ***tends to zero uniformly at infinity*** in $\mathbb{C}_I$ if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists a compact set $K \subseteq I$ such that $|F(z)| < \epsilon$ for $z \in \mathbb{C}_I \setminus \mathbb{C}_K$.

7.4.4 Theorem (Hardy spaces are defined on closed strips) *Let* $\mathrm{I} \subseteq \mathbb{R}$ *be an interval.*

(i) if $\mathrm{F} \in \overline{\mathsf{H}}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}};\mathbb{C})$ *for* $\mathrm{p} \in [1,\infty]$*, then there exists* $\overline{\mathrm{F}} \in \mathsf{H}^{\mathrm{p}}(\mathbb{C}_{\mathrm{cl(I)}};\mathbb{C})$ *such that*

(a) $\mathrm{F}|\mathbb{C}_{\mathrm{int(I)}} = \overline{\mathrm{F}}|\mathbb{C}_{\mathrm{int(I)}}$*,*

(b) $\overline{\mathrm{F}}$ *has nontangential limits;*

(c) $\lim_{\mathrm{x}\to\mathrm{x}_0}\|\overline{\mathrm{F}}_{\mathrm{x}_0} - \mathrm{F}_{\mathrm{x}}\|_{\mathrm{p}} = 0$ *for every* $\mathrm{x}_0 \in \mathrm{cl(I)}$*;*

(ii) if $\mathrm{F} \in \overline{\mathsf{H}}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}};\mathbb{C})$ *for* $\mathrm{p} \in [1,\infty)$*, then* F *tends to zero uniformly at infinity.*

In particular, by part (i), the spaces $\overline{\mathsf{H}}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}};\mathbb{C})$ *and* $\mathsf{H}^{\mathrm{p}}(\mathbb{C}_{\mathrm{cl(I)}};\mathbb{C})$ *are isometric.*

Proof Page 124 of Hoffmann

(i) First we prove the existence of $\overline{F}$ by proving that, for $x_0 \in \mathrm{bd}(I)$, there exists $\overline{F}_{x_0} \in \mathsf{L}^p(\mathbb{R};\mathbb{C})$ such that, if we define

$$\overline{F}(z) = \begin{cases} F(z), & z \in \mathbb{C}_I, \\ \overline{F}_{x_0}(\mathrm{Im}(z)), & z \in \mathrm{bd}(I), \end{cases}$$

then $\overline{F}$ has nontangential limits. Then we prove that $\lim_{x\to x_0} \overline{F}_x = \overline{F}_{x_0}$ in $\mathsf{L}^p(\mathbb{R};\mathbb{C})$.

This part of the result asserts that we can take nontangential limits of $F \in \overline{\mathsf{H}}^p(\mathbb{C}_I;\mathbb{C})$ at endpoints of I, and that the boundary functions are in $\mathsf{L}^p(\mathbb{R};\mathbb{C})$. As such, this part of the result is classical for $\mathrm{cl}(I) = [0,\infty)$ [cf. Garnett 2007, Corollary II.3.2]. By elementary changes of variable, the result also holds for intervals I for which $\mathrm{cl}(I) = [a,\infty)$ or $\mathrm{cl}(I) = (-\infty, b]$ for any $a, b \in \mathbb{R}$. For $I = \{x_0\}$ or $I = \mathbb{R}$, the result is vacuous. The case of $\mathrm{cl}(I) = [a,b]$ follows from [Bakan and Kaijser 2007, Theorem 1 and Corollary 1].

(ii) For $\mathrm{cl}(I) = [a,\infty)$ and $\mathrm{cl}(I) = (-\infty, b]$, this follows from the Theorem on page 125 of [Hoffmann 1962]. For $I = \mathbb{R}$ the result is vacuous, since $\overline{\mathsf{H}}^p(\mathbb{C};\mathbb{C}) = \{0\}$. For $I = \{x_0\}$, the result is vacuous by definition. So the only case remaining to consider is $\mathrm{cl}(I) = [a,b]$. Here we prove a lemma.

1 Lemma *Let* $p \in [1,\infty)$, *let* $I \subseteq \mathbb{R}$ *be an interval with* $\mathrm{cl}(I) = [a,b]$, *let* $a', b' \in \mathbb{R}$ *be such that*

$$[a', b'] \subseteq (a, b),$$

and let $\rho \in \mathbb{R}_{>0}$ *be such that* $[a'-\rho, b'+\rho] \subseteq (a,b)$. *Then, for* $z \in \mathbb{C}_{[a',b']}$,

$$|F(z)| \le \left(\frac{b-a}{\pi\rho^2}\right)^{1/p} \|F\|_{\mathsf{H}^p, I}.$$

Proof Let $z \in \mathbb{C}_{[a',b']}$. By Cauchy's Integral Formula [Conway 1978, Theorem IV.5.4] we have

$$F(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(z + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta$$

for every $r \in (0,\rho]$. Thus

$$\begin{aligned}\frac{1}{\mathrm{vol}(\overline{\mathsf{D}}(\rho,z))}\int_{\overline{\mathsf{D}}(\rho,z)} F(\zeta)\,\mathrm{d}\zeta &= \frac{1}{\pi\rho^2}\int_0^{\rho}\int_{-\pi}^{\pi} F(z + r\mathrm{e}^{\mathrm{i}\theta}) r\,\mathrm{d}\theta\,\mathrm{d}r \\ &= \frac{2}{\rho^2}\int_0^{\rho} F(z) r\,\mathrm{d}r = F(z).\end{aligned}$$

Therefore, by Hölder's Inequality [Cohn 2013, Proposition 3.3.2],

$$\begin{aligned}|F(z)| &\le \frac{1}{\mathrm{vol}(\overline{\mathsf{D}}(\rho,z))}\int_{\overline{\mathsf{D}}(\rho,z)} |F(\zeta)|\,\mathrm{d}\zeta \\ &\le \frac{1}{\pi\rho^2}\left(\int_{\overline{\mathsf{D}}(\rho,z)} |F(\zeta)|^p\,\mathrm{d}\zeta\right)^{1/p} \mathrm{vol}(\overline{\mathsf{D}}(\rho,z))^{1/p'} \\ &\le \left(\frac{1}{\pi\rho^2}\int_{\mathbb{C}_I} |F(\zeta)|^p\,\mathrm{d}\zeta\right)^{1/p} \\ &\le \left(\frac{1}{\pi\rho^2}\right)^{1/p}\left(\int_a^b\int_{\mathbb{R}} |F(x+\mathrm{i}y)|^p\,\mathrm{d}y\,\mathrm{d}x\right)^{1/p} \\ &\le \left(\frac{b-a}{\pi\rho^2}\right)^{1/p} \|F\|_{\mathsf{H}^p, I},\end{aligned}$$

as claimed. ▼

Now let $a', b' \in \mathbb{R}_{>0}$ be such that

$$[a', b'] \subseteq (a, b)$$

and let $\rho \in \mathbb{R}_{>0}$ be sufficiently small that $[a' - \rho, b' + \rho] \subseteq (a, b)$. As in the proof of the lemma above, we have

$$|F(z)| \le \left(\frac{1}{\pi\rho^2}\int_{\overline{\mathsf{D}}(\rho,z)} |F(\zeta)|^p \,\mathrm{d}\zeta\right)^{1/p}$$

for every $z \in \mathbb{C}_{[a',b']}$. Also,

$$\int_{\mathbb{C}_{[a',b']}} |F(\zeta)|^p \,\mathrm{d}\zeta = \int_{a'}^{b'}\int_{\mathbb{R}} |F(x+\mathrm{i}y)|^p \,\mathrm{d}y\mathrm{d}x \le (b'-a')\|F\|_{\mathsf{H}^p,I}^p < \infty.$$

Therefore, given $\epsilon \in \mathbb{R}_{>0}$, there exists $M \in \mathbb{R}_{>0}$ sufficiently large that

$$\int_{a'}^{b'}\int_{|x|\ge M} |F(x+\mathrm{i}y)|^p \,\mathrm{d}y\mathrm{d}x < \pi\rho^2\epsilon^p.$$

If $z \in \mathbb{C}_{[a,b]}$ satisfies

$$\overline{\mathsf{D}}(\rho, z) \subseteq \{\zeta \in \mathbb{C}_{[a',b']} \mid \operatorname{Im}(\zeta) > M\},$$

then

$$|F(z)| \le \int_{a'}^{b'}\int_{|x|\ge M} |F(x+\mathrm{i}y)|^p \,\mathrm{d}y\mathrm{d}x < \epsilon,$$

as desired. ∎

7.4.5 Example ($\overline{\mathsf{H}}^p(\mathbb{C}_I;\mathbb{C}) \subset \mathsf{H}^p(\mathbb{C}_I;\mathbb{C})$ if I is not closed) If I is not closed, then it is easy to see that $\overline{\mathsf{H}}^p(\mathbb{C}_I;\mathbb{C})$ is strictly contained in $\mathsf{H}^p(\mathbb{C}_I;\mathbb{C})$. To see this, suppose that I has an open endpoint at x_0 and note that $F \in \mathsf{H}^p(\mathbb{C}_I;\mathbb{C})$ defined by $F_p(z) = (z - x_0)^{-2p}$ is in $\mathsf{H}^p(\mathbb{C}_I;\mathbb{C})$, $p \in [1,\infty)$ but not in $\overline{\mathsf{H}}^p(\mathbb{C}_I;\mathbb{C})$. For every $p \in [1,\infty)$, $F_p \in \mathsf{H}^\infty(\mathbb{C}_I;\mathbb{C})$ but $F_p \notin \overline{\mathsf{H}}^\infty(\mathbb{C}_I;\mathbb{C})$. •

We now show that the norms for Hardy spaces can be simplified to the computation of L^p-norms on the boundary.

7.4.6 Proposition (Norms for Hardy spaces) *For an interval $\mathrm{I} \subseteq \mathbb{R}$, $\mathrm{p} \in [1,\infty]$, and $\mathrm{F} \in \overline{\mathsf{H}}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}};\mathbb{C})$, let $\overline{\mathrm{F}} \in \mathsf{H}^{\mathrm{p}}(\mathbb{C}_{\mathrm{cl(I)}};\mathbb{C})$ be as given by Theorem 7.4.4. Then the following statements hold:*

(i) if $\mathrm{cl(I)} = [\mathrm{a},\mathrm{b}]$ for $\mathrm{a},\mathrm{b} \in \mathbb{R}$, $\mathrm{a} < \mathrm{b}$, then $\|\mathrm{F}\|_{\mathsf{H}^{\mathrm{p}},\mathrm{I}} = \max\{\|\overline{\mathrm{F}}_{\mathrm{a}}\|_{\mathrm{p}}, \|\overline{\mathrm{F}}_{\mathrm{b}}\|_{\mathrm{p}}\}$;

(ii) if $\sup \mathrm{I} = \infty$ and $\inf \mathrm{I} = \mathrm{a} \in \mathbb{R}$, then $\|\mathrm{F}\|_{\mathsf{H}^{\mathrm{p}},\mathrm{I}} = \|\overline{\mathrm{F}}_{\mathrm{a}}\|_{\mathrm{p}}$;

(iii) if $\inf \mathrm{I} = -\infty$ and $\sup \mathrm{I} = \mathrm{b} \in \mathbb{R}$, then $\|\mathrm{F}\|_{\mathsf{H}^{\mathrm{p}},\mathrm{I}} = \|\overline{\mathrm{F}}_{\mathrm{b}}\|_{\mathrm{p}}$;

(iv) if $\mathrm{I} = \mathbb{R}$ then $\overline{\mathsf{H}}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}};\mathbb{C})$ is comprised of constant functions that are necessarily zero if $\mathrm{p} \in [1,\infty)$.

Proof (i)

(ii) and (iii) These follow from part (ii) of the Theorem on page 123 of [Hoffmann 1962].

(iv) By Liouville's Theorem [Conway 1978, Theorem IV.3.4] we have that $\overline{\mathsf{H}}^p(\mathbb{C};\mathbb{C})$ consists of constant functions that, and so these must be zero if $p \in [1,\infty)$. ■

We conclude this discussion by proving that the Hardy spaces are Banach spaces.

7.4.7 Theorem (Hardy spaces are Banach spaces) *If* $I \subseteq \mathbb{R}$ *is an interval and if* $p \in [1,\infty]$*, then* $\overline{\mathsf{H}}^p(\mathbb{C}_I;\mathbb{C})$ *is a Banach space.*

Proof For intervals satisfying $\mathrm{cl}(I) = [a,\infty)$ or $\mathrm{cl}(I) = (-\infty,b]$, the result follows from classical results [e.g., Garnett 2007, Theorem II.1.3]. If $I = \mathbb{R}$, the result follows from Proposition 7.4.6(iv). If $I = \{x_0\}$ then the result is clear by convention. It remains to consider the case $\mathrm{cl}(I) = [a,b]$, where $a,b \in \mathbb{R}$ satisfy $a < b$. In this case let $(F_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\overline{\mathsf{H}}^p(\mathbb{C}_I;\mathbb{C})$. By Lemma 1 from the proof of Theorem 7.4.4, if $p \in [1,\infty)$ we have that $(F_j|K)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in the ∞-norm for any compact $K \subseteq \mathbb{C}_{\mathrm{int}(I)}$. Of course, this is also true when $p = \infty$, and so we conclude that (F_j) converges in the compact-open topology of $\mathsf{H}(\mathbb{C}_{\mathrm{int}(I)};\mathbb{C})$ to an holomorphic limit F [Rudin 1991, §1.45]. Next let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that $\|F_j - F_k\|_{\mathsf{H}^p,I} < \epsilon$ for $j,k \geq N$. For $y \in \mathrm{int}(I)$ and for $j \geq N$ we have

$$\|F_j - F\|_{\{y\},p} = \lim_{k\to\infty}\|F_j - F_k\|_{\{y\},p} \leq \lim_{k\to\infty}\|F_j - F_k\|_{\mathsf{H}^p,I} < \epsilon.$$

As this holds for every $y \in \mathrm{int}(I)$ we conclude that $\|F_j - F\|_{\mathsf{H}^p,I} < \epsilon$ for $j \geq N$, giving the desired convergence. ■

7.4.4 Locally convex topological vector spaces of holomorphic functions on vertical strips

In the preceding section we considered the classical Hardy spaces and showed that these are essentially the big Hardy spaces for closed intervals. In this section we use this structure to provide a locally convex topology for the big Hardy spaces.

We begin by making a simple observation concerning the relationship between big Hardy spaces defined on nested intervals.

7.4.8 Proposition (Hardy spaces defined on nested intervals) *Let* $p \in [1,\infty]$*. If intervals* $I, J \subseteq \mathbb{R}$ *satisfy* $J \supseteq I$*, then* $\mathsf{H}^p(\mathbb{C}_J;\mathbb{C}) \subseteq \mathsf{H}^p(\mathbb{C}_J;\mathbb{C})$*. Moreover, if* I *and* J *are closed, then* $\|F\|_{\mathsf{H}^p,J} \geq \|I\|_{\mathsf{H}^F,p}$ *for every* $F \in \mathsf{H}^p(\mathbb{C}_J;\mathbb{C})$ *and, in particular, the inclusion map from* $\mathsf{H}^p(\mathbb{C}_J;\mathbb{C})$ *into* $\mathsf{H}^p(\mathbb{C}_I;\mathbb{C})$ *is continuous for closed intervals* I *and* J*.* follows like LT case

Proof ■

For a fixed interval I we have a family of inclusions $i_{I,J}\colon \mathsf{H}^p(\mathbb{C}_I;\mathbb{C}) \to \mathsf{H}^p(\mathbb{C}_J;\mathbb{C})$, where J runs over compact intervals $J \subseteq I$. We then have the initial topology defined by the mappings $i_{I,J}$, $J \subseteq I$ be compact, which is the coarsest topology for which the mappings are continuous. This is a locally convex topology [Horváth 1966, §2.11]. Let us characterise this topology.

7.4.9 Theorem (The locally convex topology for big Hardy spaces) *Let* $\mathrm{I} \subseteq \mathbb{R}$ *be an interval and let* $\mathrm{p} \in [1, \infty]$*. Then* $\mathsf{H}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}}; \mathbb{C})$*, with the preceding locally convex topology, is a Fréchet space. Moreover:*

(i) if $(\mathrm{I}_{\mathrm{j}})_{\mathrm{j} \in \mathbb{Z}_{>0}}$ *is a sequence of compact subintervals of* I *such that (a)* $\mathrm{I}_{\mathrm{j}} \subseteq \mathrm{I}_{\mathrm{j}+1}$ *and (b)* $\mathrm{I} = \cup_{\mathrm{j} \in \mathbb{Z}_{>0}} \mathrm{I}_{\mathrm{j}}$*, then the locally convex topology for* $\mathsf{H}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}}; \mathbb{C})$ *agrees with the locally convex inverse limit* $(\mathsf{H}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}_{\mathrm{j}}}; \mathbb{C}))_{\mathrm{j} \in \mathbb{Z}_{>0}}$*, where the connecting map from* $\mathsf{H}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}_{\mathrm{j}+1}}; \mathbb{C})$ *to* $\mathsf{H}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}_{\mathrm{j}}}; \mathbb{C})$*,* $\mathrm{j} \in \mathbb{Z}_{>0}$*, is the inclusion map;*

(ii) if I *is closed, then the locally convex topology on* $\mathsf{H}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}}; \mathbb{C})$ *is the same as the norm topology induced of* $\overline{\mathsf{H}}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}}; \mathbb{C})$*;*

(iii) if I *is not closed, then the locally convex topology on* $\mathsf{H}^{\mathrm{p}}(\mathbb{C}_{\mathrm{I}}; \mathbb{C})$ *is not normable.*

the:Lpfrechet *Proof* (i) The argument here follows in the same manner as the proof of .

the:Lpfrechet (ii) The argument here follows in the same manner as the proof of .

(iii) Let $(I_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence of intervals as in part (i). Arguing as in the proof of

the:Lpfrechet , it suffices to show that each of the sets

$$\mathcal{U}(j, r) = \{F \in \mathsf{H}^p(\mathbb{C}_I; \mathbb{C}) \mid \|F\|_{\mathsf{H}^p, I} < r\}, \qquad j \in \mathbb{Z}_{>0},\ r \in \mathbb{R}_{>0},$$

is unbounded. We consider the five possible cases of not closed intervals.

First we take $I = (a, \infty)$ for $a \in \mathbb{R}$. We let $a_j = a + \frac{1}{j}$, $j \in \mathbb{Z}_{>0}$, and define $I_j = [a_j, a + j]$, $j \in \mathbb{Z}_{>0}$. Let $p \in [1, \infty)$, $j \in \mathbb{Z}_{>0}$, $r \in \mathbb{R}_{>0}$, and $M \in \mathbb{R}_{>0}$. Define

$$F_{k,b}(z) = \frac{b}{(z-a)^k}$$

for $k \in \mathbb{Z}_{>0}$ and $b \in \mathbb{R}_{>0}$. We compute

$$\|F_{k,b,x}\|_p^p = \int_{\mathbb{R}} \left| \frac{b}{((x-a) + \mathrm{i}y)^2} \right|^p \mathrm{d}y = \frac{\sqrt{\pi} b^p (x-a)^{1-kp} \Gamma(\frac{1}{2}(kp-1))}{\Gamma(\frac{kp}{2})}, \qquad x \in I.$$

From this we conclude that $F_{k,b} \in \mathsf{H}^p(\mathbb{C}_I; \mathbb{C})$ if $kp > 1$. For each $k \in \mathbb{Z}_{>0}$ for which $kp > 1$, we choose b_k such that

$$\int_{\mathbb{R}} |F_{k,b_k}(\tfrac{1}{2}((a + \tfrac{1}{j}) + (a + \tfrac{1}{j+1}) + x) + \mathrm{i}y)|^p \,\mathrm{d}y = 1.$$

Then, since $x \mapsto \|F_{k,b,x}\|_p$ is strictly monotonically decreasing in x we have

$$\|F_{k,b_k}\|_{I_j,p}^p < 1 < \|F_{k,b_k}\|_{I_{j+1},p}^p.$$

By taking k sufficiently large we can ensure that

$$\|F_{k,b_k}\|_{I_j,p} < r, \quad \|F_{k,b_k}\|_{I_{j+1},p} > M,$$

as desired. For $p = \infty$ a similar argument with the function $\frac{b}{z-a}$ gives the desired conclusion.

The same sort of argument with the same functions gives the result for the non-closed intervals of the form (a, b) and $(a, b]$. By a change of variable, we get the conclusion for intervals of the form $(-\infty, b)$ and $[a, b)$.

It remains to consider the case $I = \mathbb{R}$. Here we take $I_j = [-j, j]$, $j \in \mathbb{Z}_{>0}$. We define $F_{a,b} = b\mathrm{e}^{az^2}$, $a, b \in \mathbb{R}_{>0}$, and calculate

$$\|F_{a,b,x}\|_p^p = \int_{\mathbb{R}} |b\mathrm{e}^{a(x+\mathrm{i}y)^2}|^p \,\mathrm{d}y = \frac{\sqrt{\pi}b^b\mathrm{e}^{apx^2}}{\sqrt{ap}}.$$

For $a \in \mathbb{R}_{>0}$ let $b_a \in \mathbb{R}_{>0}$ be such that $|F_{a,b,\pm(j+\frac{1}{2})}|_p^p = 1$. Since $x \mapsto \|F_{a,b,x}\|_p$ is monotonically decreasing on $\mathbb{R}_{>0}$ and monotonically increasing on $\mathbb{R}_{<0}$, we have

$$\|F_{a,b}\|_{\mathsf{H}^p,I_j}^p < 1 < \|F_{a,b}\|_{\mathsf{H}^p,I_{j+1}}.$$

By taking a sufficiently small we can ensure that

$$\|F_{a,b}\|_{\mathsf{H}^p,I_j}^p < r, \quad \|F_{a,b}\|_{\mathsf{H}^p,I_{j+1}} > M,$$

which is the desired conclusion in this case for $p \in [1, \infty)$. For $p = \infty$ a similarly styled argument with the same function $F_{a,b}$ will give the same conclusion. ■

Section 7.5

Hardy spaces of holomorphic functions defined on annuli

We now adapt our constructions of Section 7.4 to holomorphic functions defined on annuli in the complex plane. As with our constructions with harmonic functions on annuli in Section 7.3, some of these constructions can be built upon those for vertical strips in Section 7.4.

Do I need to read this section? The definitions in this section will be used in Chapter IV-9 and Section V-7.2. So these can certainly be read when they are needed. However, it is not necessary to understand a lot of the technical material to understand the definitions. •

7.5.1 Definitions

By this point, we can get right to it.

7.5.1 Definition (Functions holomorphic on an annulus) For an interval $I \subseteq \mathbb{R}_{\geq 0}$, we denote by $\mathsf{H}(\mathbb{A}_I;\mathbb{C})$ the mappings $F\colon \mathbb{A}_I \to \mathbb{C}$ that are holomorphic on $\mathrm{int}(\mathbb{A}_I)$. If $\mathrm{int}(\mathbb{A}_I) = \emptyset$, we denote take the convention that $\mathsf{H}(\mathbb{A}_I;\mathbb{C}) = \mathbb{C}^{\mathbb{A}_I}$. •

Note that, for each interval $I \subseteq \mathbb{R}_{\geq 0}$, $\mathsf{H}(\mathbb{A}_I;\mathbb{C})$ is a $\mathbb{C}$-vector space with respect to the operations of pointwise addition and scalar multiplication. Note that if $\mathbb{A}_I$ is not open, then we place no restrictions on the values of functions in $\mathsf{H}(\mathbb{A}_I;\mathbb{C})$ when restricted to $\mathrm{bd}(\mathbb{A}_I)$. Indeed, as with our investigation of harmonic functions in Section 7.3, the matter of boundary values is a subject of considerable independent interest.

With the preceding definitions and discussion, we make the following definition.

7.5.2 Definition (Big holomorphic Hardy spaces on an annulus) Let $I \subseteq \mathbb{R}_{\geq 0}$ be an interval and let $p \in [1,\infty]$. For $F\colon \mathbb{A}_I \to \mathbb{C}$ and $r \in I$ denote $F_r\colon \mathbb{S} \to \mathbb{C}$ by $F_r(\mathrm{e}^{\mathrm{i}\theta}) = F(r\mathrm{e}^{\mathrm{i}\theta})$. Denote by $\mathsf{H}^p(\mathbb{C}_I;\mathbb{C})$ the mappings $F\colon \mathbb{A}_I \to \mathbb{C}$ such that

(i) $F \in \mathsf{H}(\mathbb{A}_I;\mathbb{C})$,

(ii) $F_r \in \mathsf{L}^p(\mathbb{S};\mathbb{C})$ for each $r \in I$, and

(iii) F has nontangential limits.

The space $\mathsf{H}^p(\mathbb{A}_I;\mathbb{C})$ is the ***big holomorphic Hardy space***. •

disk version

Corresponding to Remark 7.1.7–3, we *do not* have that the big holomorphic Hardy spaces are the same as the classical Hardy spaces, e.g., $\mathsf{H}^p(\mathbb{A}_{[0,1)};\mathbb{C})$ is not the same as the Hardy space of the unit disk as there is no requirement that the norms $\|F_r\|_p$ be uniformly bounded in r.

7.5.2 Poisson integral representations of holomorphic functions on annuli

7.5.3 Banach spaces of holomorphic functions on annuli

Now we study topological vector spaces of holomorphic functions whose domain is an annulus. We begin, in this section, by consider Banach spaces of such functions. For such spaces we shall see that there is a useful theory of nontangential limits, such as we presented in Section 7.3.3 for harmonic functions. As we shall see, we can do better for these nontangential limits in the holomorphic case than can be done in the harmonic case.

First we define the spaces we will investigate.

7.5.3 Definition (Holomorphic Hardy spaces on an annukus) Let $I \subseteq \mathbb{R}_{\geq 0}$ be an interval and let $p \in [1, \infty]$. For $F \in \mathsf{H}^p(\mathbb{A}_I; \mathbb{C})$ we denote

$$\|F\|_{\mathsf{H}^p,I} = \sup\{\|F_r\|_p \mid r \in I\}.$$

The subset $\{F \in \mathsf{H}(\mathbb{A}_I; \mathbb{C}) \mid \|F\|_{\mathsf{H}^p,I} < \infty\}$ is denoted by $\overline{\mathsf{H}}^p(\mathbb{A}_I; \mathbb{C})$ and we call these spaces the ***holomorphic Hardy spaces***. •

Let us prove the essential properties of the holomorphic Hardy spaces.

7.5.4 Theorem (Hardy spaces are defined on closed annuli) *Let* $I \subseteq \mathbb{R}_{\geq 0}$ *be an interval.*

(i) if $F \in \overline{\mathsf{H}}^p(\mathbb{A}_I; \mathbb{C})$ *for* $p \in [1, \infty]$*, then there exists* $\overline{F} \in \mathsf{H}^p(\mathbb{A}_{\mathrm{cl}(I)}; \mathbb{C})$ *such that*

(a) $F|\mathbb{C}_{\mathrm{int}(I)} = \overline{F}|\mathbb{C}_{\mathrm{int}(I)}$*,*

(b) $\overline{F}$ *has nontangential limits;*

(c) $\lim_{r\to r_0} \|\overline{F}_{r_0} - F_r\|_p = 0$ *for every* $r_0 \in \mathrm{cl}(I)$*.*

In particular, by part (i), the spaces $\overline{\mathsf{H}}^p(\mathbb{A}_I; \mathbb{C})$ *and* $\mathsf{H}^p(\mathbb{A}_{\mathrm{cl}(I)}; \mathbb{C})$ *are isometric.*

Proof ■

7.5.4 Locally convex topological vector spaces of holomorphic functions on annuli

7.5.5 Hardy spaces of vector-valued holomorphic functions defined on annuli

Bibliography

Axler, S., Bourdon, P., and Ramey, W. [2001] *Harmonic Function Theory*, 2nd edition, number 137 in Graduate Texts in Mathematics, Springer-Verlag: New York/-Heidelberg/Berlin, ISBN: 978-0-387-95218-5.

Bakan, A. and Kaijser, S. [2007] *Hardy spaces for the strip*, Journal of Mathematical Analysis and Applications, **333**(1), pages 347–364, ISSN: 0022-247X, DOI: 10.1016/j.jmaa.2006.10.088.

Banach, S. and Tarski, A. [1924] *Sur la décomposition des ensembles de points en parties respectivement congruentes*, Fundamenta Mathematicae, Polish Academy of Sciences. Institute of Mathematics, **6**, pages 244–277, ISSN: 0016-2736, DOI: 10.4064/fm-6-1-244-277.

Bauer, W. R. and Benner, R. H. [1971] *The non-existence of a Banach space of countably infinite Hamel dimension*, The American Mathematical Monthly, **78**(8), pages 895–896, ISSN: 0002-9890, DOI: 10.2307/2316492.

Börger, R. [1999] *A non-Jordan-measurable regularly open subset of the unit interval*, Archiv der Mathematik, **73**(4), page 816, ISSN: 0003-889X, DOI: 10.1007/s000130050396.

Buddenhagen, J. R. [1971] *Subsets of a countable set*, The American Mathematical Monthly, **78**(5), pages 536–537, ISSN: 0002-9890, DOI: 10.2307/2317767.

Cohn, D. L. [2013] *Measure Theory*, 2nd edition, Birkhäuser Advanced Texts, Birkhäuser: Boston/Basel/Stuttgart, ISBN: 978-1-4614-6955-1.

Conway, J. B. [1978] *Functions of One Complex Variable I*, 2nd edition, number 11 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-90328-6.

Davis, W. J., Dean, D. W., and Singer, I. [1968] *Complemented subspaces and Λ-systems in Banach spaces*, Israel Journal of Mathematics, **6**, pages 303–309, ISSN: 0021-2172.

Dekker, T. J. and de Groot, J. [1956] *Decompositions of a sphere*, Fundamenta Mathematicae, Polish Academy of Sciences. Institute of Mathematics, **43**(2), pages 185–194, ISSN: 0016-2736, DOI: 10.4064/fm-43-2-185-194.

Dvoretzky, A. [1960] *Some results on convex bodies and Banach spaces*, in *Proceedings of the International Symposium on Linear Spaces*, International Symposium on Linear Spaces, (Jerusalem, Israel, 1960), pages 123–160, Academic Press: New York, NY.

Dvoretzky, A. and Rogers, C. A. [1950] *Absolute and unconditional convergence in normed linear spaces*, Proceedings of the National Academy of Sciences of the United States of America, **36**(3), pages 192–197, ISSN: 1091-6490, URL: http://www.jstor.org/stable/88187 (visited on 07/23/2014).

Evans, J. W. and Tapia, R. A. [1970] *Hamel versus Schauder dimension*, The American Mathematical Monthly, **77**(4), pages 385–388, ISSN: 0002-9890, DOI: 10.2307/2316148.

Fischer, E. S. [1907] *Sur la convergence en moyenne*, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences. Séries A et B, **144**, page 49.

Fréchet, M. [1907] *Sur les ensembles de fonctions et les opérations linéaires*, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences. Séries A et B, **144**, pages 1414–1416.

Frink, Jr., O. [1933] *Jordan measure and Riemann integration*, Annals of Mathematics. Second Series, **34**(3), pages 518–526, ISSN: 0003-486X, DOI: 10.2307/1968175.

Garnett, J. B. [2007] *Bounded Analytic Functions*, number 236 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-33621-3.

Gudder, S. [1974] *Inner product spaces*, The American Mathematical Monthly, **81**(1), pages 29–36, ISSN: 0002-9890, DOI: 10.2307/2318908.

Halmos, P. R. [1950] *Measure Theory*, Litton Education Publishing, Inc.: New York, NY, New edition: [Halmos 1974].

— [1974] *Measure Theory*, number 18 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-90088-9, Original: [Halmos 1950].

Hamming, R. W. [1980] *The unreasonable effectiveness of mathematics*, The American Mathematical Monthly, **87**(2), pages 81–90, ISSN: 0002-9890, DOI: 10.2307/2321982.

Hardy, G. H. [1949] *Divergent Series*, Cambridge University Press: New York/Port Chester/Melbourne/Sydney, Reprint: [Hardy 1992].

— [1992] *Divergent Series*, AMS Chelsea Publishing: Providence, RI, ISBN: 978-0-8218-2649-2, Original: [Hardy 1949].

Hoffmann, K. [1962] *Banach Spaces of Analytic Functions*, Prentice-Hall Series in Mathematical Analysis, Prentice-Hall: Englewood Cliffs, NJ, Reprint: [Hoffmann 2007].

— [2007] *Banach Spaces of Analytic Functions*, Dover Publications, Inc.: New York, NY, ISBN: 978-0-486-45874-8, Original: [Hoffmann 1962].

Horváth, J. [1966] *Topological Vector Spaces and Distributions*, volume 1, Addison Wesley: Reading, MA, Reprint: [Horváth 2012].

— [2012] *Topological Vector Spaces and Distributions*, Dover Publications, Inc.: New York, NY, ISBN: 978-0-486-48850-9, Original: [Horváth 1966].

John, F. [1948] *Extremum problems with inequalities as subsidiary conditions*, in *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, Interscience Publishers: New York, NY.

Joichi, J. T. [1966] *Normed linear spaces equivalent to inner product spaces*, Proceedings of the American Mathematical Society, **17**(2), pages 423–426, ISSN: 0002-9939, DOI: 10.2307/2035180.

Kailath, T. [1980] *Linear Systems*, Information and System Sciences Series, Prentice-Hall: Englewood Cliffs, NJ, ISBN: 978-0-13-536961-6.

Lacey, H. E. [1973] *The Hamel dimension of any infinite dimensional separable Banach space is c*, The American Mathematical Monthly, **80**, page 298, ISSN: 0002-9890, DOI: 10.2307/2318458.

Ledoux, M. [2001] *The Concentration of Measure Phenomenon*, number 89 in American Mathematical Society Mathematical Surveys and Monographs, American Mathematical Society: Providence, RI, ISBN: 978-0-8218-3792-4.

Lindenstrauss, J. and Tzafriri, L. [1971] *On the complemented subspaces problem*, Israel Journal of Mathematics, **9**, pages 263–269, ISSN: 0021-2172.

Mennicken, R. and Sagraloff, B. [1979] *On Banach's closed range theorem*, Archiv der Mathematik, **33**(1), pages 461–465, ISSN: 0003-889X, DOI: 10.1007/BF01222785.

Milman, V. D. [1971] *A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies*, Rossiĭskaya Akademiya Nauk. Funktsional'nyĭ Analiz i ego Prilozheniya, **5**(4), pages 28–37, ISSN: 0374-1990.

Munkres, J. R. [1984] *Elements of Algebraic Topology*, Addison Wesley: Reading, MA, ISBN: 978-0-201-04586-4.

Pisier, G. [1986] *Probabilistic methods in the geometry of Banach spaces*, in *Probability and Analysis*, First Session of the Centro Internazionale Matematico Estivo, (Varenna, Italy, June 1985), edited by G. Letta and M. Pratelli, 1206 Lecture Notes in Mathematics, pages 167–241, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-540-16787-7.

Riesz, F. [1907a] *Sur les systèmes orthogonaux de fontions*, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences. Séries A et B, **144**, pages 615–619.

— [1907b] *Sur les systèmes orthogonaux de fontions et l'équation de Fredholm*, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences. Séries A et B, **144**, pages 734–736.

— [1907c] *Sur une espèce de géométrie analytiques des systèmes de fonctions sommables*, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences. Séries A et B, **144**, pages 1409–1411.

— [1909] *Sur les opérations fonctionelles linéaires*, Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences. Séries A et B, **149**, pages 974–977.

Rudin, W. [1986] *Real and Complex Analysis*, 3rd edition, McGraw-Hill Series in Higher Mathematics, McGraw-Hill: New York, NY, ISBN: 978-0-07-054234-1.

— [1991] *Functional Analysis*, 2nd edition, International Series in Pure & Applied Mathematics, McGraw-Hill: New York, NY, ISBN: 978-0-07-054236-5.

Schep, A. R. [2003] *And still one more proof of the Radon–Nikodym theorem*, The American Mathematical Monthly, **110**(6), pages 536–538, ISSN: 0002-9890, DOI: 10.2307/3647910.

Stromberg, K. [1979] *The Banach–Tarski paradox*, The American Mathematical Monthly, **86**(3), pages 151–161, ISSN: 0002-9890, DOI: 10.2307/2321514.

Varberg, D. E. [1971] *Change of variables in multiple integrals*, The American Mathematical Monthly, **78**(1), pages 42–45, ISSN: 0002-9890, DOI: 10.2307/2317484.

Wigner, E. P. [1960] *The unreasonable effectiveness of mathematics in the natural sciences*, Communications on Pure and Applied Mathematics, **13**(1), pages 1–14, ISSN: 0010-3640, DOI: 10.1002/cpa.3160130102.

Wilansky, A. [1989] *On the proof of the Radon–Nikodym theorem*, The American Mathematical Monthly, **96**(5), page 441, ISSN: 0002-9890, DOI: 10.2307/2325154.