A Mathematical Introduction to Signals and Systems

Volume II. Intermediate real analysis and complex 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 2

In this volume, we cover topics that one might lump together as multivariable calculus, if one regards complex analysis as a form of two-variable real calculus (something one really should not do).

We begin in Chapter 1 by considering the extension to multiple dimensions of the topics considered for a single variable in Chapter I-3. The topics here are what one might encounter in a second year of studies, after a first year of introductory calculus. The coverage is given in a great deal of detail, as the subject of the chapter is an important one, and a careful systematic treatment of differential calculus in multiple variables is an important one in applications, and a good understanding of it provides an important bridge between the elementary topics of introductory calculus to the more difficult topics in function analysis covered in Chapters III-3, III-4, and III-6. In this chapter we also delve into some advanced specialised topics, both foundational and involving applications of multivariable analysis. The applications, we hope, provides some context for some of the mathematics we have been developing.

The next topic, vector calculus, is also one that is a part of the undergraduate education for mathematics students and well-educated students in engineering and the physical sciences. We have chosen to cover this material in a more modern way than is normally done. Thus we present the tools to develop a general version of "Stokes' Theorem," and then specialise this to the standard theorems of vector calculus. As with the presentation of the other "background" material in previous chapters, we do not emphasise the (very important) computational matters that arise in vector calculus, concentrating instead on establishing structure and the basic theorems that can be used subsequently.

The final Chapter 3 in this volume deals with the important subject of complex analysis. This subject will come up in various places in subsequent volumes, sometimes in ways that appear strange initially. Indeed, for students familiar with real variable calculus, complex analysis looks like a little strange. It has many similarities to the real variable case, but also many important differences. Students in mathematics will have had a reasonably good course in complex analysis that will very often be quite similar to what we cover in Chapter 3. Students in engineering and the physical sciences will very often have some background with complex numbers, and some sort of background in complex analysis. Such students will need to take some time to learn the fundamental parts of complex analysis that we present in this chapter.

Andrew D. Lewis Kingston, ON, Canada

This version: 2022/03/07

Table of Contents

Chapter 1

Multiple real variables and functions of multiple real variables

In this chapter we carry on from the preceding chapter and develop the notions of continuity, differentiability, and integrability for functions with multivariable domains and codomains. Much of this development goes in a manner that is strikingly similar to the single-variable case. Therefore, we do not spend as much time with illustrative examples and motivating discussion as we did in Chapter I-3. Also some proofs are very similar to their single-variable counterparts, and in these cases we omit detailed proofs. There are, however, some significant differences in the presentation that arise in the extension to multiple variables. For example, the Inverse Function Theorem and the change of variables formula for integrals are far more complicated in the multivariable case. Also, for the multivariable case, one has the important Fubini's Theorem for integrals. Therefore, it is not the case that everything here is simply a trivial extension of what we have already seen in Chapter I-3. But it is the case that understanding the material in Chapter I-3 will make this chapter far easier to get through.

Do I need to read this chapter? As with the material in Chapter I-3, readers who have had a decent sequence of analysis courses can probably skim this chapter on a first reading. This is particularly true if the material in Chapter I-3 has been satisfactorily digested. However, there will be occasions where we will use the results in this chapter, so it will have to be come back to at some point if it is not sufficiently well understood. •

Contents

Section 1.1

Norms on Euclidean space and related spaces

In this section we introduce the very most basic structure of Euclidean space: its algebraic structure along with the structure of a norm. Combined, this structure allows us to do analysis in n-dimensional Euclidean space, just as we did in Chapters I-2 and I-3 for $\mathbb{R}$.

Do I need to read this section? The results in Sections 1.1.1 and 1.1.2 are fundamental to everything in this chapter, and so are required reading. The material in the remaining sections on norms for linear and multilinear maps is required when we define the derivative and higher-order derivatives in Section 1.4. •

1.1.1 The algebraic structure of $\mathbb{R}^n$

We denote by $\mathbb{R}^n$ the n-fold Cartesian product of $\mathbb{R}$ with itself:

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \cdots \times \mathbb{R}}_{n \text{ copies}}.$$

We shall often refer to $\mathbb{R}^n$ as ***n-dimensional Euclidean space***. We shall denote a typical element of $\mathbb{R}^n$ by $\boldsymbol{v} = (v_1, \dots, v_n)$ when we are talking about the algebraic structure. We call the numbers $v_1, \dots,_n$ the ***components*** of $\boldsymbol{v}$. We may also use the letters $\boldsymbol{u}$ and $\boldsymbol{w}$. Later in this section, when we discuss properties of $\mathbb{R}^n$ that are not algebraic, we will denote typical points by $\boldsymbol{x} = (x_1, \dots, x_n)$, and we may also use letters like $\boldsymbol{y}$. Generally speaking, we shall attempt to distinguish between the algebraic and nonalgebraic parts of the structure of $\mathbb{R}^n$.

In $\mathbb{R}$, as we indicated in Section I-2.2.1, we can perform familiar algebraic operations like addition, multiplication, and division. Not all of these operations generally carry over to $\mathbb{R}^n$. One *can* add elements of $\mathbb{R}^n$ using the rule

$$\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1, \dots, u_n + v_n). \tag{1.1}$$

One can also multiply elements of $\mathbb{R}^n$ by an element of $\mathbb{R}$ using the rule

$$a\boldsymbol{v} = (av_1, \dots, av_n). \tag{1.2}$$

Let us summarise some of the properties of the algebraic structure of $\mathbb{R}^n$. The following result states that addition (1.1) and multiplication by scalars (1.2) satisfy the axioms for a $\mathbb{R}$-vector space.

1.1.1 Proposition ($\mathbb{R}^n$ is a $\mathbb{R}$-vector space) *The operations (1.1) and (1.2) have the following properties:*

(i) $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$, $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ (***commutativity***);

(ii) $\mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3) = (\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3$, $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in \mathbb{R}^n$ (***associativity***);

(iii) *the element* $\mathbf{0} = (0, \dots, 0) \in \mathbb{R}^n$ *has the property that* $\mathbf{v} + \mathbf{0} = \mathbf{v}$ *for every* $\mathbf{v} \in \mathbb{R}^n$ (***zero vector***);

(iv) *for every* $\mathbf{v} = (v_1, \dots, v_n) \in \mathbb{R}^n$ *the element* $-\mathbf{v} = (-v_1, \dots, -v_n) \in \mathbb{R}^n$ *has the property that* $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ (***negative vector***);

(v) $a(b\mathbf{v}) = (ab)\mathbf{v}$, $a, b \in \mathbb{R}$, $\mathbf{v} \in \mathbb{R}^n$ (***associativity*** *again*);

(vi) $1\mathbf{v} = \mathbf{v}$, $\mathbf{v} \in \mathbb{R}^n$;

(vii) $a(\mathbf{v}_1 + \mathbf{v}_2) = a\mathbf{v}_1 + a\mathbf{v}_2$, $a \in \mathbb{R}$, $\mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^n$ (***distributivity***);

(viii) $(a_1 + a_2)\mathbf{v} = a_1\mathbf{v} + a_2\mathbf{v}$, $a_1, a_2 \in \mathbb{R}$, $\mathbf{v} \in \mathbb{R}^n$ (***distributivity*** *again*).

Proof These statements all follow from the properties of algebraic operations on real numbers. ■

look for places where this is defined *ad hoc*

Let us introduce some useful notation for subsets of $\mathbb{R}^n$.

1.1.2 Definition (Dilation, sum, and difference of sets) Let $A, B \subseteq \mathbb{R}^n$ and let $\lambda \in \mathbb{R}$.

(i) The ***dilation*** of A by λ is the set

$$\lambda A = \{\lambda \boldsymbol{x} \mid \boldsymbol{x} \in A\}.$$

(ii) The ***sum*** of A and B is the set

$$A + B = \{\boldsymbol{x} + \boldsymbol{y} \mid \boldsymbol{x} \in A,\ \boldsymbol{y} \in B\}.$$

(iii) The ***difference*** of A and B is the set

$$A - B = \{\boldsymbol{x} - \boldsymbol{y} \mid \boldsymbol{x} \in A,\ \boldsymbol{y} \in B\}.$$

(iv) If $A = \{\boldsymbol{x}_0\}$ is a singleton, then we denote $A + B = \boldsymbol{x}_0 + B$ and $A - B = \boldsymbol{x}_0 - B$. •

Not all of the algebraic structure of $\mathbb{R}$ carries over to $\mathbb{R}^n$.

1. Generally, one cannot multiply or divide elements of $\mathbb{R}^n$ together in a useful way. However, for $n = 2$ it turns out that multiplication and division *are* also possible, and this is described in Section I-4.7.[1]
2. Although Zermelo's Well Ordering Theorem tells us that $\mathbb{R}^n$ possesses a well order, apart from $n = 1$ there is no useful (i.e., reacting well with the other structures of $\mathbb{R}^n$) partial order on $\mathbb{R}^n$. Thus any of the results about $\mathbb{R}$ that relate to its natural total order $\leq$ will not generally carry over to $\mathbb{R}^n$.

Let us review some other algebraic concepts and notation associated with $\mathbb{R}^n$. We refer to the general discussions in Sections I-4.5, I-5.1, and I-5.4 for more detailed and general discussions.

[1]There are other values of n for which multiplication and division are possible, but this will not interest us here.

1. The ***standard basis*** for $\mathbb{R}^n$ is the collection $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ of elements of $\mathbb{R}^n$ given by

$$\boldsymbol{e}_j = (0, \dots, 1, \dots, 0),$$

where the 1 is in the jth position. Obviously we have

$$(v_1, \dots, v_n) = v_1\boldsymbol{e}_1 + \dots + v_n\boldsymbol{e}_n.$$

2. The set of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ is denoted by $\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n; \mathbb{R}^m)$ and the set of $m \times n$ matrices with real entries is denoted by $\mathrm{Mat}_{m\times n}(\mathbb{R})$. The sets $\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n; \mathbb{R}^m)$ and $\mathrm{Mat}_{m\times n}(\mathbb{R})$ are $\mathbb{R}$-vector spaces and, moreover, are isomorphic in a natural way. Indeed, if $\boldsymbol{A} \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ the corresponding linear map is

$$\boldsymbol{v} \mapsto \Big(\sum_{j=1}^{n} \boldsymbol{A}(1, j)v_j, \dots, \sum_{j=1}^{n} \boldsymbol{A}(m, j)v_j\Big).$$

1.1.2 The Euclidean inner product and norm, and other norms

There is a generalisation to $\mathbb{R}^n$ of the absolute value function on $\mathbb{R}$. Indeed, this is one of the more valuable features of $\mathbb{R}^n$. In fact, there are many generalisations of the absolute value function which go under the name of "norms;" we shall discuss this idea in detail in Chapter III-3. For now let us just define the norm that is of interest to us. It turns out that the norm we use most in this section is a special sort of norm, derived from an inner product.

1.1.3 Definition (Euclidean inner product) The ***Euclidean inner product*** on $\mathbb{R}^n$ is the map $\langle\cdot,\cdot\rangle_{\mathbb{R}^n}$ from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}$ defined by

$$\langle \boldsymbol{x}, \boldsymbol{y}\rangle_{\mathbb{R}^n} = \sum_{j=1}^{n} x_j y_j. \quad \bullet$$

This is sometimes called the "dot product" and instead the notation $\boldsymbol{x}\cdot\boldsymbol{y}$ is used. We shall absolutely never use this notation; it is something to be used only by small children.

Let us give some properties of the Euclidean inner product.

1.1.4 Proposition (Properties of the Euclidean inner product) *The Euclidean inner product has the following properties:*

(i) $\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbb{R}^n} = \langle \mathbf{y}, \mathbf{x}\rangle_{\mathbb{R}^n}$ *for* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *(**symmetry**);*

(ii) $\langle \alpha\mathbf{x}, \mathbf{y}\rangle_{\mathbb{R}^n} = \alpha\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbb{R}^n}$ *for* $\alpha \in \mathbb{R}$ *and* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *(**linearity I**);*

(iii) $\langle \mathbf{x}_1 + \mathbf{x}_2, \mathbf{y}\rangle_{\mathbb{R}^n} = \langle \mathbf{x}_1, \mathbf{y}\rangle_{\mathbb{R}^n} + \langle \mathbf{x}_2, \mathbf{y}\rangle_{\mathbb{R}^n}$ *for* $\mathbf{x}_1, \mathbf{x}_2, \mathbf{y} \in \mathbb{R}^n$ *(**linearity II**);*

(iv) $\|\mathbf{x}\|_{\mathbb{R}^n}\mathbf{x} \geq 0$ *for* $\mathbf{x} \in \mathbb{R}^n$ *(**positivity**);*

(v) $\|\mathbf{x}\|_{\mathbb{R}^n}\mathbf{x} = 0$ *only if* $\mathbf{x} = \mathbf{0}$ *(**definiteness**).*

Proof These are all elementary deductions using the definition. ■

As we shall see in Definition III-4.1.1, a map assigning to a pair of vectors in any $\mathbb{R}$-vector space a number, with the assignment having the five properties above, is called an "inner product." These are studied in some generality in Chapter III-4.

Readers knowing a little Euclidean geometry are familiar with the notion of vectors being "perpendicular." For grownups, the word is "orthogonal."

1.1.5 Definition (Orthogonal, orthogonal complement) Two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ are ***orthogonal*** if $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathbb{R}^n} = 0$. If $S \subseteq \mathbb{R}^n$, the ***orthogonal complement*** of S is the set

$$S^{\perp} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathbb{R}^n} = 0 \text{ for all } \boldsymbol{y} \in S\}. \qquad \bullet$$

Let us explore the notion of orthogonality with some examples.

1.1.6 Examples (Orthogonality)

1. Consider two vectors $\boldsymbol{x} = (x_1, x_2), \boldsymbol{y} = (y_1, y_2) \in \mathbb{R}^2$. These vectors are orthogonal if and only if $x_1 y_1 + x_2 y_2 = 0$. Thinking of one of the vectors, say $\boldsymbol{x}$, as being fixed, this is a linear equation in $\boldsymbol{y}$; we refer to Section I-5.1.8 for a general discussion of such maps. Here we need only note that the subspace of solutions is two-dimensional when $\boldsymbol{x} = \boldsymbol{0}$ and is one-dimensional otherwise. Thus, obviously, every vector is orthogonal to $\boldsymbol{0}$. To describe the one-dimensional subspace of vectors orthogonal to $\boldsymbol{x} \neq \boldsymbol{0}$ we note that one such vector is $\boldsymbol{y} = (-x_2, x_1)$. Thus this is a basis for one-dimensional subspace of vectors orthogonal to $\boldsymbol{x}$. We show the picture in Figure 1.1, noting that, in this case, orthogonality agrees

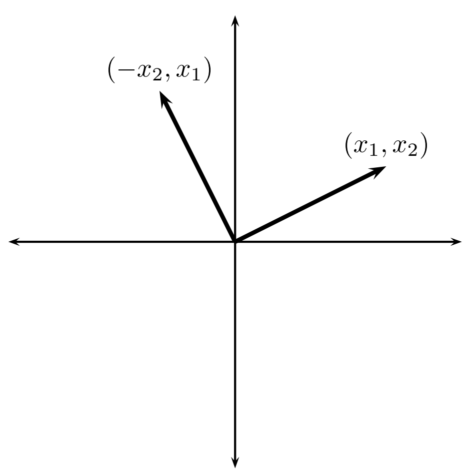

Figure 1.1 Orthogonal vectors in $\mathbb{R}^2$

with our usual notion of perpendicularity.

2. Let $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ be the standard basis for $\mathbb{R}^n$. Then one readily determines that

$$\langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle_{\mathbb{R}^n} = \begin{cases} 1, & j = k, \\ 0, & j \neq k. \end{cases}$$

A general basis for $\mathbb{R}^n$ with this property is called ***orthonormal***. Such ideas will be explored in great depth and generality in Chapter III-4. •

We shall not explore the details of what an inner product buys for us, referring the reader to for a general discussion of finite-dimensional vector spaces with inner products. For our purposes the Euclidean inner product is related to the Euclidean norm which is the generalisation of the absolute value function on $\mathbb{R}$ that we shall use to prescribe the structure of Euclidean space. fin-dim hilbert

1.1.7 Definition (Euclidean norm) The ***Euclidean norm*** on $\mathbb{R}^n$ is the function $\|\cdot\|_{\mathbb{R}^n}$ from $\mathbb{R}^n$ to $\mathbb{R}_{\geq 0}$ defined by

$$\|x\|_{\mathbb{R}^n} = \Big(\sum_{j=1}^{n} x_j^2\Big)^{1/2}. \qquad \bullet$$

Note that when $n = 1$ we have $\|\cdot\|_{\mathbb{R}^1} = |\cdot|$. When $n \in \{2, 3\}$, $\|x\|_{\mathbb{R}^n}$ is the usual notion of length in "physical space."

Let us record the properties of the Euclidean norm.

1.1.8 Proposition (Properties of the Euclidean norm) *The Euclidean norm has the following properties:*

(i) $\|\alpha\mathbf{x}\|_{\mathbb{R}^n} = |\alpha|\|\mathbf{x}\|_{\mathbb{R}^n}$ *for* $\alpha \in \mathbb{R}$ *and* $\mathbf{x} \in \mathbb{R}^n$ *(**homogeneity**);*

(ii) $\|\mathbf{x}\|_{\mathbb{R}^n} \geq 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$ *(**positivity**);*

(iii) $\|\mathbf{x}\|_{\mathbb{R}^n} = 0$ *only if* $\mathbf{x} = \mathbf{0}$ *(**definiteness**);*

(iv) $\|\mathbf{x}_1 + \mathbf{x}_2\|_{\mathbb{R}^n} \leq \|\mathbf{x}_1\|_{\mathbb{R}^n} + \|\mathbf{x}_2\|_{\mathbb{R}^n}$ *(**triangle inequality**).*

Moreover, the Euclidean norm shares the following relationships with the Euclidean inner product:

(v) $\|\mathbf{x}\|_{\mathbb{R}^n} = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle_{\mathbb{R}^n}}$ *for all* $\mathbf{x} \in \mathbb{R}^n$*;*

(vi) $|\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbb{R}^n}| \leq \|\mathbf{x}\|_{\mathbb{R}^n}\|\mathbf{y}\|_{\mathbb{R}^n}$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ *(**Cauchy–Bunyakovsky– Schwarz inequality**).*

Proof The only nontrivial properties are the fourth one and the final one. We first prove the Cauchy–Bunyakovsky–Schwarz inequality and then use it to prove the triangle inequality.

The Cauchy–Bunyakovsky–Schwarz inequality is obviously true for $y = \mathbf{0}$, so we shall suppose that $y \neq \mathbf{0}$. We first prove the result for $\|y\|_{\mathbb{R}^n} = 1$. In this case we have

$$\begin{aligned}
0 &\leq \|x - \langle x, y\rangle_{\mathbb{R}^n} y\|_{\mathbb{R}^n}^2 \\
&= \langle x - \langle x, y\rangle_{\mathbb{R}^n} y, x - \langle x, y\rangle_{\mathbb{R}^n} y\rangle_{\mathbb{R}^n} \\
&= \langle x, x\rangle_{\mathbb{R}^n} - \langle x, y\rangle_{\mathbb{R}^n}\langle y, x\rangle_{\mathbb{R}^n} - \langle x, y\rangle_{\mathbb{R}^n}\langle x, y\rangle_{\mathbb{R}^n} + \langle x, y\rangle_{\mathbb{R}^n}\langle x, y\rangle_{\mathbb{R}^n}\langle y, y\rangle_{\mathbb{R}^n} \\
&= \|x\|_{\mathbb{R}^n}^2 - \langle x, y\rangle_{\mathbb{R}^n}^2.
\end{aligned}$$

Thus we have shown that provided $\|y\|_{\mathbb{R}^n} = 1$, $\langle x, y\rangle_{\mathbb{R}^n}^2 \leq \|x\|_{\mathbb{R}^n}^2$. Taking square roots yields the result in this case. For $\|y\|_{\mathbb{R}^n} \neq 1$ we define $z = \frac{y}{\|y\|_{\mathbb{R}^n}}$ so that $\|z\|_{\mathbb{R}^n} = 1$. In this case

$$\left|\langle x, z\rangle_{\mathbb{R}^n}\right| \le \|x\|_{\mathbb{R}^n} \quad \Longrightarrow \quad \frac{\left|\langle x, y\rangle_{\mathbb{R}^n}\right|}{\|y\|_{\mathbb{R}^n}} \le \|x\|_{\mathbb{R}^n},$$

and so the inequality follows.

Now, to prove the triangle inequality, we compute

$$\begin{aligned}\|x+y\|_{\mathbb{R}^n}^2 &= \langle x+y, x+y\rangle_{\mathbb{R}^n}\\ &= \|x\|_{\mathbb{R}^n}^2 + 2\langle x, y\rangle_{\mathbb{R}^n} + \|y\|_{\mathbb{R}^n}^2\\ &\le \|x\|_{\mathbb{R}^n}^2 + 2\left|\langle x, y\rangle_{\mathbb{R}^n}\right| + \|y\|_{\mathbb{R}^n}^2\\ &\le \|x\|_{\mathbb{R}^n}^2 + 2\|x\|_{\mathbb{R}^n}\|y\|_{\mathbb{R}^n} + \|y\|_{\mathbb{R}^n}^2\\ &= (\|x\|_{\mathbb{R}^n} + \|y\|_{\mathbb{R}^n})^2,\end{aligned}$$

where we have used the lemma. The result now follows by taking square roots. ■

As we shall see in Definition III-3.1.2, a map assigning to vectors in a $\mathbb{R}$-vector space a number, with the assignment having the three properties above, is a “norm.” These are studied in detail in Chapter III-3.

Sometimes we will use other norms for $\mathbb{R}^n$. Two common norms are given in the following definition.

1.1.9 Definition (1- and ∞-norm for Euclidean space) The ***1-norm*** on $\mathbb{R}^n$ is the function $\|\cdot\|_1$ from $\mathbb{R}^n$ to $\mathbb{R}_{\ge 0}$ defined by

$$\|x\|_1 = \sum_{j=1}^{n} |x_j|,$$

and the ***∞-norm*** on $\mathbb{R}^n$ is the function $\|\cdot\|_\infty$ from $\mathbb{R}^n$ to $\mathbb{R}_{\ge 0}$ defined by

$$\|x\|_\infty = \max\{|x_1|, \dots, |x_n|\}.$$ •

The 1- and ∞-norms enjoy the following properties, as is easily verified (see also Examples III-3.1.3–3 and III-4 and Section III-3.8.1).

1.1.10 Proposition (Properties of the 1- and ∞-norms) *For* $p \in \{1, \infty\}$, *the* p*-norm has the following properties:*

(i) $\|\alpha\mathbf{x}\|_p = |\alpha|\|\mathbf{x}\|_p$ *for* $\alpha \in \mathbb{R}$ *and* $\mathbf{x} \in \mathbb{R}^n$ *(**homogeneity**);*

(ii) $\|\mathbf{x}\|_p \ge 0$ *for all* $\mathbf{x} \in \mathbb{R}^n$ *(**positivity**);*

(iii) $\|\mathbf{x}\|_p = 0$ *only if* $\mathbf{x} = \mathbf{0}$ *(**definiteness**);*

(iv) $\|\mathbf{x}_1 + \mathbf{x}_2\|_p \le \|\mathbf{x}_1\|_p + \|\mathbf{x}_2\|_p$ *(**triangle inequality**).*

When we are simultaneously discussing and contrasting the various norms, we will sometime use $\|\cdot\|_2$ rather than $\|\cdot\|_{\mathbb{R}^n}$ to denote the Euclidean norm, and we may refer to this norm as the ***2-norm***.

The following relationships between the 1-, 2-, and ∞-norms are often useful.

1.1.11 Proposition (Relationships between the 1-, 2-, and ∞-norms) *For $\mathbf{v} \in \mathbb{R}^n$ we have the following inequalities:*

(i) $\|\mathbf{v}\|_1 \le \sqrt{n}\|\mathbf{v}\|_2$;
(ii) $\|\mathbf{v}\|_1 \le n\|\mathbf{v}\|_\infty$;
(iii) $\|\mathbf{v}\|_2 \le \|\mathbf{v}\|_1$;
(iv) $\|\mathbf{v}\|_2 \le \sqrt{n}\|\mathbf{v}\|_\infty$;
(v) $\|\mathbf{v}\|_\infty \le \|\mathbf{v}\|_1$;
(vi) $\|\mathbf{v}\|_\infty \le \|\mathbf{v}\|_2$.

Moreover, the above inequalities are the best possible in the sense that, in each case, there exists a vector $\mathbf{v} \in \mathbb{R}^n$ such that equality is satisfied.

Proof (i) Note that the expression

$$\|v\|_1 = \sum_{j=1}^n |v_j|$$

means that $n\|v\|_1$ is the average of the positive numbers $|v_1|, \dots, |v_n|$. Thus we can write each of these numbers as this average divided by n plus the difference: $|v_j| = \frac{\|v\|_1}{n} + \delta_j$. Note that $\sum_{j=1}^n \delta_j = 0$. Now compute

$$\begin{aligned}\|v\|_2 &= \Big(\sum_{j=1}^n |v_j|^2\Big)^{1/2} = \Big(\sum_{j=1}^n \Big(\frac{\|v\|_1}{n} + \delta_j\Big)^2\Big)^{1/2} \\ &= \Big(\sum_{j=1}^n \Big(\frac{\|v\|_1^2}{n^2} + 2\frac{\|v\|_1\delta_j}{n} + \delta_j^2\Big)\Big)^{1/2} \ge \Big(\sum_{j=1}^n \frac{\|v\|_1^2}{n^2}\Big)^{1/2} = \frac{\|v\|_1}{\sqrt{n}},\end{aligned}$$

as desired, using the fact that $\sum_{j=1}^n \delta_j = 0$. The inequality is an equality by taking, for example, $v = (1, \dots, 1)$.

(ii) We have

$$\|v\|_1 = \sum_{j=1}^n |v_j| \le \sum_{j=1}^n \max\{|v_j| \mid j \in \{1, \dots, n\}\} = n\|v\|_\infty.$$

The inequality becomes equality, for example, for the vector $(1, \dots, 1)$.

(iii) We have

$$\|v\|_2 = \Big\|\sum_{j=1}^n v_j e_j\Big\|_2 \le \sum_{j=1}^n \|v_j e_j\|_2 = \sum_{j=1}^n |v_j| \|e_j\|_2 = \sum_{j=1}^n |v_j| = \|v\|_1.$$

The inequality becomes equality if, for example, $v = (1, 0, \dots, 0)$.

(iv) First note that the inequality is trivially satisfied when $v = \mathbf{0}_{\mathbb{F}^n}$. If $\|v\|_\infty = 1$ we have $|v_j| \le 1$ whence $|v_j|^2 \le |v_j|$ for $j \in \{1, \dots, n\}$. Therefore, in this case we have

$$\|v\|_2^2 = \sum_{j=1}^n \|v_j\|^2 \le \sum_{j=1}^n |v_j| \le \sum_{j=1}^n \max\{|v_j| \mid j \in \{1, \dots, n\}\} = n\|v\|_\infty.$$

Therefore, taking square roots, when $\|v\|_\infty = 1$ we have $\|v\|_2 \leq \sqrt{n}\|v\|_\infty$. For general nonzero v we write $v = \lambda u$ where $\|u\|_\infty = 1$ and where $\lambda = \|v\|_\infty$. We then have

$$\|v\|_2 = |\lambda|\|u\|_2 \leq \lambda \sqrt{n}\|u\|_\infty = \sqrt{n}\|v\|_\infty,$$

giving the desired result. The inequality becomes equality by taking, for example, $v = (1, \dots, 1)$.

(v) Let $j_0 \in \{1, \dots, n\}$ be such that

$$|v_{j_0}| = \max\{|v_j| \mid j \in \{1, \dots, n\}\}.$$

Then

$$\|v\|_\infty = |v_{j_0}| \leq \sum_{j=1} |v_j|.$$

The inequality becomes equality, for example, for the vector $(1, 0, \dots, 0)$.

(vi) Let $j_0 \in \{1, \dots, n\}$ be such that

$$|v_{j_0}| = \max\{|v_j| \mid j \in \{1, \dots, n\}\}.$$

Then

$$\|v\|_\infty^2 = |v_{j_0}|^2 \leq \sum_{j=1}^n |v_j|^2 = \|v\|_2^2.$$

Taking square roots gives $\|v\|_\infty \leq \|v\|_2$.

The inequality becomes equality, for example, for the vector $(1, 0, \dots, 0)$. ■

The ideas of norms and inner products are explored in some detail in Chapters III-3 and III-4.

1.1.3 Norms for multilinear maps

One of the places in the development of multivariable differentiation in Section 1.4 departs from the single-variable case is in higher-order derivatives. In the single-variable case, the derivative of a function is again a function, and so higher-order derivatives can be defined inductively as functions. But in the multivariable case, the derivative is a linear map as we shall see, and so to talk about higher-order derivatives one must talk intelligently about functions taking values in the set of linear maps. There are two facets to this. Firstly we must be comfortable with the algebraic aspects of multilinear maps. These are dealt with in Section I-5.6, and the reader will have to understand some material from this section before proceeding. Secondly, in order to inductively define higher-order derivatives we must have norms on sets of multilinear maps. We implicitly identify the set $\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n; \mathbb{R}^m)$ of linear maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ and the set $\mathrm{Mat}_{m\times n}(\mathbb{R})$ of $m \times n$ matrices with entries in $\mathbb{R}$; see Definition I-5.4.20. Thus for linear maps, the norms are sometimes called matrix norms.

First of all, we shall use somewhat more compact notation for multilinear maps than is used in Section I-5.6. Namely, we denote by $\mathrm{L}(\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_k}; \mathbb{R}^m)$ the set of

$\mathbb{R}$-multilinear maps from $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ to $\mathbb{R}^m$. (In Section I-5.6 we denoted this set of multilinear maps by $\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_k}; \mathbb{R}^m)$.) In the particular (and in this section usual) case when $n_1 = \cdots = n_k = n$ then we denote the multilinear maps from $(\mathbb{R}^n)^k$ to $\mathbb{R}^m$ by $\mathrm{L}^k(\mathbb{R}^n; \mathbb{R}^m)$. We also recall that a multilinear map $\mathsf{L} \in \mathrm{L}^k(\mathbb{R}^n; \mathbb{R}^m)$ is ***symmetric*** if

$$\mathsf{L}(v_{\sigma(1)}, \dots, v_{\sigma(k)}) = \mathsf{L}(v_1, \dots, v_k)$$

for every permutation $\sigma \in \mathfrak{S}_k$. We denote the set of symmetric multilinear maps from $(\mathbb{R}^n)^k$ to $\mathbb{R}^m$ by $S^k(\mathbb{R}^n; \mathbb{R}^m)$.

Our notation for multilinear maps will come back to us in when we talk about what continuous linear maps between normed vector spaces and in when we talk about what linear maps between topological vector spaces. In finite-dimensions all multilinear maps are continuous and so our notationally identifying $\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_k}; \mathbb{R}^m)$ with the continuous multilinear maps is justified. All that justification aside, all we care about is that $\mathrm{L}(\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_k}; \mathbb{R}^m)$ denotes the set of $\mathbb{R}$-multilinear maps from $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ to $\mathbb{R}^m$. Now we need to put norms on sets of linear and multilinear maps. The reader may well wish to refer ahead to Section III-3.1 for a general introduction to norms. Only the elementary definitions and examples from that section are needed here.

We will let $\|\cdot\|$ denote an arbitrary norm on $\mathbb{R}^n$. In practice, we shall most often take $\|\cdot\|$ to be the Euclidean norm, but we stick to a more general setup for simplicity. When talking about maps between $\mathbb{R}^n$ and $\mathbb{R}^m$, we will have norms on both spaces, and we shall denote both of these norms, and any norm induced by them, by $\|\cdot\|$, accepting an abuse of notation that does not cause problems.

With all of this preamble, we can now make the following definition.

1.1.12 Definition (Induced norm on the set of multilinear maps) Let $\|\cdot\|_{\alpha_1}, \dots, \|\cdot\|_{\alpha_k}$ be norms on $\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_k}$, respectively, and let $\|\cdot\|_{\beta}$ be a norm on $\mathbb{R}^m$. For $\mathsf{L} \in \mathrm{L}(\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_k}; \mathbb{R}^m)$ the ***induced norm*** of L is

$$\|\mathsf{L}\|_{\boldsymbol{\alpha},\beta} = \inf\{M \in \mathbb{R}_{>0} \mid \ \|\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_{\beta} \le M \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},\ \boldsymbol{x}_j \in \mathbb{R}^{n_j},\ j \in \{1, \dots, k\}\}.$$

•

Let us verify that the proposed norm is indeed a norm. The reader may wish to refer to Section III-3.5.3 for more information in the case of linear maps.

1.1.13 Proposition (The induced norm is a norm) *The induced norm defined in Definition 1.1.12 is a norm on* $\mathrm{L}(\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_k}; \mathbb{R}^m)$*. Moreover, for every* $\mathbf{x}_j \in \mathbb{R}^{n_j}$, $j \in \{1, \dots, k\}$,

$$\|\mathsf{L}(\mathbf{x}_1, \dots, \mathbf{x}_k)\|_{\beta} \le \|\mathsf{L}\|_{\boldsymbol{\alpha},\beta} \|\mathbf{x}_1\|_{\alpha_1} \cdots \|\mathbf{x}_k\|_{\alpha_k}.$$

Proof Let $\{\boldsymbol{e}_1, \dots, vecte_d\}$ be the standard basis for $\mathbb{R}^d$. For $\mathsf{L} \in \mathrm{L}(\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_k}; \mathbb{R}^m)$ define $\mathsf{L}^l_{j_1 \cdots j_k}$, $j_1 \in \{1, \dots, n_1\}, \dots, j_k \in \{1, \dots, n_k\}$, $l \in \{1, \dots, m\}$, by

$$\mathsf{L}(\boldsymbol{e}_{j_1}, \dots, \boldsymbol{e}_{j_k}) = \sum_{l=1}^{m} \mathsf{L}^m_{j_1 \cdots j_k} \boldsymbol{e}_l.$$

For $\boldsymbol{x}_j \in \mathbb{R}^{n_j}$, $j \in \{1, \dots, k\}$, let us write

$$\boldsymbol{x}_j = x_j^1 \boldsymbol{e}_1 + \cdots + x_j^{n_j} \boldsymbol{e}_{n_j}.$$

Then we have, by multilinearity of L,

$$\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \sum_{j_1=1}^{n_1} \cdots \sum_{j_k=1}^{n_k} \sum_{l=1}^{m} \mathsf{L}^l_{j_1 \cdots j_k} x_1^{j_1} \cdots x_k^{j_k} \boldsymbol{e}_l.$$

This shows that L is continuous since its components are polynomial functions of the components, and such functions are continuous.

Let us denote by $\overline{\mathsf{B}}(r, \boldsymbol{x})$ the closed ball of radius r centred at $\boldsymbol{x}$. We shall use the same notation for balls in any norm. Since L is continuous, by Theorem 1.3.31 it is bounded when restricted to the compact set $\overline{\mathsf{B}}(1, \boldsymbol{0}) \times \cdots \times \overline{\mathsf{B}}(1, \boldsymbol{0})$. Let

$$M = \sup\{\|\mathsf{L}(\boldsymbol{u}_1, \dots, \boldsymbol{u}_k)\|_\beta \mid \|\boldsymbol{u}_j\|_{\alpha_j} = 1,\ j \in \{1, \dots, k\}\}.$$

For $\boldsymbol{x}_j \in \mathbb{R}^{n_j} \setminus \{\boldsymbol{0}\}$, $j \in \{1, \dots, k\}$, we then have

$$\|\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta = \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k} \mathsf{L}\Big(\frac{\boldsymbol{x}_1}{\|\boldsymbol{x}_1\|_{\alpha_1}}, \dots, \frac{\boldsymbol{x}_k}{\|\boldsymbol{x}_k\|_{\alpha_k}}\Big) \le M \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k}.$$

This shows that $\|\mathsf{L}\|_{\boldsymbol{\alpha},\beta} < \infty$ and so is well-defined.

Let us next verify the final assertion of the proposition. Suppose that there exists $\boldsymbol{x}_j \in \mathbb{R}^{n_j}$, $j \in \{1, \dots, k\}$, such that

$$\|\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta > \|\mathsf{L}\|_{\boldsymbol{\alpha},\beta} \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k}.$$

Then there exists $\epsilon \in \mathbb{R}_{>0}$ such that

$$\|\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta > (\|\mathsf{L}\|_{\boldsymbol{\alpha},\beta} - \epsilon) \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},$$

and this contradicts the definition of $\|\mathsf{L}\|_{\boldsymbol{\alpha},\beta}$. Thus we must have

$$\|\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \le \|\mathsf{L}\|_{\boldsymbol{\alpha},\beta} \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k}, \tag{1.3}$$

as desired.

Now we show that $\mathsf{L} \mapsto \|\mathsf{L}\|_{\boldsymbol{\alpha},\beta}$ has the properties of a norm. It is clear that $\|\mathsf{L}\|_{\boldsymbol{\alpha},\beta} \ge 0$ and that $\|\mathsf{L}\|_{\boldsymbol{\alpha},\beta} = 0$ when $\mathsf{L} = 0$. Suppose that $\|\mathsf{L}\|_{\boldsymbol{\alpha},\beta} = 0$. Then, by (1.3), for every $\boldsymbol{x}_j \in \mathbb{R}^{n_j}$, $j \in \{1, \dots, k\}$,

$$\|\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \le \|\mathsf{L}\|_{\boldsymbol{\alpha},\beta} \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k} = 0,$$

giving $\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = 0$, and so $\mathsf{L} = 0$. Note that $\|0\mathsf{L}\|_{\boldsymbol{\alpha},\beta} = |0| \|\mathsf{L}\|_{\boldsymbol{\alpha},\beta}$. Also, if $a \in \mathbb{R} \setminus \{0\}$, then

$$\begin{aligned}
\|a\mathsf{L}\|_{\boldsymbol{\alpha},\beta} &= \inf\{M \in \mathbb{R}_{>0} \mid \ \|a\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \le M \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},\ \boldsymbol{x}_j \in \mathbb{R}^{n_j},\ j \in \{1, \dots, k\}\} \\
&= \inf\{M \in \mathbb{R}_{>0} \mid |a| \|\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \le M \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},\ \boldsymbol{x}_j \in \mathbb{R}^{n_j},\ j \in \{1, \dots, k\}\} \\
&= \inf\Big\{M \in \mathbb{R}_{>0} \ \Big|\ \|\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \le \frac{M}{|a|} \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},\ \boldsymbol{x}_j \in \mathbb{R}^{n_j},\ j \in \{1, \dots, k\}\Big\} \\
&= \inf\{|a|M' \in \mathbb{R}_{>0} \mid \|\mathsf{L}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \le M' \|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},\ \boldsymbol{x}_j \in \mathbb{R}^{n_j},\ j \in \{1, \dots, k\}\} \\
&= |a| \|\mathsf{L}\|_{\boldsymbol{\alpha},\beta},
\end{aligned}$$

using Proposition I-2.2.28. Finally, if $\mathsf{L}_1, \mathsf{L}_2 \in \mathrm{L}(\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_k}; \mathbb{R}^m)$, then

$$\begin{aligned}
\|\mathsf{L}_1 + \mathsf{L}_2\|_{\boldsymbol{\alpha},\beta} &= \inf\{M \in \mathbb{R}_{>0} \mid \ \|(\mathsf{L}_1 + \mathsf{L}_2)(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \\
&\qquad \le M\|\boldsymbol{x}\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},\ \boldsymbol{x}_j \in \mathbb{R}^{n_j},\ j \in \{1, \dots, k\}\} \\
&\le \inf\{M \in \mathbb{R}_{>0} \mid \ \|\mathsf{L}_1(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \\
&\qquad + \|\mathsf{L}_2(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \le M\|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},\ \boldsymbol{x}_j \in \mathbb{R}^{n_j},\ j \in \{1, \dots, k\}\} \\
&= \inf\{M_1 + M_2 \in \mathbb{R}_{>0} \mid \ \|\mathsf{L}_1(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \le M_1\|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k}, \\
&\qquad \|\mathsf{L}_2(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \le M_2\|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},\ \boldsymbol{x}_j \in \mathbb{R}^{n_j},\ j \in \{1, \dots, k\}\} \\
&= \inf\{M \in \mathbb{R}_{>0} \mid \ \|\mathsf{L}_1(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \le \\
&\qquad M\|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},\ \boldsymbol{x}_j \in \mathbb{R}^{n_j},\ j \in \{1, \dots, k\}\} \\
&\qquad + \inf\{M \in \mathbb{R}_{>0} \mid \ \|\mathsf{L}_2(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_\beta \\
&\qquad \le M\|\boldsymbol{x}_1\|_{\alpha_1} \cdots \|\boldsymbol{x}_k\|_{\alpha_k},\ \boldsymbol{x}_j \in \mathbb{R}^{n_j},\ j \in \{1, \dots, k\}\} \\
&= \|\mathsf{L}_1\|_{\boldsymbol{\alpha},\beta} + \|\mathsf{L}_2\|_{\boldsymbol{\alpha},\beta},
\end{aligned}$$

using Proposition I-2.2.28. ■

1.1.4 The nine common induced norms for linear maps

Let us consider a collection of special cases for linear maps. We use the three norms

$$\|\boldsymbol{x}\|_1 = \sum_{j=1}^n |x_j|, \quad \|\boldsymbol{x}\|_2 = \Big(\sum_{j=1}^n x_j^2\Big)^{1/2}, \quad \|\boldsymbol{x}\|_\infty = \max\{|x_1|, \dots, |x_n|\}$$

on $\mathbb{R}^n$, noting that $\|\cdot\|_2$ is the Euclidean norm, which we have also denoted by $\|\cdot\|_{\mathbb{R}^n}$. Let us characterise the nine possible induced norms

$$\|\mathsf{L}\|_{p,q} \triangleq \inf\{M \in \mathbb{R}_{>0} \mid \|\mathsf{L}(\boldsymbol{x})\|_q \le M\|\boldsymbol{x}\|_p,\ \boldsymbol{x} \in \mathbb{R}^n\}, \qquad p, q \in \{1, 2, \infty\},$$

on $\mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$ induced by these three norms. In the statement of the following theorem, recall from Definition I-5.1.4 that $\boldsymbol{c}(\mathsf{L}, j) \in \mathbb{R}^m$, $j \in \{1, \dots, n\}$, denotes the jth column vector of L and $\boldsymbol{r}(\mathsf{L}, a) \in \mathbb{R}^n$, $a \in \{1, \dots, m\}$, denotes the ath row vector of L, where we recall from Theorem I-5.1.13 that there is a natural correspondence between finite matrices and linear maps.

1.1.14 Theorem (Induced norms for linear maps) *Let* $p, q \in \{1, 2, \infty\}$ *and let* $\mathsf{L} \in \mathrm{L}(\mathbb{R}^m; \mathbb{R}^m)$. *The induced norm* $\|\cdot\|_{p,q}$ *satisfies the following formulae:*

(i) $\|\mathsf{L}\|_{1,1} = \max\{\|\mathbf{c}(\mathsf{L}, j)\|_1 \mid j \in \{1, \dots, n\}\}$;

(ii) $\|\mathsf{L}\|_{1,2} = \max\{\|\mathbf{c}(\mathsf{L}, j)\|_2 \mid j \in \{1, \dots, n\}\}$;

(iii) $$\begin{aligned}\|\mathsf{L}\|_{1,\infty} &= \max\{|\mathsf{L}(a, j)| \mid a \in \{1, \dots, m\},\ j \in \{1, \dots, n\}\}; \\ &= \max\{\|\mathbf{c}(\mathsf{L}, j)\|_\infty \mid j \in \{1, \dots, n\}\} \\ &= \max\{\|\mathbf{r}(\mathsf{L}, a)\|_\infty \mid a \in \{1, \dots, m\}\}\end{aligned}$$

(iv) $\|\mathsf{L}\|_{2,1} = \max\{\|\mathsf{L}^T(\mathbf{u})\|_2 \mid \mathbf{u} \in \{-1, 1\}^m\}$;

(v) $\|\mathsf{L}\|_{2,2} = \max\{\sqrt{\lambda} \mid \lambda$ *is an eigenvalue for* $\mathsf{L}^T\mathsf{L}\}$;

(vi) $\|\mathsf{L}\|_{2,\infty} = \max\{\|\mathbf{r}(\mathsf{L},a)\|_2 \mid a \in \{1,\dots,m\}\}$;

(vii) $\|\mathsf{L}\|_{\infty,1} = \max\{\|\mathsf{L}(\mathbf{u})\|_1 \mid \mathbf{u} \in \{-1,1\}^n\}$;

(viii) $\|\mathsf{L}\|_{\infty,2} = \max\{\|\mathsf{L}(\mathbf{u})\|_2 \mid \mathbf{u} \in \{-1,1\}^n\}$;

(ix) $\|\mathsf{L}\|_{\infty,\infty} = \max\{\|\mathbf{r}(\mathsf{L},a)\|_1 \mid a \in \{1,\dots,m\}\}$.

Proof In the proof we make free use of results we have not yet proved. We also make frequent use of the obvious formula

$$\mathsf{L}(\boldsymbol{x}) = \big(\langle \boldsymbol{r}(\mathsf{L},1), \boldsymbol{x}\rangle_{\mathbb{R}^n}, \dots, \langle \boldsymbol{r}(\mathsf{L},m), \boldsymbol{x}\rangle_{\mathbb{R}^n}\big).$$

Let $\mathsf{L} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ and note that

$$\begin{aligned}\|\mathsf{L}\| &= \inf\{M \in \mathbb{R}_{>0} \mid \|\mathsf{L}(\boldsymbol{x})\| \le M\|\boldsymbol{x}\|,\ \boldsymbol{x} \in \mathbb{R}^n\}\\ &= \{M \in \mathbb{R}_{>0} \mid \|\mathsf{L}(\boldsymbol{x})\| \le M\|\boldsymbol{x}\|,\ \boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}\}\\ &= \{M \in \mathbb{R}_{>0} \mid \|\mathsf{L}(\tfrac{\boldsymbol{x}}{\|\boldsymbol{x}\|})\| \le M,\ \boldsymbol{x} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}\}\\ &= \sup\{\|\mathsf{L}(\boldsymbol{x})\| \mid \|\boldsymbol{x}\| = 1\}.\end{aligned}$$

We shall use this characterisation of the norm below.

In the proof, we also let $\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_d\}$ be the standard basis for $\mathbb{R}^d$.

(i) We compute

$$\begin{aligned}\|\mathsf{L}\|_{1,1} &= \sup\{\|\mathsf{L}(\boldsymbol{x})\|_1 \mid \|\boldsymbol{x}\|_1 = 1\}\\ &= \sup\Big\{\sum_{a=1}^m |\langle \boldsymbol{r}(\mathsf{L}(\boldsymbol{x})), \boldsymbol{x}\rangle_{\mathbb{R}^n}| \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\}\\ &\le \sup\Big\{\sum_{a=1}^m \sum_{j=1}^n |\mathsf{L}(a,j)||x_j| \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\}\\ &= \sup\Big\{\sum_{j=1}^n |x_j|\Big(\sum_{a=1}^m |\mathsf{L}(a,j)|\Big) \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\}\\ &\le \max\Big\{\sum_{a=1}^m |\mathsf{L}(a,j)| \;\Big|\; j \in \{1,\dots,n\}\Big\}\\ &= \max\{\|\boldsymbol{c}(\mathsf{L},j)\|_1 \mid j \in \{1,\dots,n\}\}.\end{aligned}$$

To establish the opposite inequality, suppose that $k \in \{1,\dots,n\}$ is such that

$$\|\boldsymbol{c}(\mathsf{L},k)\|_1 = \max\{\|\boldsymbol{c}(\mathsf{L},j)\|_1 \mid j \in \{1,\dots,n\}\}.$$

Then,

$$\|\mathsf{L}(\boldsymbol{e}_k)\|_1 = \sum_{a=1}^m \Big|\Big(\sum_{j=1}^n \mathsf{L}(a,j)\boldsymbol{e}_k(j)\Big)\Big| = \sum_{a=1}^m |\mathsf{L}(a,k)| = \|\boldsymbol{c}(\mathsf{L},k)\|_1.$$

Thus

$$\|\mathsf{L}\|_{1,1} \ge \max\{\|\boldsymbol{c}(\mathsf{L},j)\|_1 \mid j \in \{1,\dots,n\}\},$$

since $\|\boldsymbol{e}_k\|_1 = 1$.

(ii) We compute

$$\begin{aligned}
\|\mathsf{L}\|_{1,2} &= \sup\{\|\mathsf{L}(\boldsymbol{x})\|_2 \mid \|\boldsymbol{x}\|_1 = 1\} \\
&= \sup\Big\{\Big(\sum_{a=1}^{m} \langle \boldsymbol{r}(\mathsf{L}, a), \boldsymbol{x}\rangle_{\mathbb{R}^n}^2\Big)^{1/2} \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\} \\
&\le \sup\Big\{\Big(\sum_{a=1}^{m}\Big(\sum_{j=1}^{n} |\mathsf{L}(a, j)x_j|\Big)^2\Big)^{1/2} \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\} \\
&\le \sup\Big\{\Big(\sum_{a=1}^{m} (\max\{|\mathsf{L}(a, j)| \mid j \in \{1, \dots, n\}\})^2 \Big(\sum_{j=1}^{n} |x_j|\Big)^2\Big)^{1/2} \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\} \\
&= \Big(\sum_{a=1}^{m} (\max\{|\mathsf{L}(a, j)| \mid j \in \{1, \dots, n\}\})^2\Big)^{1/2} \\
&= \Big(\max\Big\{\sum_{a=1}^{m} \mathsf{L}(a, j)^2 \;\Big|\; j \in \{1, \dots, n\}\Big\}\Big)^{1/2} = \max\{\|\boldsymbol{c}(\mathsf{L}, j)\|_2 \mid j \in \{1, \dots, n\}\},
\end{aligned}$$

using Proposition I-2.2.27 and the fact that

$$\sup\{\|\boldsymbol{x}\|_2 \mid \|\boldsymbol{x}\|_1 = 1\} = 1.$$

To establish the other inequality, note that if we take $k \in \{1, \dots, n\}$ such that

$$\|\boldsymbol{c}(\mathsf{L}, k)\|_2 = \max\{\|\boldsymbol{c}(\mathsf{L}, j)\|_2 \mid j \in \{1, \dots, n\}\},$$

then we have

$$\|\mathsf{L}(\boldsymbol{e}_k)\|_2 = \Big(\sum_{a=1}^{m}\Big(\sum_{j=1}^{n} \mathsf{L}(a, j)\boldsymbol{e}_k(j)\Big)^2\Big)^{1/2} = \Big(\sum_{a=1}^{m} \mathsf{L}(a, k)^2\Big)^{1/2} = \|\boldsymbol{c}(\mathsf{L}, k)\|_2.$$

Thus

$$\|\mathsf{L}\|_{1,2} \ge \max\{\|\boldsymbol{c}(\mathsf{L}, j)\|_2 \mid j \in \{1, \dots, n\}\},$$

since $\|\boldsymbol{e}_k\|_1 = 1$.

(iii) Here we compute

$$\begin{aligned}
\|\mathsf{L}\|_{1,\infty} &= \sup\{\|\mathsf{L}(\boldsymbol{x})\|_\infty \mid \|\boldsymbol{x}\|_1 = 1\} \\
&= \sup\Big\{\max\Big\{\Big|\sum_{j=1}^{n} \mathsf{L}(a, j)x_j\Big| \;\Big|\; a \in \{1, \dots, m\}\Big\} \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\} \\
&\le \sup\Big\{\max\Big\{|\mathsf{L}(a, j)| \;\Big|\; j \in \{1, \dots, n\},\ a \in \{1, \dots, m\}\Big\}\Big(\sum_{j=1}^{n} |x_j|\Big) \;\Big|\; \|\boldsymbol{x}\|_1 = 1\Big\} \\
&= \max\{|\mathsf{L}(a, j)| \mid j \in \{1, \dots, n\},\ a \in \{1, \dots, m\}\}.
\end{aligned}$$

For the converse inequality, let $k \in \{1, \dots, n\}$ be such that

$$\max\{|\mathsf{L}(a, k)| \mid a \in \{1, \dots, m\}\} = \max\{|\mathsf{L}(a, j)| \mid j \in \{1, \dots, n\},\ a \in \{1, \dots, m\}\}.$$

Then

$$\begin{aligned}\|\mathsf{L}(\boldsymbol{e}_k)\|_\infty &= \max\left\{\left|\sum_{j=1}^n \mathsf{L}(a,j)\boldsymbol{e}_k(j)\right| \;\middle|\; a\in\{1,\dots,m\}\right\}\\ &= \max\{|\mathsf{L}(a,k)| \mid a\in\{1,\dots,m\}\}.\end{aligned}$$

Thus

$$\|\mathsf{L}\|_{1,\infty} \geq \max\{|\mathsf{L}(a,j)| \mid j\in\{1,\dots,n\},\ a\in\{1,\dots,m\}\},$$

since $\|\boldsymbol{e}_k\|_1 = 1$.

(iv) In this case we maximise the function $\boldsymbol{x}\mapsto\|\mathsf{L}(\boldsymbol{x})\|_1$ subject to the constraint that $\|\boldsymbol{x}\|_2 = 1$, or equivalently, subject to the constraint that $\|\boldsymbol{x}\|_2^2 = 1$. We shall do this using Theorem 1.4.44 and defining

$$f(\boldsymbol{x}) = \|\mathsf{L}(\boldsymbol{x})\|_1, \quad g(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2 - 1.$$

Let us first assume that none of the rows of L are zero. We must exercise some care because f is not differentiable on $\mathbb{R}^n$. Note that

$$\|\mathsf{L}(\boldsymbol{x})\|_1 = \sum_{a=1}^m |\langle \boldsymbol{r}(\mathsf{L},a),\boldsymbol{x}\rangle_{\mathbb{R}^n}|.$$

Thus f is differentiable at points off the set

$$B_{\mathsf{L}} = \{\boldsymbol{x}\in\mathbb{R}^n \mid \text{there exists } a\in\{1,\dots,m\} \text{ such that } \langle \boldsymbol{r}(\mathsf{L},a),\boldsymbol{x}\rangle_{\mathbb{R}^n} = 0\}.$$

To facilitate computations, let us define $\boldsymbol{u}_{\mathsf{L}}\colon\mathbb{R}^n\to\mathbb{R}^m$ by asking that

$$u_{\mathsf{L},a}(\boldsymbol{x}) = \operatorname{sign}(\langle \boldsymbol{r}(\mathsf{L},a),\boldsymbol{x}\rangle_{\mathbb{R}^n}).$$

Note that $B_{\mathsf{L}} = \boldsymbol{u}_{\mathsf{L}}^{-1}(\boldsymbol{0})$. Note that on $\mathbb{R}^n\setminus B_{\mathsf{L}}$ the function $\boldsymbol{u}_{\mathsf{L}}$ is locally constant. That is to say, if $\boldsymbol{x}\in\mathbb{R}^n\setminus B$, then there is a neighbourhood $U\subseteq\mathbb{R}^n\setminus B_{\mathsf{L}}$ of $\boldsymbol{x}$ such that $\boldsymbol{u}_{\mathsf{L}}|U$ is constant (why?). Moreover, it is clear that

$$f(\boldsymbol{x}) = \langle \boldsymbol{u}_{\mathsf{L}}(\boldsymbol{x}),\mathsf{L}(\boldsymbol{x})\rangle_{\mathbb{R}^m}.$$

Now let $\boldsymbol{x}_0\in\mathbb{R}^n\setminus B_{\mathsf{L}}$ be a maximum of f subject to the constraint that $g(\boldsymbol{x}) = 0$. Note that

$$\boldsymbol{D}g(\boldsymbol{x})\cdot\boldsymbol{v} = \langle\boldsymbol{x},\boldsymbol{v}\rangle_{\mathbb{R}^n} + \langle\boldsymbol{v},\boldsymbol{x}\rangle_{\mathbb{R}^n} = 2\langle\boldsymbol{x},\boldsymbol{v}\rangle_{\mathbb{R}^n},$$

and so, if $\boldsymbol{x}\neq\boldsymbol{0}$, then we can conclude that $\boldsymbol{D}g(\boldsymbol{x})$ has rank 1. Thus, by Theorem 1.4.44, there exists $\lambda\in\mathbb{R}$ such that

$$\boldsymbol{D}(f-\lambda g)(\boldsymbol{x}_0) = \boldsymbol{0}.$$

Since $\boldsymbol{u}_{\mathsf{L}}$ is locally constant,

$$\boldsymbol{D}f(\boldsymbol{x}_0)\cdot\boldsymbol{v} = \langle\boldsymbol{u}_{\mathsf{L}}(\boldsymbol{x}_0),\mathsf{L}(\boldsymbol{v})\rangle_{\mathbb{R}^m}.$$

Moreover, $\boldsymbol{D}g(\boldsymbol{x})\cdot\boldsymbol{v} = 2\langle\boldsymbol{x},\boldsymbol{v}\rangle_{\mathbb{R}^n}$. Thus $\boldsymbol{D}(f-\lambda g)(\boldsymbol{x}_0) = \boldsymbol{0}$ if and only if

$$\mathsf{L}^T(\boldsymbol{u}_{\mathsf{L}}(\boldsymbol{x}_0)) = 2\lambda\boldsymbol{x}_0 \quad\Longrightarrow\quad |\lambda| = \frac{1}{2}\|\mathsf{L}^T(\boldsymbol{u}_{\mathsf{L}}(\boldsymbol{x}_0))\|_2,$$

since $\|x_0\|_2 = 1$. Thus $\lambda = 0$ if and only if $\mathsf{L}^T(u_{\mathsf{L}}(x_0)) = \mathbf{0}$. Therefore, if $\lambda = 0$ then

$$f(x_0) = \langle u_{\mathsf{L}}(x_0), \mathsf{L}(x_0)\rangle_{\mathbb{R}^m} = \langle \mathsf{L}^T(u_{\mathsf{L}}(x_0)), x_0\rangle_{\mathbb{R}^n} = 0.$$

If $\lambda \neq 0$ then

$$f(x_0) = \langle \mathsf{L}^T(u_{\mathsf{L}}(x_0)), x_0\rangle_{\mathbb{R}^n} = \frac{1}{2\lambda}\|\mathsf{L}^T(u_{\mathsf{L}}(x_0))\|_2^2 = \frac{2}{\lambda}\lambda^2 = 2\lambda.$$

Observing that $|\lambda| = \|\mathsf{L}^T(u_{\mathsf{L}}(x_0))\|_2$ and that f is nonnegative-valued, we can conclude that, at solutions of the constrained maximisation problem, we must have

$$f(x_0) = \|\mathsf{L}^T(u)\|_2,$$

where u varies over the nonzero points in the image of u_{L}, i.e., over points from $\{-1, 1\}^m$.

This would conclude the proof of this part of the theorem in the case that L has no zero rows, but for the fact that it is possible that f attains its maximum on B_{L}. We now show that this does not happen. Let $x_0 \in B_{\mathsf{L}}$ satisfy $\|x_0\|_2 = 1$ and denote

$$A_0 = \{a \in \{1, \dots, m\} \mid u_{\mathsf{L},a}(x_0) = 0\}.$$

Let $A_1 = \{1, \dots, m\} \setminus A_0$. Let $a_0 \in A_0$. For $\epsilon \in \mathbb{R}$ define

$$x_\epsilon = \frac{x_0 + \epsilon r(\mathsf{L}, a_0)}{\sqrt{1 + \epsilon^2\|r(\mathsf{L}, a_0)\|_2^2}}.$$

Note that

$$\|x_0 + \epsilon r(\mathsf{L}, a_0)\|_2^2 = \|x_0\|_2^2 + \epsilon^2\|r(\mathsf{L}, a_0)\|_2^2 = 1 + \epsilon^2\|r(\mathsf{L}, a_0)\|_2^2$$

since $\langle r(\mathsf{L}, a_0), x_0\rangle_{\mathbb{R}^n} = 0$. Thus x_ϵ satisfies the constraint $\|x_\epsilon\|_2^2 = 1$. Now let $\epsilon_0 \in \mathbb{R}_{>0}$ be sufficiently small that

$$\langle r(\mathsf{L}, a), x_\epsilon\rangle_{\mathbb{R}^n} \neq 0$$

for all $a \in A_1$ and $\epsilon \in [-\epsilon_0, \epsilon_0]$; this is possible since x_ϵ depends continuously on ϵ. Then we compute

$$\begin{aligned}\|\mathsf{L}(x_\epsilon)\|_1 &= \sum_{a=1}^{m} |\langle r(\mathsf{L}, a), x_\epsilon\rangle_{\mathbb{R}^n}|\\ &= \frac{1}{\sqrt{1 + \epsilon^2\|r(\mathsf{L}, a_0)\|_2^2}} \sum_{a=1}^{m} |\langle r(\mathsf{L}, a), x_0\rangle_{\mathbb{R}^n} + \epsilon\langle r(\mathsf{L}, a), r(\mathsf{L}, a_0)\rangle_{\mathbb{R}^n}|.\end{aligned}$$

Note that, by Taylor Theorem, , we can write what?

$$\frac{1}{\sqrt{1 + \epsilon^2\|r(\mathsf{L}, a_0)\|_2^2}} = 1 - \epsilon^2\frac{\|r(\mathsf{L}, a_0)\|_2^2}{2} + O(\epsilon^3),$$

so that, for ϵ sufficiently small,

$$\begin{aligned}\|\mathsf{L}(\boldsymbol{x}_\epsilon)\|_1 &= \sum_{a=1}^m |\langle \boldsymbol{r}(\mathsf{L},a), \boldsymbol{x}_0\rangle_{\mathbb{R}^n} + \epsilon\langle \boldsymbol{r}(\mathsf{L},a), \boldsymbol{r}(\mathsf{L},a_0)\rangle_{\mathbb{R}^n}| + O(\epsilon^2)\\ &= \sum_{a\in A_0} |\epsilon||\langle \boldsymbol{r}(\mathsf{L},a), \boldsymbol{r}(\mathsf{L},a_0)\rangle_{\mathbb{R}^n}|\\ &\quad + \sum_{a\in A_1} |\langle \boldsymbol{r}(\mathsf{L},a), \boldsymbol{x}_0\rangle_{\mathbb{R}^n} + \epsilon\langle \boldsymbol{r}(\mathsf{L},a), \boldsymbol{r}(\mathsf{L},a_0)\rangle_{\mathbb{R}^n}| + O(\epsilon^2).\end{aligned} \tag{1.4}$$

Since we are assuming that none of the rows of L are zero,

$$\sum_{a\in A_0} |\epsilon||\langle \boldsymbol{r}(\mathsf{L},a), \boldsymbol{r}(\mathsf{L},a_0)\rangle_{\mathbb{R}^n}| > 0 \tag{1.5}$$

for $\epsilon \in [-\epsilon_0, \epsilon_0]$. Now take $a \in A_1$. If ϵ is sufficiently small we can write

$$|\langle \boldsymbol{r}(\mathsf{L},a), \boldsymbol{x}_0\rangle_{\mathbb{R}^n} + \epsilon\langle \boldsymbol{r}(\mathsf{L},a), \boldsymbol{r}(\mathsf{L},a_0)\rangle_{\mathbb{R}^n}| = |\langle \boldsymbol{r}(\mathsf{L},a), \boldsymbol{x}_0\rangle_{\mathbb{R}^n}| + \epsilon C_a$$

for some $C_a \in \mathbb{R}$. As a result, and using (1.4), we have

$$\|\mathsf{L}(\boldsymbol{x}_\epsilon)\|_1 = \|\mathsf{L}(\boldsymbol{x}_0)\|_1 + \sum_{a\in A_0} (|\epsilon||\langle \boldsymbol{r}(\mathsf{L},a), \boldsymbol{r}(\mathsf{L},a_0)\rangle_{\mathbb{R}^n}| + \epsilon \sum_{a\in A_1} C_a + O(\epsilon^2).$$

It therefore follows, possibly by again choosing ϵ_0 to be sufficiently small, that we have

$$\|\mathsf{L}(\boldsymbol{x}_\epsilon)\|_1 > \|\mathsf{L}(\boldsymbol{x}_0)\|_1$$

either for all $\epsilon \in [-\epsilon_0, 0)$ or for all $\epsilon \in (0, \epsilon_0]$, taking (1.5) into account. Thus if $\boldsymbol{x}_0 \in B_{\mathsf{L}}$ then $\boldsymbol{x}_0$ is not a local maximum for f subject to the constraint $g^{-1}(0)$.

Finally, suppose that L has some rows that are zero. Let

$$A_0 = \{a \in \{1, \dots, m\} \mid \boldsymbol{r}(\mathsf{L}, a) = \boldsymbol{0}\}$$

and let $A_1 = \{1, \dots, m\} \setminus A_0$. Let $A_1 = \{a_1, \dots, a_k\}$ with $a_1 < \dots < a_k$, and define $\hat{\mathsf{L}} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^k)$ by

$$\hat{\mathsf{L}}(\boldsymbol{x}) = \sum_{r=1}^k \langle \boldsymbol{r}(\mathsf{L}, a_r), \boldsymbol{x}\rangle_{\mathbb{R}^n} \boldsymbol{e}_r,$$

and note that $\|\mathsf{L}(\boldsymbol{x})\|_1 = \|\hat{\mathsf{L}}(\boldsymbol{x})\|_1$ for every $\boldsymbol{x} \in \mathbb{R}^n$. If $\boldsymbol{y} \in \mathbb{R}^m$ define $\hat{\boldsymbol{y}} \in \mathbb{R}^k$ by removing from $\boldsymbol{y}$ the elements corresponding to the zero rows of L:

$$\hat{\boldsymbol{y}} = (y_{a_1}, \dots, y_{a_k}).$$

Then we compute

$$\begin{aligned}\mathsf{L}^T(\boldsymbol{y}) &= \sum_{j=1}^n \langle \boldsymbol{r}(\mathsf{L}^T, j), \boldsymbol{y}\rangle_{\mathbb{R}^n} \boldsymbol{e}_j = \sum_{j=1}^n \Big(\sum_{a=1}^m \mathsf{L}(a, j) y_a\Big) \boldsymbol{e}_j\\ &= \sum_{j=1}^n \Big(\sum_{r=1}^k \mathsf{L}(a_r, j) y_{a_r}\Big) \boldsymbol{e}_j = \sum_{j=1}^n \langle \boldsymbol{c}(\hat{\mathsf{L}}, r), \hat{\boldsymbol{y}}\rangle_{\mathbb{R}^n} \boldsymbol{e}_j\\ &= \sum_{j=1}^n \langle \boldsymbol{t}(\hat{\mathsf{L}}^T, r), \hat{\boldsymbol{y}}\rangle_{\mathbb{R}^n} \boldsymbol{e}_j = \hat{\mathsf{L}}^T(\hat{\boldsymbol{y}}).\end{aligned}$$

Therefore,

$$\begin{aligned}\|\mathsf{L}\|_{2,1} &= \sup\{\|\mathsf{L}(x)\|_1 \mid \|x\|_2 = 1\}\\ &= \sup\{\|\hat{\mathsf{L}}(x)\|_1 \mid \|x\|_2 = 1\} = \|\hat{\mathsf{L}}\|_{2,1}\\ &= \max\{\|\hat{\mathsf{L}}^T(\hat{u})\|_2 \mid \hat{u} \in \{-1,1\}^k\}\\ &= \max\{\|\mathsf{L}^T(u)\|_2 \mid u \in \{-1,1\}^m\},\end{aligned}$$

and this finally gives the result.

(v) Note that, in this case, we wish to maximise the function $x \mapsto \|\mathsf{L}(x)\|_2$ subject to the constraint that $\|x\|_2 = 1$. However, this is equivalent to maximising $x \mapsto \|\mathsf{L}(x)\|_2^2$ subject to the constraint that $\|x\|_2^2 = 1$. In this case, the function we are maximising and the function defining the constraint are infinitely differentiable. Therefore, we can use Theorem 1.4.44 below to determine the character of the maxima. Thus we define

$$f(x) = \|\mathsf{L}(x)\|_2^2, \quad g(x) = \|x\|_2^2 - 1.$$

Note that

$$\boldsymbol{D}g(x) \cdot v = \langle x, v\rangle_{\mathbb{R}^n} + \langle v, x\rangle_{\mathbb{R}^n} = 2\langle x, v\rangle_{\mathbb{R}^n},$$

and so, if $x \neq \mathbf{0}$, then we can conclude that $\boldsymbol{D}g(x)$ has rank 1. Thus, by Theorem 1.4.44, if a point $x_0 \in \mathbb{R}^n$ solves the constrained maximisation problem, then there exists $\lambda \in \mathbb{R}$ such that

$$\boldsymbol{D}(f - \lambda g)(x_0) = 0.$$

Since

$$f(x) = \langle \mathsf{L}(x), \mathsf{L}(x)\rangle_{\mathbb{R}^n} = \langle \mathsf{L}^T \circ \mathsf{L}(x), x\rangle_{\mathbb{R}^n},$$

we compute

$$\boldsymbol{D}f(x) \cdot v = \langle \mathsf{L}^T \circ \mathsf{L}(x), v\rangle_{\mathbb{R}^n} + \langle \mathsf{L}^T \circ \mathsf{L}(v), x\rangle_{\mathbb{R}^n} = 2\langle \mathsf{L}^T \circ \mathsf{L}(x), v\rangle_{\mathbb{R}^n}.$$

We also have $\boldsymbol{D}g(x) \cdot v = 2\langle x, v\rangle_{\mathbb{R}^n}$. Thus $\boldsymbol{D}(f - \lambda g)(x_0) = 0$ implies that

$$\mathsf{L}^T \circ \mathsf{L}(x_0) = \lambda x_0.$$

Thus it must be the case that λ is an eigenvalue for $\mathsf{L}^T \circ \mathsf{L}$ with eigenvector x_0. Let us record some facts about this eigenvalue/eigenvector combination.

1 **Lemma** *If* $\mathsf{L} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$ *then the linear map* $\mathsf{L}^T \circ \mathsf{L} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ *has the following properties:*

(i) all eigenvalues of $\mathsf{L}^T \circ \mathsf{L}$ *are real and nonnegative;*

(ii) the exists a basis for $\mathbb{R}^n$*, orthonormal with respect to the Euclidean inner product, consisting of eigenvectors of* $\mathsf{L}^T \circ \mathsf{L}$*.*

Proof First of all, note that

$$(\mathsf{L}^T \circ \mathsf{L})^T = \mathsf{L}^T \circ \mathsf{L},$$

and so, by , the linear map $\mathsf{L}^T \circ \mathsf{L}$ is symmetric with respect to the Euclidean inner product. Thus the eigenvalues of $\mathsf{L}^T \circ \mathsf{L}$ are real. Also note that what?

$$\langle \mathsf{L}^T \circ \mathsf{L}(x), x\rangle_{\mathbb{R}^n} = \langle \mathsf{L}(x), \mathsf{L}(x)\rangle_{\mathbb{R}^n} \geq 0$$

by , and so the eigenvalues of $\mathsf{L}^T \circ \mathsf{L}$ are nonnegative by . what

That there is a basis of eigenvectors for $\mathbb{R}^n$, orthonormal with respect to $\langle\cdot,\cdot\rangle_{\mathbb{R}^n}$, what? follows from . ▼ what?

Let us proceed with our analysis. The lemma implies that there exist $\lambda_1, \dots, \lambda_n \in \mathbb{R}_{\geq 0}$ and vectors $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$ such that

$$\lambda_1 \leq \cdots \leq \lambda_n,$$

such that $\mathsf{L}^T \circ \mathsf{L}(\boldsymbol{x}_j) = \lambda_j \boldsymbol{x}_j$, $j \in \{1, \dots, n\}$, and such that a solution to the problem of maximising f with the constraint $g^{-1}(0)$ is obtained by evaluating f at one of the points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_n$. Thus the problem can be solved by evaluating f at this finite collection of points, and determining at which of these f has its largest value. Thus we compute

$$f(\boldsymbol{x}_j) = \|\mathsf{L}(\boldsymbol{x}_j)\|_2^2 = \langle \mathsf{L}(\boldsymbol{x}), \mathsf{L}(\boldsymbol{x}_j)\rangle_{\mathbb{R}^m} = \langle \mathsf{L}^T \circ \mathsf{L}(\boldsymbol{x}_j), \boldsymbol{x}_j\rangle_{\mathbb{R}^n} = \lambda_j \|\boldsymbol{x}_j\|_2^2 = \lambda_j.$$

The maximum value of f subject to the constraint $g^{-1}(0)$ is then attained at $\boldsymbol{x}_n$ and this maximum value is λ_n. Thus the maximum value of the function $\boldsymbol{x} \mapsto \|\mathsf{L}(\boldsymbol{x})\|_2$ subject to the constraint that $\|\boldsymbol{x}\|_2 = 1$ is $\sqrt{\lambda_n}$, and this gives the desired result.

(vi) First of all, we note that this part of the theorem certainly holds when $\mathsf{L} = 0$. Thus we shall freely assume that L is nonzero when convenient. We maximise the function $\boldsymbol{x} \mapsto \|\mathsf{L}(\boldsymbol{x})\|_\infty$ subject to the constraint that $\|\boldsymbol{x}\|_2 = 1$, or equivalently subject to the constraint that $\|\boldsymbol{x}\|_2^2 = 1$. We shall use Theorem 1.4.44, defining

$$f(\boldsymbol{x}) = \|\mathsf{L}(\boldsymbol{x})\|_\infty, \quad g(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2 - 1.$$

Note that L is not differentiable on $\mathbb{R}^n$, so we first restrict to a subset where f is differentiable. Let us define

$$\begin{aligned} A_{\mathsf{L}} \colon \mathbb{R}^n &\to \mathbf{2}^{\{1,\dots,m\}} \\ \boldsymbol{x} &\mapsto \{a \in \{1, \dots, m\} \mid \langle \boldsymbol{r}(\mathsf{L}, a), \boldsymbol{x}\rangle_{\mathbb{R}^n} = \|\mathsf{L}(\boldsymbol{x})\|_\infty\}. \end{aligned}$$

Then denote

$$B_{\mathsf{L}} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \operatorname{card}(A_{\mathsf{L}}(\boldsymbol{x})) > 1\}.$$

Since

$$\|\mathsf{L}(\boldsymbol{x})\|_\infty = \max\{\langle \boldsymbol{r}(\mathsf{L}, 1), \boldsymbol{x}\rangle_{\mathbb{R}^n}, \dots, \langle \boldsymbol{r}(\mathsf{L}m), \boldsymbol{x}\rangle_{\mathbb{R}^n}\},$$

we see that f is differentiable at points that are not in the set B_{L}.

Let us first suppose that $\boldsymbol{x}_0 \in \mathbb{R}^n \setminus B_{\mathsf{L}}$ is a maximum of f subject to the constraint that $g(\boldsymbol{x}) = 0$. Then there exists a unique $a_0 \in \{1, \dots, m\}$ such that $f(\boldsymbol{x}_0) = \langle \boldsymbol{r}(\mathsf{L}, a_0), \boldsymbol{x}_0\rangle_{\mathbb{R}^n}$. Since we are assuming that L is nonzero, it must be that $\boldsymbol{r}(\mathsf{L}, a_0)$ is nonzero. Moreover, there exists a neighbourhood U of $\boldsymbol{x}_0$ such that

$$\operatorname{sign}(\langle \boldsymbol{r}(\mathsf{L}, a_0), \boldsymbol{x}\rangle_{\mathbb{R}^n}) = \operatorname{sign}(\langle \boldsymbol{r}(\mathsf{L}, a_0), \boldsymbol{x}_0\rangle_{\mathbb{R}^n})$$

and

$$f(\boldsymbol{x}) = \langle \boldsymbol{r}(\mathsf{L}, a_0), \boldsymbol{x}\rangle_{\mathbb{R}^n}$$

for each $\boldsymbol{x} \in U$. Abbreviating

$$u_{\mathsf{L},a_0}(\boldsymbol{x}) = \operatorname{sign}(\langle \boldsymbol{r}(\mathsf{L}, a_0), \boldsymbol{x}\rangle_{\mathbb{R}^n}),$$

we have

$$f(\boldsymbol{x}) = u_{\mathsf{L},j}(\boldsymbol{x}_0)\langle \boldsymbol{r}(\mathsf{L}, a_0), \boldsymbol{x}\rangle_{\mathbb{R}^n}$$

for every $x \in U$. Note that, as in the proofs of parts (iv) and (v) above, $Dg(x)$ has rank 1 for $x \neq 0$. Therefore, by Theorem 1.4.44, there exists $\lambda \in \mathbb{R}$ such that

$$D(f - \lambda g)(x_0) = \mathbf{0}.$$

We compute

$$D(f - \lambda g)(x_0) \cdot v = u_{\mathsf{L},j}(x_0)\langle r(\mathsf{L}, a_0), v\rangle_{\mathbb{R}^n} - 2\lambda\langle x_0, v\rangle_{\mathbb{R}^n}$$

for every $v \in \mathbb{R}^n$. Thus we must have

$$2\lambda x_0 = u_{\mathsf{L},a_0}(x_0) r(\mathsf{L}, a_0).$$

This implies that x_0 and $r(\mathsf{L}, a_0)$ are linearly dependent and that

$$|\lambda| = \frac{1}{2}\|r(\mathsf{L}, a_0)\|_2$$

since $\|x_0\|_2 = 1$. Therefore,

$$f(x_0) = u_{\mathsf{L},a_0}(x_0)\langle r(\mathsf{L}, a_0), \tfrac{1}{2\lambda} u_{\mathsf{L},a_0}(x_0) r(\mathsf{L}, a_0)\rangle_{\mathbb{R}^n} = \frac{2}{\lambda}\lambda^2 = 2\lambda.$$

Since $|\lambda| = \frac{1}{2}\|r(\mathsf{L}, a_0)\|_2$ it follows that

$$f(x_0) = \|r(\mathsf{L}, a_0)\|_2.$$

This completes the proof, but for the fact that maxima of f may occur at points in B_{L}. Thus let $x_0 \in B_{\mathsf{L}}$ be such that $\|x_0\|_2 = 1$. For $a \in A_{\mathsf{L}}(x_0)$ let us write

$$r(\mathsf{L}, a) = \rho_a x_0 + y_a,$$

where $\langle x_0, y_a\rangle_{\mathbb{R}^n} = 0$. Therefore,

$$\langle r(\mathsf{L}, a), x_0\rangle_{\mathbb{R}^n} = \rho_a.$$

We claim that if there exists $a_0 \in A_{\mathsf{L}}(x_0)$ for which $y_{a_0} \neq \mathbf{0}$, then x_0 cannot be a maximum of f subject to the constraint $g^{-1}(0)$. Indeed, if $y_{a_0} \neq 0$ then define

$$x_\epsilon = \frac{x_0 + \epsilon y_{a_0}}{\sqrt{1 + \epsilon^2\|y_{a_0}\|_2^2}}.$$

As in the proof of part (iv) above, one shows that $\|x_\epsilon\|_2 = 1$, and so x_ϵ satisfies the constraint for every $\epsilon \in \mathbb{R}$. Also as in the proof of part (iv), we have

$$x_\epsilon = x_0 + \epsilon y_0 + O(\epsilon^2).$$

Thus

$$\langle r(\mathsf{L}, a_0), x_\epsilon\rangle_{\mathbb{R}^n} = \rho_a + \epsilon\|y_{a_0}\|_2^2 + O(\epsilon^2)$$

and so, for ϵ sufficiently small,

$$|\langle \boldsymbol{r}(\mathsf{L}, a_0), \boldsymbol{x}_\epsilon \rangle_{\mathbb{R}^n}| = |\langle \boldsymbol{r}(\mathsf{L}, a_0), \boldsymbol{x}_0 \rangle_{\mathbb{R}^n}| + \epsilon C_{a_0} + O(\epsilon^2)$$

where C_{a_0} is nonzero. Therefore, there exists $\epsilon_0 \in \mathbb{R}_{>0}$ such that

$$|\langle \boldsymbol{r}(\mathsf{L}, a_0), \boldsymbol{x}_\epsilon \rangle_{\mathbb{R}^n}| > |\langle \boldsymbol{r}(\mathsf{L}, a_0), \boldsymbol{x}_0 \rangle_{\mathbb{R}^n}|$$

either for all $\epsilon \in [-\epsilon_0, 0)$ or for all $\epsilon \in (0, \epsilon_0]$. In either case, $\boldsymbol{x}_0$ cannot be a maximum for f subject to the constraint $g^{-1}(0)$.

Finally, suppose that $\boldsymbol{x}_0 \in B_{\mathsf{L}}$ is a maximum for f subject to the constraint $g^{-1}(0)$. Then, as we saw in the preceding paragraph, for each $a \in A_{\mathsf{L}}(\boldsymbol{x}_0)$, we must have

$$\boldsymbol{r}(\mathsf{L}, a) = \langle \boldsymbol{r}(\mathsf{L}, a), \boldsymbol{x}_0 \rangle_{\mathbb{R}^n} \boldsymbol{x}_0.$$

It follows that $\|\boldsymbol{r}(\mathsf{L}, a)\|_2^2 = \langle \boldsymbol{r}(\mathsf{L}, a), \boldsymbol{x}_0 \rangle_{\mathbb{R}^n}^2$. Moreover, by definition of $A_{\mathsf{L}}(\boldsymbol{x}_0)$ and since we are supposing that $\boldsymbol{x}_0$ is a maximum for f subject to the constraint $g^{-1}(0)$, we have

$$\begin{aligned} & |\langle \boldsymbol{r}(\mathsf{L}, a), \boldsymbol{x}_0 \rangle_{\mathbb{R}^n}| = \|\mathsf{L}\|_{2,\infty} \\ \implies \quad & \langle \boldsymbol{r}(\mathsf{L}, a), \boldsymbol{x}_0 \rangle_{\mathbb{R}^n}^2 = \|\mathsf{L}\|_{2,\infty}^2 \\ \implies \quad & \|\boldsymbol{r}(\mathsf{L}, a)\|_2 = \|\mathsf{L}\|_{2,\infty}. \end{aligned} \tag{1.6}$$

Now, if $a \in \{1, \dots, m\}$, we claim that

$$\|\boldsymbol{r}(\mathsf{L}, a)\|_2 \le \|\mathsf{L}\|_{2,\infty}. \tag{1.7}$$

Indeed suppose that $a \in \{1, \dots, m\}$ satisfies

$$\|\boldsymbol{r}(\mathsf{L}, a)\|_2 > \|\mathsf{L}\|_{2,\infty}.$$

Define $\boldsymbol{x} = \frac{\boldsymbol{r}(\mathsf{L}, a)}{\|\boldsymbol{r}(\mathsf{L}, a)\|_2}$ so that $\boldsymbol{x}$ satisfies the constraint $g(\boldsymbol{x}) = 0$. Moreover,

$$f(\boldsymbol{x}) \ge \langle \boldsymbol{r}(\mathsf{L}, a), \boldsymbol{x} \rangle_{\mathbb{R}^n} = \|\boldsymbol{r}(\mathsf{L}, a)\|_2 > \|\mathsf{L}\|_{2,\infty},$$

contradicting the assumption that $\boldsymbol{x}_0$ is a maximum for f. Thus, given that (1.6) holds for every $a \in A_{\mathsf{L}}(\boldsymbol{x}_0)$ and (1.7) holds for every $a \in \{1, \dots, m\}$, we have

$$\|\mathsf{L}\|_{2,\infty} = \max\{\|\boldsymbol{r}(\mathsf{L}, a)\|_2 \mid a \in \{1, \dots, m\}\},$$

as desired.

For the last three parts of the theorem, the following result is useful.

2 Lemma *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and let $|||\cdot|||_\infty$ be the norm induced on $\mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$ by the norm $\|\cdot\|_\infty$ on $\mathbb{R}^n$ and the norm $\|\cdot\|$ on $\mathbb{R}^m$. Then*

$$|||\mathsf{L}|||_\infty = \max\{\|\mathsf{L}(\mathbf{u})\| \mid \mathbf{u} \in \{-1, 1\}^n\}.$$

Proof Note that the set

$$\{x \in \mathbb{R}^n \mid \|x\|_\infty \leq 1\}$$

is a convex polytope. Therefore, by (1) from the proof of Theorem 1.9.50, this set is the convex hull of $\{-1, 1\}^n$. Thus, if $\|x\|_\infty = 1$ we can write

$$x = \sum_{u \in \{-1,1\}^n} \lambda_u u$$

where $\lambda_u \in [0, 1]$ for each $u \in \{-1, 1\}^n$ and

$$\sum_{u \in \{-1,1\}^n} \lambda_u = 1.$$

Therefore,

$$\begin{aligned} \|\mathsf{L}(x)\| = \Big\| \sum_{u \in \{-1,1\}^n} \lambda_u \mathsf{L}(u) \Big\| &\leq \sum_{u \in \{-1,1\}^n} \lambda_u \|\mathsf{L}(u)\| \\ &\leq \Big(\sum_{u \in \{-1,1\}^n} \lambda_u \Big) \max\{\|\mathsf{L}(u)\| \mid u \in \{-1, 1\}^n\} \\ &= \max\{\|\mathsf{L}(u)\| \mid u \in \{-1, 1\}^n\}. \end{aligned}$$

Therefore,

$$\sup\{\|\mathsf{L}(x)\| \mid \|x\|_\infty = 1\} \leq \max\{\|\mathsf{L}(u)\| \mid u \in \{-1, 1\}^n\} \leq \sup\{\|\mathsf{L}(x)\| \mid \|x\|_\infty = 1\},$$

the last inequality holding since if $u \in \{-1, 1\}^n$ then $\|u\|_\infty = 1$. The result follows since the previous inequalities must be equalities. ▼

(vii) This follows immediately from the preceding lemma.
(viii) This too follows immediately from the preceding lemma.
(ix) Note that for $u \in \{-1, 1\}^n$ we have

$$|\langle r(\mathsf{L}, a), u\rangle_{\mathbb{R}^n}| = \Big|\sum_{j=1}^{n} \mathsf{L}(a, j) u_j\Big| \leq \sum_{j=1}^{n} |\mathsf{L}(a, j)| = \|r(\mathsf{L}, a)\|_1.$$

Therefore, using the previous lemma,

$$\begin{aligned} \|\mathsf{L}\|_{\infty,\infty} &= \max\{\|\mathsf{L}(u)\|_\infty \mid u \in \{-1, 1\}^n\} \\ &= \max\{\max\{|\langle r(\mathsf{L}, a), u\rangle_{\mathbb{R}^n}| \mid a \in \{1, \dots, m\}\} \mid u \in \{-1, 1\}^n\} \\ &\leq \max\{\|r(\mathsf{L}, a)\|_1 \mid a \in \{1, \dots, m\}\}. \end{aligned}$$

To establish the other inequality, for $a \in \{1, \dots, m\}$ define $u_a \in \{-1, 1\}^n$ by

$$u_{a,j} = \begin{cases} 1, & \mathsf{L}(a, j) \geq 0, \\ -1, & \mathsf{L}(a, j) < 0 \end{cases}$$

and note that a direct computation gives the ath component of $\mathsf{L}(\boldsymbol{u}_a)$ as $\|\boldsymbol{r}(\mathsf{L},a)\|_1$. Therefore,

$$\begin{aligned}\max\{\|\boldsymbol{r}(\mathsf{L},a)\|_1 \mid a\in\{1,\dots,m\}\} &= \max\{|\mathsf{L}(\boldsymbol{u}_a)_a| \mid a\in\{1,\dots,m\}\}\\ &\leq \max\{\|\mathsf{L}(\boldsymbol{u}_a)\|_\infty \mid a\in\{1,\dots,m\}\}\\ &\leq \max\{\|\mathsf{L}(\boldsymbol{u})\|_\infty \mid \boldsymbol{u}\in\{-1,1\}^n\} = \|\mathsf{L}\|_{\infty,\infty},\end{aligned}$$

giving this part of the theorem. ■

Having characterised the nine possible norms on $\mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ corresponding to the norms $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_\infty$, we shall always use the norm $\|\cdot\|_{2,2}$, unless explicitly stated to the contrary. And, as we do for the 2-norm for $\mathbb{R}^n$, we will adopt particular notation for the $(2,2)$-norm on $\mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$, denoting it by $\|\cdot\|_{\mathbb{R}^n,\mathbb{R}^m}$.

1.1.5 The Frobenius norm

Next let us consider a different norm for the set of linear maps. First of all, note that there is an identification of $\mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ with $\mathbb{R}^{mn}$. Indeed, there are many such identifications; for example, one could assemble the m rows of $\boldsymbol{A}$, each consisting of n numbers, consecutively to get a vector of length mn. On $\mathbb{R}^{mn}$ one has the Euclidean norm $\|\cdot\|_{\mathbb{R}^{mn}}$, and this then defines a norm on $\mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ using whatever identification one chooses. Moreover, since the Euclidean norm is "unbiased" in terms of the ordering of the indices (i.e., the Euclidean norm of a vector is independent on the order of its components), this norm on $\mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ will be independent of how one chooses to assemble the components of a matrix into a vector of length mn. Thus, waiting for the dust to settle, we have the following definition.

1.1.15 Definition (Frobenius[2] norm) The ***Frobenius norm*** of $\boldsymbol{A}\in\mathrm{Mat}_{m\times n}(\mathbb{R})$ is

$$\|\boldsymbol{A}\|_{\mathrm{Fr}} = (\mathrm{tr}(\boldsymbol{A}^T\boldsymbol{A}))^{1/2}$$ •

Note that, using the definition of transpose, of matrix multiplication, and of trace we have following formula for the Frobenius norm:

$$\|\boldsymbol{A}\|_{\mathrm{Fr}} = \Big(\sum_{a=1}^{m}\sum_{j=1}^{n}\boldsymbol{A}(a,j)^2\Big)^{1/2}.$$

Thus the Frobenius norm is indeed just the square root of the sum of the squares of the components of $\boldsymbol{A}$, just as suggested before the definition.

Let us give some properties of the Frobenius norm, including the assertion that it is indeed a norm.

[2]Ferdinand Georg Frobenius (1849–1917) was a German mathematician whose primary contributions were to the fields of group theory, operator theory, differential geometry, and other.

1.1.16 Proposition (Properties of the Frobenius norm) *If* $\mathbf{A}, \mathbf{A}_1, \mathbf{A}_2 \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$*, if* $\mathbf{B} \in \mathrm{L}(\mathbb{R}^k; \mathbb{R}^n)$*, if* $a \in \mathbb{R}$*, and if* $\mathbf{x} \in \mathbb{R}^n$ *then the following statements hold:*

(i) $\|a\mathbf{A}\|_{\mathrm{Fr}} = |a|\|\mathbf{A}\|_{\mathrm{Fr}}$*;*

(ii) $\|\mathbf{A}\|_{\mathrm{Fr}} \geq 0$*;*

(iii) $\|\mathbf{A}\|_{\mathrm{Fr}} = 0$ *only if* $\mathbf{A} = \mathbf{0}_{m\times n}$*;*

(iv) $\|\mathbf{A}_1 + \mathbf{A}_2\|_{\mathrm{Fr}} \leq \|\mathbf{A}_1\|_{\mathrm{Fr}} + \|\mathbf{A}_2\|_{\mathrm{Fr}}$*;*

(v) $\|\mathbf{A}\mathbf{x}\|_{\mathbb{R}^m} \leq \|\mathbf{A}\|_{\mathrm{Fr}}\|\mathbf{x}\|_{\mathbb{R}^n}$*;*

(vi) $\|\mathbf{A}\mathbf{B}\|_{\mathrm{Fr}} \leq \|\mathbf{A}\|_{\mathrm{Fr}}\|\mathbf{B}\|_{\mathrm{Fr}}$*.*

Proof The first four properties of the Frobenius norm follow from the corresponding properties for the Euclidean norm on $\mathbb{R}^{mn}$. Thus we prove only the last two.

For the fifth property we adopt the notation of Proposition 1.3.16 and compute

$$\|\boldsymbol{A}\boldsymbol{x}\|_{\mathbb{R}^m} = \Big(\sum_{a=1}^{m} \langle \boldsymbol{r}(\boldsymbol{A}, a), \boldsymbol{x}\rangle_{\mathbb{R}^n}^2\Big)^{1/2} \leq \Big(\sum_{a=1}^{m} \|\boldsymbol{r}(\boldsymbol{A}, a)\|_{\mathbb{R}^n}^2 \|\boldsymbol{x}\|_{\mathbb{R}^n}^2\Big)^{1/2}$$
$$= \Big(\sum_{a=1}^{m} \|\boldsymbol{r}(\boldsymbol{A}, a)\|_{\mathbb{R}^n}^2\Big)^{1/2} \|\boldsymbol{x}\|_{\mathbb{R}^n}.$$

The result follows after we notice, and verify via a direct computation, that

$$\|\boldsymbol{A}\|_{\mathrm{Fr}} = \Big(\sum_{a=1}^{m} \|\boldsymbol{r}(\boldsymbol{A}, a)\|_{\mathbb{R}^n}^2\Big)^{1/2}.$$

For the final assertion we first note that

$$\|\boldsymbol{A}\|_{\mathrm{Fr}} = \Big(\sum_{j=1}^{n} \|\boldsymbol{c}(\boldsymbol{A}, j)\|_{\mathbb{R}^m}^2\Big)^{1/2},$$

where, as in Definition I-5.1.4, $\boldsymbol{c}(\boldsymbol{A}, j)$ is the jth column of $\boldsymbol{A}$. Also note that the sth column of $\boldsymbol{A}\boldsymbol{B}$ is given by $\boldsymbol{A}\boldsymbol{c}(\boldsymbol{B}, s)$. Thus we compute

$$\|\boldsymbol{A}\boldsymbol{B}\|_{\mathrm{Fr}} = \Big(\sum_{s=1}^{k} \|\boldsymbol{c}(\boldsymbol{A}\boldsymbol{B}, s)\|_{\mathbb{R}^m}^2\Big)^{1/2} = \Big(\sum_{s=1}^{k} \|\boldsymbol{A}\boldsymbol{c}(\boldsymbol{B}, s)\|_{\mathbb{R}^n}^2\Big)^{1/2}$$
$$\leq \Big(\sum_{s=1}^{k} \|\boldsymbol{A}\|_{\mathrm{Fr}}^2 \|\boldsymbol{c}(\boldsymbol{B}, s)\|_{\mathbb{R}^n}^2\Big)^{1/2} \leq \|\boldsymbol{A}\|_{\mathrm{Fr}} \Big(\sum_{s=1}^{k} \|\boldsymbol{c}(\boldsymbol{B}, s)\|_{\mathbb{R}^n}^2\Big)^{1/2}$$
$$= \|\boldsymbol{A}\|_{\mathrm{Fr}} \|\boldsymbol{B}\|_{\mathrm{Fr}},$$

as desired, and where we have used the result from the previous part. ■

It is natural to ask whether the Frobenius norm is the induced norm for some pair of norms, one on $\mathbb{R}^n$ and one on $\mathbb{R}^m$.

1.1.17 Proposition (The Frobenius norm is not often induced) *If* $\mathrm{m}, \mathrm{n} \in \mathbb{Z}_{>0}$, *then the Frobenius norm on* $\mathrm{L}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{m}})$ *is the induced norm for any pair of norms, one on* $\mathbb{R}^{\mathrm{n}}$ *and the other on* $\mathbb{R}^{\mathrm{m}}$, *if and only if* m *or* n *are equal to* 1.

Proof If $\|\cdot\|$ is a norm on $\mathbb{R}^n$, then let us define a norm $\|\cdot\|^*$ on $\mathbb{R}^n$ by

$$\|\boldsymbol{x}\|^* = \sup\{|\langle \boldsymbol{x}, \boldsymbol{v}\rangle_{\mathbb{R}^n}| \mid \|\boldsymbol{v}\| = 1\}.$$

It is easy to verify $\|\cdot\|^*$ is indeed a norm. Moreover, it is easy to verify that $\|\cdot\|^{**} = \|\cdot\|$.

Let us give a few lemmata that we will use in the proof. For the following lemma, if $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$ then $\boldsymbol{y}\boldsymbol{x}^T$ denotes the linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ defined by

$$\boldsymbol{y}\boldsymbol{x}^T(\xi) = \langle \boldsymbol{x}, \xi\rangle_{\mathbb{R}^n}\boldsymbol{y}.$$

It is evident that $\operatorname{rank}(\boldsymbol{y}\boldsymbol{x}^T) = 1$.

1 Lemma *Let* $\|\cdot\|_\alpha$ *and* $\|\cdot\|_\beta$ *be norms on* $\mathbb{R}^{\mathrm{n}}$ *and* $\mathbb{R}^{\mathrm{m}}$, *respectively, and let* $\|\cdot\|_{\alpha,\beta}$ *be the induced norm on* $\mathrm{L}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{m}})$. *Then*

$$\|\mathbf{y}\mathbf{x}^{\mathrm{T}}\|_{\alpha,\beta} = \|\mathbf{x}\|_\alpha^*\|\mathbf{y}\|_\beta$$

for every $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$ *and* $\mathbf{y} \in \mathbb{R}^{\mathrm{m}}$.

Proof We compute

$$\|\boldsymbol{y}\boldsymbol{x}^T\|_{\alpha,\beta} = \sup\{\|\boldsymbol{y}\boldsymbol{x}^T(\boldsymbol{v})\|_\beta \mid \|\boldsymbol{v}\|_\alpha = 1\} = \sup\{|\langle \boldsymbol{x}, \boldsymbol{v}\rangle_{\mathbb{R}^n}|\|\boldsymbol{y}\|_\beta \mid \|\boldsymbol{v}\|_\alpha\} = \|\boldsymbol{x}\|_\alpha^*\|\boldsymbol{y}\|_\beta \qquad ▼$$

For the following lemma, we refer ahead to Definition 1.3.19 for the notion of
what an orthogonal matrix, or equivalently linear map. We also recall from the notion of singular values for a linear map between inner product spaces.

2 Lemma *If* $\|\cdot\|$ *is a norm on* $\mathrm{L}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{m}})$ *such that*

$$\|\mathsf{U} \circ \mathsf{L}\mathsf{V}\| = \|\mathsf{L}\|$$

for every $\mathsf{U} \in \mathsf{O}(\mathrm{m})$ *and every* $\mathsf{V} \in \mathsf{O}(\mathrm{n})$, *then there exists* $\mathrm{c} \in \mathbb{R}_{>0}$ *such that, if* $\mathsf{L} \in \mathrm{L}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{m}})$ *has rank* 1, *it holds that* $\|\mathsf{L}\| = \mathrm{c}\sigma_{\max}(\mathsf{L})$.

Proof Let us denote by $\mathsf{L}_{11} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$ the linear map defined by

$$\mathsf{L}(x_1, \dots, x_n) = (x_1, 0, \dots, 0).$$

what As we show in , if L has rank 1, then there exists $\mathsf{U} \in \mathsf{O}(m)$ and $\mathsf{V} \in \mathsf{O}(n)$ such that $\mathsf{L} = \sigma_{\max}(\mathsf{L})\mathsf{U} \circ \mathsf{L}_{11} \circ \mathsf{V}$. It therefore follows that if L has rank 1 then $\|\mathsf{L}\| = \sigma_{\max}\|\mathsf{L}_{11}\|$, giving the result by taking $c = \|\mathsf{L}_{11}\|$. ▼

Now the following lemma is key.

3 Lemma *Let $\|\cdot\|$ be a norm on $\mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ satisfy*

$$\|\mathsf{U} \circ \mathsf{L}\mathsf{V}\| = \|\mathsf{L}\|$$

for every $\mathsf{U} \in \mathsf{O}(m)$ and every $\mathsf{V} \in \mathsf{O}(n)$. Then the following statements are equivalent:

(i) there exist norms $\|\cdot\|_\alpha$ on $\mathbb{R}^n$ and $\|\cdot\|_\beta$ on $\mathbb{R}^m$ such that $\|\cdot\|$ is the corresponding induced norm;

(ii) there exists $c \in \mathbb{R}_{>0}$ such that $\|\mathsf{L}\| = c\sigma_{\max}(\mathsf{L})$ for every $\mathsf{L} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$.

Proof From Theorem 1.1.14(v), the norm on $\mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ induced by the norm $\|\cdot\|_2$ on $\mathbb{R}^n$ and $c\|\cdot\|_2$ satisfies $\|\mathsf{L}\| = c\sigma_{\max}(\mathsf{L})$ for every $\mathsf{L} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$. Moreover, since

$$\sigma_{\max}(\mathsf{U} \circ \mathsf{L} \circ \mathsf{V}) = \sigma_{\max}(\mathsf{L})$$

for every $\mathsf{U} \in \mathsf{O}(m)$ and $\mathsf{V} \in \mathsf{O}(n)$, we arrive at the implication (ii) $\Longrightarrow$ (i).

For the converse implication, suppose that $\|\cdot\|$ is induced by $\|\cdot\|_\alpha$ and $\|\cdot\|_\beta$ on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively. By Lemma 2 there exists $c \in \mathbb{R}_{>0}$ such that $\|\mathsf{L}\| = c\sigma_{\max}(\mathsf{L})$ for every $\mathsf{L} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ having rank 1. From Lemma 1 and we also have singular value of yx^T

$$c\|\boldsymbol{x}\|_2\|\boldsymbol{y}\|_2 = c\sigma_{\max}(\boldsymbol{y}\boldsymbol{x}^T) = \|\boldsymbol{x}\|_\alpha^*\|\boldsymbol{y}\|_\beta$$

for every $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$. By fixing $\boldsymbol{y} \in \mathbb{R}^m$ we see that there exists $c_1 \in \mathbb{R}_{>0}$ such that $\|\boldsymbol{x}\|_\alpha^* = c_1\|\boldsymbol{x}\|_2$ for every $\boldsymbol{x} \in \mathbb{R}^n$. Similarly, by fixing $\boldsymbol{x}$ there exists $c_2 \in \mathbb{R}_{>0}$ such that $\|\boldsymbol{y}\|_\beta = c_2\|\boldsymbol{y}\|_2$ for every $\boldsymbol{y} \in \mathbb{R}^m$. Since $\|\cdot\|_\alpha^{**} = \|\cdot\|\alpha$ and since $\|\cdot\|_2^* = \|\cdot\|_2$ (verify this), we conclude that $\|\cdot\|_\alpha = c_2\|\cdot\|_2$. From Theorem 1.1.14(v) we conclude that $\|\mathsf{L}\| = \frac{c_2}{c_1}\sigma_{\max}(\mathsf{L})$, giving the lemma. ▼

Now we prove the proposition. First of all, note that if $n = 1$ or if $m = 1$, then $\|\cdot\|_{\mathrm{Fr}} = \|\cdot\|_{2,2}$ by Theorem 1.1.14. Conversely, suppose that neither n nor m is equal to 1. For $a \in \mathbb{R}_{>0}$ define $\mathsf{L}_a \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^m)$ by

$$\mathsf{L}_a(x_1, x_2, x_3, \dots, x_n) = (x_1, ax_2, 0, \dots, 0).$$

Note that $\sigma_{\max}(\mathsf{L}_a) = \max\{1, a\}$. However, $\|\mathsf{L}_a\|_{\mathrm{Fr}} = \sqrt{1 + a^2}$. Thus we cannot have $\|(\|_{\mathrm{Fr}}\mathsf{L}_a) = c\sigma_{\max}(\mathsf{L}_a)$ for every $a \in \mathbb{R}_{>0}$. By Lemma 3 the theorem follows. ■

1.1.6 Notes

Some parts of the proof we give of Theorem 1.1.14 are new, although much of the result is classically known; see [Horn and Johnson 2013]. The proof of part (iv) of Theorem 1.1.14 comes from [Drakakis and Pearlmutter 2009]. The proof of part (vii) of Theorem 1.1.14 comes from [Rohn 2000]. Note that there is a somewhat different character in certain of the induced norm computations in Theorem 1.1.14. In particular, the induced norms $\|\cdot\|_{2,1}$, $\|\cdot\|_{\infty,1}$, and $\|\cdot\|_{\infty,2}$ involve a search over the 2^m points in $\{-1, 1\}^m$ (in the first two cases) or the 2^n points in $\{-1, 1\}^n$ in the third case. The computations of these norms is correspondingly more involved in terms of the numbers of computations that must be performed. This is discussed by Rohn [2000] for the norm $\|\cdot\|_{\infty,1}$.

The proof we give of Proposition 1.1.17 follows [Chellaboina and Haddad 1995].

Exercises

1.1.1 Show that $S^\perp$ is a subspace of $\mathbb{R}^n$ for every nonempty subset $S \subseteq \mathbb{R}^n$.

1.1.2 Let $r_1, r_2 \in \mathbb{R}_{>0}$ satisfy $r_2 \leq r_1$ and let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$. Show that if $\overline{\mathsf{B}}(r_1, \boldsymbol{x}_1) \cap \overline{\mathsf{B}}(r_2, \boldsymbol{x}_2) \neq \emptyset$ then $\overline{\mathsf{B}}(r_2, \boldsymbol{x}_2) \subseteq \overline{\mathsf{B}}(3r_1, \boldsymbol{x}_1)$. Show that you understand your proof by drawing a picture.

1.1.3 Show that for each $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$, $\left| \|\boldsymbol{x}_1\|_{\mathbb{R}^n} - \|\boldsymbol{x}_2\|_{\mathbb{R}^n} \right| \leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathbb{R}^n}$.

Section 1.2

The structure of $\mathbb{R}^n$

In this section we summarise the topological (see Chapter III-1) properties of $\mathbb{R}^n$. Many of the properties here are discussed in a more general context in Chapter III-3. Therefore, we limit ourselves here to those features of $\mathbb{R}^n$ that we will make use of without needing the abstract development of Chapter III-3. For example, some of what we do here will be used in Chapter 3. Because some of what we say here bears a strong resemblance to some of the results of Chapter I-2, and because we shall generalise much of this structure in Chapter III-3, we shall omit some of the proofs that resemble their counterparts of Chapters I-2 and III-3.

Do I need to read this section? Much of what we say in this section follows in the same vein as does much of Chapter I-2. Therefore, perhaps a reader can overlook some of the details of what we say here until specific parts of it are needed. •

1.2.1 Sequences in $\mathbb{R}^n$

Note that for $\mathbb{R}$ the discussion of sequences and their convergence is reliant on the absolute value function. Since this can be generalised to $\mathbb{R}^n$, the ideas of Cauchy sequences and convergent sequences carries over to $\mathbb{R}^n$. Let us give the definitions in this case.

1.2.1 Definition (Cauchy sequence, convergent sequence, bounded sequence) Let $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}^n$. The sequence:

(i) is a ***Cauchy sequence*** if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $\|\boldsymbol{x}_j - \boldsymbol{x}_k\|_{\mathbb{R}^n} < \epsilon$ for $j, k \geq N$;

(ii) ***converges to*** $\mathbf{x}_0$ if, for each $\epsilon \in \mathbb{R}_{>0}$, there exists $N \in \mathbb{Z}_{>0}$ such that $\|\boldsymbol{x}_j - \boldsymbol{x}_0\|_{\mathbb{R}^n} < \epsilon$ for $j \geq N$;

(iii) ***diverges*** if it does not converge to any element in $\mathbb{R}^n$;

(iv) is ***bounded*** if there exists $M \in \mathbb{R}_{>0}$ such that $\|\boldsymbol{x}_j\|_{\mathbb{R}^n} < M$ for each $j \in \mathbb{Z}_{>0}$;

(v) is ***constant*** if $\boldsymbol{x}_j = \boldsymbol{x}_1$ for every $j \in \mathbb{Z}_{>0}$;

(vi) is ***eventually constant*** if there exists $N \in \mathbb{Z}_{>0}$ such that $\boldsymbol{x}_j = \boldsymbol{x}_N$ for every $j \geq N$. •

One can show, just as for sequences of real numbers, that convergent sequences are Cauchy and that Cauchy sequences are bounded. Let us state these results here.

1.2.2 Proposition (Convergent sequences are Cauchy) *If a sequence $(\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $\mathbf{x}_0$ then it is Cauchy.*

Proof Let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $\|\boldsymbol{x}_j - \boldsymbol{x}_0\|_{\mathbb{R}^n} < \frac{\epsilon}{2}$ for $j \geq N$. Then, for $j,k \geq N$ we have

$$\|\boldsymbol{x}_j - \boldsymbol{x}_k\|_{\mathbb{R}^n} \leq \|\boldsymbol{x}_j - \boldsymbol{x}_0\|_{\mathbb{R}^n} + \|\boldsymbol{x}_0 - \boldsymbol{x}_k\|_{\mathbb{R}^n} < \tfrac{\epsilon}{2} + \tfrac{\epsilon}{2} = \epsilon,$$

as desired. ■

1.2.3 Proposition (Cauchy sequences are bounded) *If $(\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequences then it is bounded.*

Proof Let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $\|\boldsymbol{x}_j - \boldsymbol{x}_k\|_{\mathbb{R}^n} < 1$ for $j,k \geq N$. Let $M_N = \max\{\|\boldsymbol{x}_1\|_{\mathbb{R}^n}, \dots, \|\boldsymbol{x}_N\|_{\mathbb{R}^n}\}$. For $j \geq N$ we have

$$|\boldsymbol{x}_j| \leq \|\boldsymbol{x}_j - \boldsymbol{x}_N\|_{\mathbb{R}^n} + \|\boldsymbol{x}_N\|_{\mathbb{R}^n} < 1 + M_N,$$

showing that $\|\boldsymbol{x}_j\|_{\mathbb{R}^n} < 1 + M_N$ for each $j \in \mathbb{Z}_{>0}$. ■

The following result indicates that, to show the convergence of a sequence in $\mathbb{R}^n$, it suffices to show the convergence of the sequence of components.

1.2.4 Proposition (Convergence of a sequence in $\mathbb{R}^n$ equals convergence of each of the components) *Let $(\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}^n$ and denote $\mathbf{x}_j = (x_j^1, \dots, x_j^n)$, $j \in \mathbb{Z}_{>0}$. Then the sequence $(\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $\mathbf{x}_0 = (x_0^1, \dots, x_0^n)$ if and only if each of the sequences $(x_j^l)_{j\in\mathbb{Z}_{>0}}$, $l \in \{1, \dots, n\}$, converges to x_0^l.*

Proof Suppose that $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $\boldsymbol{x}_0$. For $\epsilon \in \mathbb{R}_{>0}$ let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $\|\boldsymbol{x}_j - \boldsymbol{x}_0\|_{\mathbb{R}^n} < \epsilon$ for $j \geq N$. Then

$$\left|x_j^l - x_0^l\right| \leq \Big(\sum_{m=1}^n (x_j^m - x_0^m)^2\Big)^{1/2} = \|\boldsymbol{x}_j - \boldsymbol{x}_0\|_{\mathbb{R}^n} < \epsilon,$$

showing that $(x_j^l)_{j\in\mathbb{Z}_{>0}}$ converges to x_0^l.

Now suppose that $(x_j^l)_{j\in\mathbb{Z}_{>0}}$ converges to x_0^l for $l \in \{1, \dots, n\}$. Let $\epsilon \in \mathbb{R}_{>0}$ and let N be sufficiently large that $\left|x_j^m - x^m\right| < \frac{\epsilon}{\sqrt{n}}$ for $j \geq N$ and for $m \in \{1, \dots, n\}$. Then

$$\|\boldsymbol{x}_j - \boldsymbol{x}_0\|_{\mathbb{R}^n} = \Big(\sum_{m=1}^n (x_j^m - x^m)^2\Big)^{1/2} < \Big(\sum_{m=1}^n \frac{\epsilon^2}{n}\Big)^{1/2} = \epsilon,$$

as desired. ■

Thus the convergence tests for sequences in Section I-2.3.3 can be used to prove convergence of sequences in $\mathbb{R}^n$ by applying them componentwise.

It is also true that Cauchy sequences converge in $\mathbb{R}^n$. As we see in the proof of the following result, this is reliant on the completeness of $\mathbb{R}$. This notion of completeness is explored in detail in more generality in Section III-3.3.

1.2.5 Theorem (Cauchy sequences in $\mathbb{R}^n$ converge) *If $(\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathbb{R}^n$ then it converges.*

Proof Let $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ be a Cauchy sequence in $\mathbb{R}^n$; we write $\boldsymbol{x}_j = (x_j^1, \dots, x_j^n)$, $j \in \mathbb{Z}_{>0}$. We claim that $(x_j^l)_{j\in\mathbb{Z}_{>0}}$ is a Cauchy sequence in $\mathbb{R}$ for $l \in \{1, \dots, n\}$. Indeed, for $\epsilon \in \mathbb{R}_{>0}$ let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $\|\boldsymbol{x}_j - \boldsymbol{x}_k\|_{\mathbb{R}^n} < \epsilon$ for $j, k \geq N$. Then

$$\left|x_j^l - x_k^l\right| \leq \Big(\sum_{m=1}^n (x_j^m - x_k^m)^2\Big)^{1/2} = \|\boldsymbol{x}_j - \boldsymbol{x}_k\|_{\mathbb{R}^n} < \epsilon$$

for all $l \in \{1, \dots, n\}$ and $j, k \geq N$. By Theorem I-2.3.5 there exists $x^l \in \mathbb{R}$ to which the sequence $(x_j^l)_{j\in\mathbb{Z}_{>0}}$ converges. By Proposition 1.2.4 it follows that $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $(x^1, \dots, x^l)$. ■

It is also possible to discuss convergence of multiple sequences in $\mathbb{R}^n$. The definitions and results are just like those in Section I-2.3.5 for multiple sequences in $\mathbb{R}$. Multiple sequences are also discussed in Section III-3.2.3 in a more general context. The reader who wants to use multiple sequences in $\mathbb{R}^n$, and is somehow unable to extrapolate from the results of Section I-2.3.5 will find the appropriate definitions in this more general setting.

It is useful to know the relationship between limits and algebraic operations.

1.2.6 Proposition (Algebraic operations on sequences) *Let $(\mathbf{x}_j)_{j\in\mathbb{Z}}$ and $(\mathbf{y}_j)_{j\in\mathbb{Z}_{>0}}$ be sequences in $\mathbb{R}^n$ converging to $\mathbf{x}_0$ and $\mathbf{y}_0$, respectively, let $(a_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}$ converging to a_0, and let $a \in \mathbb{R}$. Then the following statements hold:*

(i) *the sequence $(a\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $a\mathbf{x}_0$;*

(ii) *the sequence $(\mathbf{x}_j + \mathbf{y}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $\mathbf{x}_0 + \mathbf{y}_0$;*

(iii) *the sequence $(a_j\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $a_0\mathbf{x}_0$.*

Proof This proof will be given in a more general context, but with essentially identical notation, for Proposition III-3.2.6. The proof is also quite similar to the proof for Proposition I-2.3.23. Thus we forgo giving the details here. ■

1.2.2 Series in $\mathbb{R}^n$

The extension of series of real numbers to series in $\mathbb{R}^n$ is fairly easily achieved. One begins by considering a ***series*** in $\mathbb{R}^n$ to bean expression of the form

$$\sum_{j=1}^\infty \boldsymbol{x}_j,$$

where $\boldsymbol{x}_j \in \mathbb{R}^n$, $j \in \mathbb{Z}_{>0}$. As we discussed at the beginning of Section I-2.4.1, one needs to interpret this expression carefully as it is meaningless as a sum until one says something about its convergence. However, as a formal expression involving the elements of the sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}}$ it is sensible, and the summation sign is just a convenience to indicate in what we are interested.

Let us define the sorts of convergence one can consider for series.

1.2.7 Definition (Convergence and absolute convergence of series) Let $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}^n$ and consider the series

$$S = \sum_{j=1}^{\infty} \boldsymbol{x}_j.$$

The corresponding sequence of ***partial sums*** is the sequence $(S_k)_{k\in\mathbb{Z}_{>0}}$ defined by

$$S_k = \sum_{j=1}^{k} \boldsymbol{x}_j.$$

Let $\boldsymbol{x}_0 \in \mathbb{R}^n$. The series:

(i) ***converges to*** $\mathbf{x}_0$, and we write $\sum_{j=1}^{\infty} \boldsymbol{x}_j = \boldsymbol{x}_0$, if the sequence of partial sums converges to $\boldsymbol{x}_0$;

(ii) has $\boldsymbol{x}_0$ as a ***limit*** if it converges to $\boldsymbol{x}_0$;

(iii) is ***convergent*** if it converges to some member of $\mathbb{R}^n$;

(iv) ***converges absolutely***, or is ***absolutely convergent***, if the series

$$\sum_{j=1}^{\infty} \|\boldsymbol{x}_j\|_{\mathbb{R}^n}$$

converges;

(v) ***converges conditionally***, or is ***conditionally convergent***, if it is convergent, but not absolutely convergent;

(vi) ***diverges*** if it does not converge;

(vii) has a limit that ***exists*** if $\lim_{j\to\infty} S_j \in \mathbb{R}^n$. •

We have the following correspondence between convergence and absolute convergence.

1.2.8 Proposition (Absolutely convergent series are convergent) *If a series* $\sum_{j=1}^{\infty} \mathrm{x_j}$ *is absolutely convergent, then it is convergent.*

Proof Let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be such that

$$\sum_{j=N}^{\infty} \|\boldsymbol{x}_j\|_{\mathbb{R}^n} < \epsilon;$$

this is possible by absolute convergence (why?). Let $k, l \geq N$ with $l > k$ and compute

$$\Big\| \sum_{j=k+1}^{l} \boldsymbol{x}_j \Big\| \leq \sum_{j=l+1}^{k} \|\boldsymbol{x}_j\|_{\mathbb{R}^n} \leq \sum_{j=N}^{\infty} \|\boldsymbol{x}_j\|_{\mathbb{R}^n} < \epsilon,$$

showing that the sequence of partial sums is Cauchy. By Theorem 1.2.5 it follows that the sequence is convergent. ■

The importance of the concept of absolute convergence is perhaps not perfectly clear at a first glance. One of the reasons it is important is that absolutely convergent series have the property that if you reorder their terms in an arbitrary way, the resulting series still converges and converges to the same limit. This is shown for real series in Theorem I-2.4.5 and is explored in detail in a more general setting in Section III-3.4.2.

The following property of absolutely convergent series is often important,

1.2.9 Proposition (Swapping summation and norm) *For a sequence* $(\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$*, if the series* $\mathbf{S} = \sum_{j=1}^{\infty} \mathbf{x}_j$ *is absolutely convergent, then*

$$\Big\|\sum_{j=1}^{\infty} \mathbf{x}_j\Big\|_{\mathbb{R}^n} \leq \sum_{j=1}^{\infty} \|\mathbf{x}_j\|_{\mathbb{R}^n}.$$

Proof Define

$$S_m^1 = \Big\|\sum_{j=1}^{m} \boldsymbol{x}_j\Big\|_{\mathbb{R}^n}, \quad S_m^2 = \sum_{j=1}^{m} \|\boldsymbol{x}_j\|_{\mathbb{R}^n}, \qquad m \in \mathbb{Z}_{>0}.$$

By Exercise 1.2.1 we have $S_m^1 \leq S_m^2$ for each $m \in \mathbb{Z}_{>0}$. Moreover, by Proposition 1.2.8 and Theorem 1.2.5 the sequences $(S_m^1)_{m\in\mathbb{Z}_{>0}}$ and $(S_m^2)_{m\in\mathbb{Z}_{>0}}$ are Cauchy sequences in $\mathbb{R}^n$ 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. ∎

One can also talk about multiple series in $\mathbb{R}^n$. The definitions are just like those in Section I-2.4.5 for multiple series in $\mathbb{R}$. We shall also give these definitions in a more general setting in Section III-3.4.4, so the reader can refer ahead if need be.

We can also give results analogous to those in Section I-2.3.6 for series in $\mathbb{R}$. First we give some notation for products of series.

1.2.10 Definition (Scalar multiplication of series) Let $S = \sum_{j=0}^{\infty} \boldsymbol{x}_j$ be a series in $\mathbb{R}^n$ and let $s = \sum_{j=0}^{\infty} a_j$ be series in $\mathbb{R}$.

(i) The ***product*** of s and $\boldsymbol{S}$ is the double series $\sum_{j,k=0}^{\infty} a_j \boldsymbol{v}_k$.

(ii) The ***Cauchy product*** of s and $\boldsymbol{S}$ is the series $\sum_{k=0}^{\infty}\Big(\sum_{j=0}^{k} a_j \boldsymbol{v}_{k-j}\Big)$. •

Now we can state the interaction between convergence of series and the vector space operations.

1.2.11 Proposition (Algebraic operations on series) *Let* $\mathbf{S} = \sum_{j=0}^{\infty} \mathbf{x}_j$ *and* $\mathbf{T} = \sum_{j=0}^{\infty} \mathbf{y}_j$ *be series in* $\mathbb{R}^n$ *converging to* $\mathbf{X}_0$ *and* $\mathbf{Y}_0$*, respectively, let* $s = \sum_{j=0}^{\infty} a_j$ *be a series in* $\mathbb{F}$ *converging to* A_0*, and let* $a \in \mathbb{F}$*. Then the following statements hold:*

(i) the series $\sum_{j=0}^{\infty} a\mathbf{x}_j$ *converges to* $a\mathbf{X}_0$*;*

(ii) the series $\sum_{j=0}^{\infty}(\mathbf{x}_j + \mathbf{y}_j)$ *converges to* $\mathbf{X}_0 + \mathbf{Y}_0$*;*

(iii) *if* s *and* **S** *are absolutely convergent, then the product of* s *and* **S** *is absolutely convergent and converges to* $\mathrm{A}_0\mathbf{X}_0$*;*

(iv) *if* s *and* **S** *are absolutely convergent, then the Cauchy product of* s *and* **S** *is absolutely convergent and converges to* $\mathrm{A}_0\mathbf{X}_0$*;*

(v) *if* s *or* **S** *are absolutely convergent, then the Cauchy product of* s *and* **S** *is convergent and converges to* $\mathrm{A}_0\mathbf{X}_0$*.*

Proof The proof is identical, except for slight notational changes, to that for Proposition III-3.4.10. It also bears a resemblance to the proof of Proposition I-2.4.30. Thus we do not repeat the proof here. ■

1.2.3 Open and closed balls, rectangles

Note that the definition of open (and therefore closed) sets in $\mathbb{R}$ relies on the absolute value function. Therefore, since the absolute value function has an appropriate generalisation to $\mathbb{R}^n$ as the Euclidean norm, the ideas of open and closed sets carry over to $\mathbb{R}^n$. The key idea is the generalisation of the notion of an open ball as seen in Example III-1.1.2–3. Here we simply make the following definition.

1.2.12 Definition (Open ball, closed ball) Let $\boldsymbol{x}_0 \in \mathbb{R}^n$ and let $r \in \mathbb{R}_{\geq 0}$.

(i) The ***open ball*** centred at $\boldsymbol{x}_0$ of radius r is the set

$$\mathsf{B}^n(r, \boldsymbol{x}_0) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n} < r\}.$$

(ii) The ***closed ball*** centred at $\boldsymbol{x}_0$ of radius r is the set

$$\overline{\mathsf{B}}^n(r, \boldsymbol{x}_0) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n} \leq r\}.$$ •

For example, in the case when $n = 1$, we have

$$\mathsf{B}^1(r, x_0) = (x_0 - r, x_0 + r), \quad \overline{\mathsf{B}}^1(r, x_0) = [x_0 - r, x_0 + r].$$

Thus open and closed balls can be thought of as generalisations of open and closed intervals. In Figure 1.2 we depict how one should think of open and closed balls.

1.2.13 Notation ("Balls" versus "spheres") Note that we have defined a ball of radius r as containing all points that are a distance at most r from the centre. It is also interesting to talk about the points that are a distance *exactly* r from the centre. Thus we define

$$\mathbb{S}(r, \boldsymbol{x}_0) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_{\mathbb{R}^n} = r\},$$

which is the ***sphere*** of radius r and centre $\boldsymbol{x}_0$. In common language, "sphere" is often used where we mean "ball." The reader should be aware of our precise convention as we will never violate it, even casually. •

Another natural generalisation of an interval is the following.

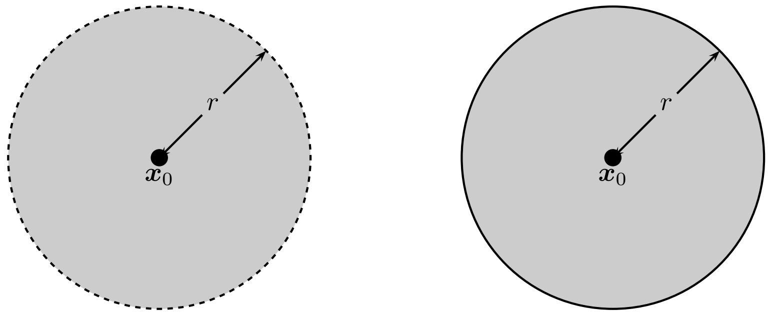

Figure 1.2 Open (left) and closed (right) balls in $\mathbb{R}^n$

1.2.14 Definition (Rectangle, cube) A ***rectangle*** in $\mathbb{R}^n$ is a subset of the form

$$R = I_1 \times \cdots \times I_n$$

where $I_1, \dots, I_n \subseteq \mathbb{R}$ are intervals. A rectangle $R = I_1 \times \cdots \times I_n$ is ***fat*** if $\mathrm{int}(I_j) \neq \emptyset$ for each $j \in \mathbb{Z}_{>0}$. If each of the intervals $I_1, \dots, I_+n$ is bounded and has the same length, the resulting rectangle is called a ***cube***. •

A rectangle is, somehow, a more faithful generalisation of the notion of an interval, it being a product of intervals. Both balls (as we have defined then) and rectangles can serve as the building blocks for what we do in the remainder of this section. This is made precise only after one knows a little about topology and norm topologies; we refer to Section III-3.1 for more details. For now we simply stick to using balls to define many of the useful structural properties of $\mathbb{R}^n$.

However, since we will use rectangles in Section 1.6 to define the Riemann integral, let us engage in a discussion of some useful constructions involving rectangles. These are direct generalisations of corresponding notions for intervals.

1.2.15 Definition (Partition of a compact rectangle) If

$$R = [a_1, b_1] \times \cdots \times [a_n, b_n],$$

with $a_j < b_j$, $j \in \{1, \dots, n\}$ is a fat compact rectangle, a ***partition*** of R is an n-tuple $\boldsymbol{P} = (P_1, \dots, P_n)$ where $P_j = (I_{j1}, \dots, I_{jk_j})$ is a partition of the interval $[a_j, b_j]$, $j \in \{1, \dots, n\}$. The rectangles

$$R_{l_1,\dots,l_n} = I_{1l_1} \times \cdots \times I_{nl_n}, \qquad l_j \in \{1, \dots, k_j\},\ j \in \{1, \dots, n\},$$

are the ***subrectangles*** of the partition. •

Thus the partition is applied to each of the coordinate axes of the rectangle R. In Figure 1.3 we depict a partition of a two-dimensional rectangle. Note that

$$R = \mathop{\overset{\circ}{\bigcup}}_{\substack{l_j \in \{1,\dots,k_j\} \\ j \in \{1,\dots,n\}}} R_{l_1,\dots,l_n}.$$

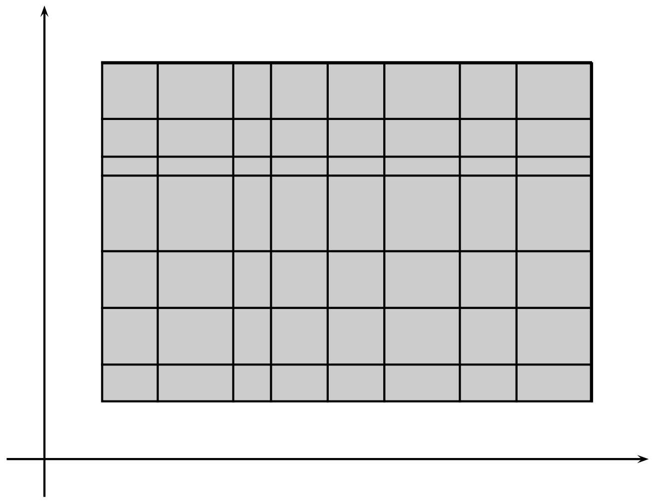

Figure 1.3 A partition of a two-dimensional rectangle

As with a partition of an interval we can define a "length" of a partition $\boldsymbol{P} = (P_1, \dots, P_n)$. We suppose that $\mathrm{EP}_j = (x_{j0}, \dots, x_{jk_j})$ and then define

$$|\boldsymbol{P}| = \min\{|x_{jl} - x_{jm}| \mid j \in \{1, \dots, n\},\ l \in \{1, \dots, k_j\}\}.$$

Thus $|\boldsymbol{P}|$ is the length of the smallest side of each of the rectangles whose union is R.

It is also possible to say when one partition is contained in another.

1.2.16 Definition (Refinement of a partition) Let $R \subseteq \mathbb{R}^n$ be a fat rectangle and let $\boldsymbol{P} = (P_1, \dots, P_n)$ and $\boldsymbol{P}' = (P'_1, \dots, P'_n)$ be partitions of R. Then $\boldsymbol{P}'$ is a ***refinement*** of $\boldsymbol{P}$ if P'_j is a refinement of P_j for each $j \in \{1, \dots, n\}$. •

The idea is that each of the rectangles from $\boldsymbol{P}'$ is a subset of a rectangle from $\boldsymbol{P}$.

1.2.4 Open and closed subsets

We now use open balls to define the notion of open and closed subsets of $\mathbb{R}^n$, just as we used intervals in Section I-2.5.1 to define open and closed subsets of $\mathbb{R}$.

1.2.17 Definition (Open and closed sets in $\mathbb{R}^n$) A subset $A \subseteq \mathbb{R}^n$

(i) is ***open*** if, for every $\boldsymbol{x} \in A$, there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\epsilon, \boldsymbol{x}) \subseteq A$ and

(ii) is ***closed*** if $\mathbb{R}^n \setminus A$ is open. •

1.2.18 Remark (Use of the words "topology" and "topological") We shall on occasion, and sometimes more frequently than that, make use of words like "topology" and "topological" in our discussion, although we will not formally introduce such terminology until Chapter III-1. The way to read our use of such words is this: They refer to things broadly related to the use of open subsets of $\mathbb{R}^n$. As we shall see, almost everything we shall say in this chapter depends in some way on open sets, their definition, and their properties. This is exactly what the study of topology consists of. •

The following properties of open and closed sets arise in the general presentation of topological spaces in Chapter III-1.

1.2.19 Proposition (Properties of open and closed sets) *For an arbitrary collection* $(\mathrm{U}_a)_{a\in A}$ *of open sets and an arbitrary collection* $(\mathrm{C}_b)_{b\in B}$ *of closed sets the following statements hold:*

(i) $\cup_{a\in A}\mathrm{U}_a$ *is open;*

(ii) $\cap_{b\in B}\mathrm{C}_b$ *is closed.*

Moreover, for open sets U_1 *and* U_2 *and closed sets* C_1 *and* C_2*, the following statements hold:*

(iii) $\mathrm{U}_1 \cap \mathrm{U}_2$ *is open;*

(iv) $\mathrm{C}_1 \cup \mathrm{C}_2$ *is closed.*

Proof This is Exercise 1.2.3. ■

As with open subsets of $\mathbb{R}$ the language "neighbourhood" is often useful.

1.2.20 Definition (Neighbourhood) A ***neighbourhood*** of $x \in \mathbb{R}^n$ is an open set U for which $x \in U$. More generally, a ***neighbourhood*** of a subset $A \subseteq \mathbb{R}^n$ is an open set U for which $A \subseteq U$. •

Many of the properties of open sets in $\mathbb{R}$ also hold for open subsets of $\mathbb{R}^n$.

1.2.21 Proposition (Open subsets of $\mathbb{R}^n$ are unions of open balls) *If* $\mathrm{U} \subseteq \mathbb{R}^n$ *is a nonempty open set then* U *is a countable union of open balls.*

Proof Let $x \in U$ so that there exists $r_x \in \mathbb{R}_{>0}$ for which $\mathsf{B}^n(r_x, x) \subseteq U$. By Proposition I-2.2.15 there exists $q_x \in \mathbb{Q}_{>0}$ such that $q_x < r_x$. Therefore, $\mathsf{B}^n(q_x, x) \subseteq U$. Also by Proposition I-2.2.15 there exists $\boldsymbol{q}_x \in \mathbb{R}^n$ with rational components such that $\|x - \boldsymbol{q}_x\|_{\mathbb{R}^n} < \frac{q_x}{s}$. For $y \in \mathsf{B}^n(\frac{q_x}{2}, \boldsymbol{q}_x)$ we have

$$\|y - x\|_{\mathbb{R}^n} \leq \|y - \boldsymbol{q}_x\|_{\mathbb{R}^n} + \|\boldsymbol{q}_x - x\|_{\mathbb{R}^n} < \frac{q_x}{2} + \frac{q_x}{2} = q_x,$$

and so $y \in \mathsf{B}^n(q_x, x) \subseteq U$. Thus $\mathsf{B}^n(\frac{q_x}{2}, \boldsymbol{q}_x)$ is a ball of rational radius centred at a point with rational components, contained in U and containing x. Doing this for each x gives a collection of open balls of rational radius centred at points with rational components that covers U. The result will follow is we can show that the set of balls with rational radius with centres having rational components is countable. For fixed $x \in \mathbb{R}^n$ the set of balls centred at x with rational radius is certainly countable since $\mathbb{Q}_{>0}$ is countable. The subset $\mathbb{Q}^n \subseteq \mathbb{R}^n$ is also countable by since it has cardinality $n \cdot \mathrm{card}(\mathbb{Q})$ which is equal to $\mathrm{card}(\mathbb{Q})$ by Theorem I-1.7.17(ii). Thus the set of balls with rational radius centred at points with rational coordinates is a countable union of countable sets. Such sets are countable by Proposition I-1.7.16. ■

result about unions of rectangles?

1.2.5 Interior, closure, boundary, etc.

The definitions and results here are similar to those for $\mathbb{R}$ given in Section I-2.5.3. Moreover, they will be discussed in a more general setting in Section III-3.6.2. The proofs in the most general setting in Section III-3.6.2 are virtually identical to the proofs in the least general case in Section I-2.5.3. Therefore, we elect to omit the proofs in this section, and merely state the results for reference. Readers unable to translate the results from Section I-2.5.3 to this section can refer ahead to Section III-3.6.2; the only difference between the proofs in that section and what would appear here are trivial differences in notation. Moreover, examples, discussion, and motivation can be found in Section I-2.5.3.

1.2.22 Definition (Accumulation point, cluster point, limit point) For a subset $A \subseteq \mathbb{R}^n$, a point $\boldsymbol{x} \in \mathbb{R}^n$ is:

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

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

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

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

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.

1.2.23 Proposition ("Accumulation point" equals "cluster point") *For a set* $A \subseteq \mathbb{R}^n$, $\mathbf{x} \in \mathbb{R}^n$ *is an accumulation point for* A *if and only if it is a cluster point for* A.

1.2.24 Proposition (Properties of the derived set) *For* $A, B \subseteq \mathbb{R}^n$ *and for a family of subsets* $(A_i)_{i\in I}$ *of* $\mathbb{R}^n$, *the following statements hold:*

(i) $\operatorname{der}(\emptyset) = \emptyset$;

(ii) $\operatorname{der}(\mathbb{R}^n) = \mathbb{R}^n$;

(iii) $\operatorname{der}(\operatorname{der}(A)) = \operatorname{der}(A)$;

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

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

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

1.2.25 Definition (Interior, closure, and boundary) Let $A \subseteq \mathbb{R}^n$.

(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 $\mathrm{bd}(A) = \mathrm{cl}(A) \cap \mathrm{cl}(\mathbb{R}^n \setminus A)$. •

1.2.26 Proposition (Characterisation of interior, closure, and boundary) *For $A \subseteq \mathbb{R}^n$, the following statements hold:*

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

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

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

1.2.27 Proposition (Properties of interior) *For $A, B \subseteq \mathbb{R}^n$ and for a family of subsets $(A_i)_{i\in I}$ of $\mathbb{R}^n$, the following statements hold:*

(i) $\mathrm{int}(\emptyset) = \emptyset$;

(ii) $\mathrm{int}(\mathbb{R}^n) = \mathbb{R}^n$;

(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 $A \subseteq \mathbb{R}^n$ *is open if and only if* $\mathrm{int}(A) = A$.

1.2.28 Proposition (Properties of closure) *For $A, B \subseteq \mathbb{R}^n$ and for a family of subsets $(A_i)_{i\in I}$ of $\mathbb{R}^n$, the following statements hold:*

(i) $\mathrm{cl}(\emptyset) = \emptyset$;

(ii) $\mathrm{cl}(\mathbb{R}^n) = \mathbb{R}^n$;

(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 $A \subseteq \mathbb{R}^n$ *is closed if and only if* $\mathrm{cl}(A) = A$.

1.2.29 Proposition (Joint properties of interior, closure, boundary, and derived set) *For* $A \subseteq \mathbb{R}^n$*, the following statements hold:*

(i) $\mathbb{R}^n \setminus \mathrm{int}(A) = \mathrm{cl}(\mathbb{R}^n \setminus A)$;

(ii) $\mathbb{R}^n \setminus \mathrm{cl}(A) = \mathrm{int}(\mathbb{R}^n \setminus A)$.

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

(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) $\mathbb{R}^n = \mathrm{int}(A) \cup \mathrm{bd}(A) \cup \mathrm{int}(\mathbb{R}^n \setminus A)$.

We close this section by defining a useful notion related to the topics of this section.

1.2.30 Definition (Dense subset) A subset $D \subseteq \mathbb{R}^n$ is ***dense*** if $\mathrm{cl}(D) = \mathbb{R}^n$. •

There is a simple example of a countable dense subset of $\mathbb{R}^n$.

1.2.31 Example (Countable dense subset) The set $\mathbb{Q}^n$ is a dense subset of $\mathbb{R}^n$. To verify this one needs only, for $\boldsymbol{x} \in \mathbb{R}^n$, to construct a sequence $(\boldsymbol{q}_j)_{j\in\mathbb{Z}_{>0}}$ converging to $\boldsymbol{x}$. That this is possible follows from the fact that $\mathbb{Q} \subseteq \mathbb{R}$ is dense, along with Proposition 1.2.4. Moreover, note that $\mathbb{Q}^n$ is countable by Theorem I-1.7.17. •

1.2.6 Compact subsets

The notion of compactness, relying as it does only on the idea of an open set, is transferable from $\mathbb{R}$ to $\mathbb{R}^n$, and indeed to the general setting of Chapter III-1 (see Section III-1.6). That is to say, the idea of an open cover of a subset of $\mathbb{R}^n$ transfers directly from $\mathbb{R}$, and, therefore, the definition of a compact set as being a set for which every open cover possesses a finite subcover also generalises. In this section we explore the details of this for $\mathbb{R}^n$.

We begin with some notions associated to open covers.

1.2.32 Definition (Open cover of a subset of $\mathbb{R}^n$) Let $A \subseteq \mathbb{R}^n$.

(i) An ***open cover*** for A is a family $(U_i)_{i\in I}$ of open subsets of $\mathbb{R}^n$ 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 $(V_j)_{j\in J} \subseteq (U_i)_{i\in I}$. •

The following property of open covers of subsets of $\mathbb{R}^n$ is useful.

1.2.33 Lemma (Lindelöf Lemma for $\mathbb{R}^n$) *If* $(U_i)_{i\in I}$ *is an open cover of* $A \subseteq \mathbb{R}^n$*, then there exists a countable subcover of* A*.*

Proof Let $\mathscr{B} = \{\mathsf{B}^n(r, x) \mid r \in \mathbb{Q},\ x \in \mathbb{Q}^n\}$. Note that $\mathscr{B}$ is a countable union of countable sets, and so is countable by Proposition I-1.7.16 (also see the last part of the proof of Proposition 1.2.21). Therefore, we can write $\mathscr{B} = (\mathsf{B}^n(r_j, x_j))_{j\in\mathbb{Z}_{>0}}$. Now define

$$\mathscr{B}' = \{\mathsf{B}^n(r_j, x_j) \mid \mathsf{B}^n(r_j, x_j) \subseteq U_i \text{ for some } i \in I\}.$$

Let us write $\mathscr{B}' = (\mathsf{B}^n(r_{j_k}, x_{j_k}))_{k\in\mathbb{Z}_{>0}}$. We claim that $\mathscr{B}'$ covers A. Indeed, if $x \in A$ then $x \in U_i$ for some $i \in I$. Then there exists $k \in \mathbb{Z}_{>0}$ such that $x \in \mathsf{B}^n(r_{j_k}, x_{j_k}) \subseteq U_i$. Now, for each $k \in \mathbb{Z}_{>0}$, let $i_k \in I$ satisfy $\mathsf{B}^n(r_{j_k}, x_{j_k}) \subseteq U_{i_k}$. Then the countable collection of open sets $(U_{i_k})_{k\in\mathbb{Z}_{>0}}$ clearly covers A since $\mathscr{B}'$ covers A. ■

Now we define the important notion of compactness, along with some other related useful concepts.

1.2.34 Definition (Bounded, compact, and totally bounded in $\mathbb{R}^n$) A subset $A \subseteq \mathbb{R}^n$ is:

(i) ***bounded*** if there exists $M \in \mathbb{R}_{>0}$ such that $A \subseteq \overline{\mathsf{B}}^n(M, \mathbf{0})$;

(ii) ***compact*** if every open cover $(U_i)_{i\in I}$ of A possesses a finite subcover;

(iii) ***precompact***[3] if $\operatorname{cl}(A)$ is compact;

(iv) ***totally bounded*** if, for every $\epsilon \in \mathbb{R}_{>0}$ there exists $x_1, \dots, x_k \in \mathbb{R}^n$ such that $A \subseteq \cup_{j=1}^k \mathsf{B}^n(\epsilon, x_j)$. •

The simplest characterisation of compact subsets of $\mathbb{R}^n$ is the following. We shall freely interchange our use of the word compact between the definition given in Definition 1.2.34 and the conclusions of the following theorem.

1.2.35 Theorem (Heine–Borel Theorem in $\mathbb{R}^n$) *A subset $K \subseteq \mathbb{R}^n$ is compact if and only if K is closed and bounded.*

Proof We first prove a couple of lemmata.

1 Lemma *If $K \subseteq \mathbb{R}^m$ is compact and if $L \subseteq \mathbb{R}^n$ is compact then $K \times L \subseteq \mathbb{R}^{m+n}$ is compact.*

Proof Let us denote points in $\mathbb{R}^{m+n}$ by $(x, y) \in \mathbb{R}^m \times \mathbb{R}^n$.

Let $(U_a)_{a\in A}$ be an open cover of $K \times L$. For $(x, y) \in K \times L$, let $a(x, y) \in A$ be such that $(x, y) \in U_{a(x,x)}$. Let $V_{(x,y)} \subseteq \mathbb{R}^m$ and $W_{(x,y)} \subseteq \mathbb{R}^n$ be open sets such that $x \in V_{(x,y)}$, $y \in W_{(x,y)}$, and $V_{(x,y)} \times W_{(x,y)} \subseteq U_{a(x,y)}$.

We claim that, for fixed $x \in K$, $(W_{(x,y)})_{y\in L}$ is an open cover of L. Indeed, if $y \in L$, then $(x, y) \in K \times L$ and so $(x, y) \in U_{a(x,y)}$. By construction, we have $y \in W_{(x,y)}$, giving the claim. By compactness of L, let $k(x) \in \mathbb{Z}_{>0}$ and $y(x)_1, \dots, y(x)_{k(x)} \in L$ be such that $L \subseteq \cup_{j=1}^{k(x)} W_{x,y_j(x)}$.

Let $V_x = \cap_{j=1}^{k(x)} V_{(x,y_j(x))}$, noting that V_x is open by Proposition 1.2.19. Moreover, $(V_x)_{x\in K}$ covers K since $x \in V_x$. Let $x_1, \dots, x_l \in K$ be such that $K \subseteq \cup_{r=1}^l V_{x_r}$.

Finally, we claim that

$$\{U_{a(x_r, y(x_r)_j)} \mid r \in \{1, \dots, l\},\ j \in \{1, \dots, k(x_r)\}\}$$

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

is an open cover of $K \times L$. Indeed, if $(x, y) \in K \times L$, then $x \in V_{x_r}$ for some $r \in \{1, \dots, l\}$. Moreover, $y \in W_{(x_r, y(x_r)_j)}$ for some $j \in \{1, \dots, k(x_r)\}$. Therefore,

$$(x, y) \in V_{x_r} \times W_{(x_r, y(x_r)_j)} \subseteq V_{(x_r, y(x_r)_j)} \times W_{(x_r, y(x_r)_j)} \subseteq U_{a(x_r, y(x_r)_j)},$$

giving the desired conclusion. ▼

2 Lemma *If* A *is compact and if* $B \subseteq A$ *is closed, then* B *is compact.*

Proof Let $(U_i)_{i \in I}$ be an open cover for B and define $V = \mathbb{R}^m \setminus B$. Since B is closed, $(U_i)_{i \in I} \cup (V)$ is an open cover for A. Since A is compact there exists $i_1, \dots, i_k \in I$ such that $A \subseteq \cup_{j=1}^k U_{i_j} \cup V$. Therefore, $B \subseteq \cup_{j=1}^k U_{i_k}$, giving a finite subcover of B. ▼

Suppose that K is closed and bounded. Let $R \in \mathbb{R}_{>0}$ be sufficiently large that $K \subseteq [-R, R] \times \cdots \times [-R, R]$. By Theorem I-2.5.27 it follows that $[-R, R]$ is compact. By induction using Lemma 1 it follows that $[-R, R] \times \cdots \times [-R, R]$ is compact. By Lemma 2 it follows that K is compact.

Next suppose that K is compact. For $\epsilon \in \mathbb{R}_{>0}$ consider the open cover $(\mathsf{B}^n(\epsilon, x))_{x \in K}$ of K. Since K is compact there exists $x_1, \dots, x_k \in K$ such that $K \subseteq \cup_{j=1}^k \mathsf{B}^n(\epsilon, x_j)$. If

$$M_0 = \max\{\|x_j - x_l\|_{\mathbb{R}^n} \mid j, l \in \{i, \dots, k\}\} + 2\epsilon,$$

then it is easy to see that $A \subseteq \mathsf{B}^n(M, \mathbf{0})$ for any $M > M_0$; thus K is bounded. Now suppose that K is compact but not closed. Then, by Proposition 1.2.28, there exists $x_0 \in \mathrm{cl}(K) \setminus K$. For each $x \in K$ let $r_x \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^n(\epsilon_x, x) \cap \mathsf{B}^n(\epsilon_x, x_0) = \emptyset$. Then $(\mathsf{B}^n(\epsilon_x, x))_{x \in K}$ is an open cover of K. Therefore, there exists $x_1, \dots, x_k \in K$ such that $K \subseteq \cup_{j=1}^k \mathsf{B}^n(\epsilon_{x_j}, x_j)$. But this means that K does not intersect the open subset $\cap_{j=1}^k \mathsf{B}^n(\epsilon_{x_j}, x_0)$, so contradicting the existence of $x \in \mathrm{cl}(K) \setminus K$. Thus $K = \mathrm{cl}(K)$, giving the result. ■

The Heine–Borel Theorem has the following useful corollary.

1.2.36 Corollary (Closed subsets of compact sets in $\mathbb{R}^n$ are compact) *If* $A \subseteq \mathbb{R}^n$ *is compact and if* $B \subseteq A$ *is closed, then* B *is compact.*

Proof This was proved as Lemma 2 in the proof of the Heine–Borel Theorem. ■

As we warned the reader in Section I-2.5.4, care must be taken when generalising the notion of compactness from $\mathbb{R}^n$ to the more general notion of a topological space as defined in Chapter III-1. A key fact is that compactness and closed and boundedness are not generally equivalent. Perhaps the nicest illustration of this is given in Theorem III-3.6.15 where it is shown that, for Banach spaces, this equivalence happens only in finite dimensions.

The following result is another equivalent characterisation of compact subsets of $\mathbb{R}^n$, and is often useful.

1.2.37 Theorem (Bolzano–Weierstrass Theorem in $\mathbb{R}^n$) *A subset* $K \subseteq \mathbb{R}^n$ *is compact if and only if every sequence in* K *has a subsequence which converges in* K.

Proof Suppose that there exists a sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in K having no convergent subsequence. This means that for each $j \in \mathbb{Z}_{>0}$ there exists $\epsilon_j \in \mathbb{R}_{>0}$ such that $x_k \notin \mathsf{B}^n(\epsilon_j, \boldsymbol{x}_j)$ for $k \neq j$. Let $X \triangleq \{\boldsymbol{x}_j \mid j \in \mathbb{Z}_{>0}\}$. The open cover $(\mathsf{B}^n(\epsilon_j, \boldsymbol{x}_j))_{j\in\mathbb{Z}_{>0}}$ of X possesses no finite subcover and so X is not compact. We claim that the set is closed. Indeed, if $\boldsymbol{x} \in \mathrm{cl}(X)$ it follows by Proposition 1.2.26 that $\boldsymbol{x}$ is the limit of a sequence in X. But the only such sequences are those that are eventually constant, and so the claim follows. By Corollary 1.2.36 it now follows that K is not compact since it possesses a closed but not compact subset.

Next suppose that every sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in K possesses a convergent subsequence. Let $(U_i)_{i\in I}$ be an open cover of K, and by Lemma 1.2.33 choose a countable subcover which we denote by $(U_j)_{j\in\mathbb{Z}_{>0}}$. Now suppose that every finite subcover of $(U_j)_{j\in\mathbb{Z}_{>0}}$ does not cover K. This means that, for every $k \in \mathbb{Z}_{>0}$, the set $C_k = K \setminus \left(\cup_{j=1}^k U_j\right)$ is nonempty. Thus we may define a sequence $(\boldsymbol{x}_k)_{k\in\mathbb{Z}_{>0}}$ in $\mathbb{R}^n$ such that $\boldsymbol{x}_k \in C_k$. Since the sequence $(x_k)_{k\in\mathbb{Z}_{>0}}$ is in K, it possesses a convergent subsequence $(\boldsymbol{x}_{k_m})_{m\in\mathbb{Z}_{>0}}$, by hypotheses. Let $\boldsymbol{x}$ be the limit of this subsequence. Since $\boldsymbol{x} \in K$ and since $K = \cup_{j\in\mathbb{Z}_{>0}} U_j$, $\boldsymbol{x} \in U_l$ for some $l \in \mathbb{Z}_{>0}$. Since the sequence $(\boldsymbol{x}_{k_m})_{m\in\mathbb{Z}_{>0}}$ converges to $\boldsymbol{x}$, it follows that there exists $N \in \mathbb{Z}_{>0}$ such that $\boldsymbol{x}_{k_m} \in U_l$ for $m \geq N$. But this contradicts the definition of the sequence $(\boldsymbol{x}_k)_{k\in\mathbb{Z}_{>0}}$, forcing us to conclude that our assumption is wrong that there is no finite subcover of K from the collection $(U_j)_{j\in\mathbb{Z}_{>0}}$. ■

The following property of compact subsets of $\mathbb{R}^n$ is useful.

1.2.38 Theorem (Lebesgue number for compact sets) *Let* $K \subseteq \mathbb{R}^n$ *be a compact set. Then for any open cover* $(U_\alpha)_{\alpha\in A}$ *of* K*, there exists* $\delta \in \mathbb{R}_{>0}$*, called the* ***Lebesgue number*** *of* K*, such that, for each* $\mathbf{x} \in K$*, there exists* $\alpha \in A$ *such that* $\mathsf{B}^n(\delta, \mathbf{x}) \cap K \subseteq U_\alpha$.

Proof Suppose there exists an open cover $(U_\alpha)_{\alpha\in A}$ such that, for all $\delta \in \mathbb{R}_{>0}$, there exists $\boldsymbol{x} \in K$ such that none of the sets U_α, $\alpha \in A$, contains $\mathsf{B}^n(\delta, \boldsymbol{x}) \cap K$. Then there exists a sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in K such that

$$\left\{\alpha \in A \;\middle|\; \mathsf{B}^n(\tfrac{1}{j}, \boldsymbol{x}_j) \subseteq U_\alpha\right\} = \emptyset$$

for each $j \in \mathbb{Z}_{>0}$. By the Bolzano–Weierstrass Theorem there exists a subsequence $(\boldsymbol{x}_{j_k})_{k\in\mathbb{Z}_{>0}}$ that converges to a point, say $\boldsymbol{x}$, in K. Then there exists $\epsilon \in \mathbb{R}_{>0}$ and $\alpha \in A$ such that $\mathsf{B}^n(\epsilon, \boldsymbol{x}) \subseteq U_\alpha$. Now let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $\|\boldsymbol{x}_{j_k} - \boldsymbol{x}\|_{\mathbb{R}^n} < \frac{\epsilon}{2}$ for $k \geq N$ and such that $\frac{1}{j_N} < \frac{\epsilon}{2}$. Now let $k \geq N$. Then, if $\boldsymbol{y} \in \mathsf{B}^n(\frac{1}{j_k}, \boldsymbol{x}_{j_k})$ we have

$$\|\boldsymbol{y} - \boldsymbol{x}\|_{\mathbb{R}^n} = \|\boldsymbol{y} - \boldsymbol{x}_{j_k} + \boldsymbol{x}_{j_k} - \boldsymbol{x}\|_{\mathbb{R}^n} \leq \|\boldsymbol{y} - \boldsymbol{x}_{j_k}\|_{\mathbb{R}^n} + \|\boldsymbol{x} - \boldsymbol{x}_{j_k}\|_{\mathbb{R}^n} < \epsilon.$$

Thus we arrive at the contradiction that $\mathsf{B}^n(\frac{1}{j_k}, \boldsymbol{x}_{j_k}) \subseteq U_\alpha$. ■

The following result is useful and is sometimes known as the ***Cantor Intersection Theorem***.

1.2.39 Proposition (Countable intersections of nested compact sets are nonempty) *Let $(K_j)_{j\in\mathbb{Z}_{>0}}$ be a collection of nonempty compact subsets of $\mathbb{R}^n$ satisfying $K_{j+1} \subseteq K_j$. Then $\cap_{j\in\mathbb{Z}_{>0}} K_j$ is nonempty.*

Proof It is clear that $K = \cap_{j\in\mathbb{Z}_{>0}} K_j$ is bounded, and moreover it is closed by Exercise 1.2.3. Thus K is compact by the Heine–Borel Theorem. Let $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence for which $\boldsymbol{x}_j \in K_j$ for $j \in \mathbb{Z}_{>0}$. This sequence is thus a sequence in K_1 and so, by the Bolzano–Weierstrass Theorem, has a subsequence $(\boldsymbol{x}_{j_k})_{k\in\mathbb{Z}_{>0}}$ converging to $\boldsymbol{x} \in K_1$. The sequence $(\boldsymbol{x}_{j_{k+1}})_{k\in\mathbb{Z}_{>0}}$ is then a sequence in K_2 which is convergent, so showing that $\boldsymbol{x} \in K_2$. Similarly, one shows that $\boldsymbol{x} \in K_j$ for all $j \in \mathbb{Z}_{>0}$, giving the result. ■

Finally, let us indicate the relationship between the notions of relative compactness and total boundedness. We see that for $\mathbb{R}^n$ these concepts are the same. This may not be true in general.

example in Chapter III-1?

1.2.40 Proposition ("Precompact" equals "totally bounded" in $\mathbb{R}^n$) *A subset of $\mathbb{R}^n$ is precompact if and only if it is totally bounded.*

Proof Let $A \subseteq \mathbb{R}^n$.

First suppose that A is precompact. Since $A \subseteq \mathrm{cl}(A)$ and since $\mathrm{cl}(A)$ is bounded by the Heine–Borel Theorem, it follows that A is bounded. We claim that A is then totally bounded. Let $M \in \mathbb{R}_{>0}$ be such that $A \subseteq \overline{\mathsf{B}}^n(M, \boldsymbol{0})$ so that $\mathrm{cl}(A) \subseteq \overline{\mathsf{B}}^n(M, \boldsymbol{0})$ by Proposition 1.2.28(iv). Thus $\mathrm{cl}(A)$ is closed and bounded, and so compact by the Heine–Borel Theorem. For $\epsilon \in \mathbb{R}_{>0}$ note that $(\mathsf{B}^n(\epsilon, \boldsymbol{x}))_{\boldsymbol{x}\in\mathrm{cl}(A)}$ is an open cover of $\mathrm{cl}(A)$. Thus there exists a finite collection $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathrm{cl}(A)$ such that $\mathrm{cl}(A) \subseteq \cup_{j=1}^k \mathsf{B}^n(\epsilon, \boldsymbol{x}_j)$. Since $A \subseteq \mathrm{cl}(A)$ this shows that A is totally bounded.

Now suppose that A is totally bounded. For $\epsilon \in \mathbb{R}_{>0}$ let $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathbb{R}^n$ have the property that $A \subseteq \cup_{j=1}^k \mathsf{B}^n(\epsilon, \boldsymbol{x}_j)$. If

$$M_0 = \max\{\|\boldsymbol{x}_j - \boldsymbol{x}_l\|_{\mathbb{R}^n} \mid j, l \in \{i, \dots, k\}\} + 2\epsilon,$$

then it is easy to see that $A \subseteq \mathsf{B}^n(M, \boldsymbol{0})$ for any $M > M_0$. Then $\mathrm{cl}(A) \subseteq \overline{\mathsf{B}}^n(M, \boldsymbol{0})$ by part (iv) of Proposition 1.2.28, and so $\mathrm{cl}(A)$ is bounded. Since $\mathrm{cl}(A)$ is closed, it follows from the Heine–Borel Theorem that A is precompact. ■

We close this section with a discussion of a notion of the size of a set.

1.2.41 Definition (Diameter of a set) The ***diameter*** of a set $A \subseteq \mathbb{R}^n$ is

$$\mathrm{diam}(A) = \sup\{\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathbb{R}^n} \mid \boldsymbol{x}_1, \boldsymbol{x}_2 \in A\}. \quad \bullet$$

The following properties of the diameter are useful.

1.2.42 Proposition (Properties of diameter) *For $A \subseteq \mathbb{R}^n$ the following statements hold:*

(i) $\mathrm{diam}(A) < \infty$ *if and only if* A *is bounded;*

(ii) $\mathrm{diam}(\mathrm{cl}(A)) = \mathrm{diam}(A)$.

Proof (i) Suppose that $\operatorname{diam}(A) = D \in \mathbb{R}_{>0}$. Let $x_0 \in A$ and define $M = D + \|x_0\|_{\mathbb{R}^n}$. Then, for $x \in A$ we have

$$\|x\|_{\mathbb{R}^n} = \|x - x_0\|_{\mathbb{R}^n} + \|x_0\|_{\mathbb{R}^n} < M$$

and so $A \subseteq \mathsf{B}^n(M, \mathbf{0})$.

Now suppose that A is bounded and let $M \in \mathbb{R}_{>0}$ be such that $A \subseteq \mathsf{B}^n(M, \mathbf{0})$. Let $x_1, x_2 \in A$ so that

$$\|x_1 - x_2\|_{\mathbb{R}^n} \leq \|x_1\|_{\mathbb{R}^n} + \|x_2\|_{\mathbb{R}^n} < 2M.$$

Therefore,

$$\sup\{\|x_1 - x_2\|_{\mathbb{R}^n} \mid x_1, x_2 \in A\} \leq 2M,$$

and so $\operatorname{diam}(A) \leq 2M$.

(ii) Let $x_1, x_2 \in \operatorname{cl}(A)$ and let $(x_{1,j})_{j\in\mathbb{Z}_{>0}}$ and $(x_{2,j})_{j\in\mathbb{Z}_{>0}}$ be sequences in A converging to x_1 and x_2, respectively. Then, for each $j \in \mathbb{Z}_{>0}$,

$$\|x_{1,j} - x_{2,j}\|_{\mathbb{R}^n} \leq \operatorname{diam}(A),$$

which gives

$$\|x_1 - x_2\|_{\mathbb{R}^n} = \lim_{j\to\infty} \|x_{1,j} - x_{2,j}\|_{\mathbb{R}^n} \leq \operatorname{diam}(A),$$

where we have swapped the limit with the norm using continuity of the norm () and Theorem 1.3.2. ■ what?

1.2.7 Connected subsets

It is pretty easy to characterise connectivity in $\mathbb{R}$, as we saw in Section I-2.5.5. Here we discuss connectedness in $\mathbb{R}^n$, and as we shall see things are a little more complicated in this case.

One of the reasons why connectedness is more complicated in dimensions higher than one is because there are two natural distinct notions of connectivity. As we shall see, these agree in one dimension, but not in higher dimensions.

The first notion we consider is fairly intuitive. It relies on the notion of paths in Euclidean spaces which are discussed in Section 2.2.1. Readers who cannot imagine what is the definition of a path can refer ahead.

1.2.43 Definition (Path-connected subset of $\mathbb{R}^n$) A subset $A \subseteq \mathbb{R}^n$ is ***path-connected*** if, for every $x_0, x_1 \in \mathbb{R}^n$ there exists a path $\gamma\colon [a, b] \to \mathbb{R}^n$ such that $\gamma(s) \in A$ for every $s \in [a, b]$ and such that $\gamma(a) = x_0$ and $\gamma(b) = x_1$. •

The idea is that the map γ is to be thought of as a curve, or path, from x_1 to x_2. Path-connectedness of A is the property of going from any point in A to any other point in A in a continuous manner while remaining in A. This is depicted in Figure 1.4.

Besides this fairly intuitive notion of path-connectedness (which, as we shall see, agrees with our notion of connectedness from Definition I-2.5.33) we can duplicate the definition we have already seen for subsets of $\mathbb{R}$.

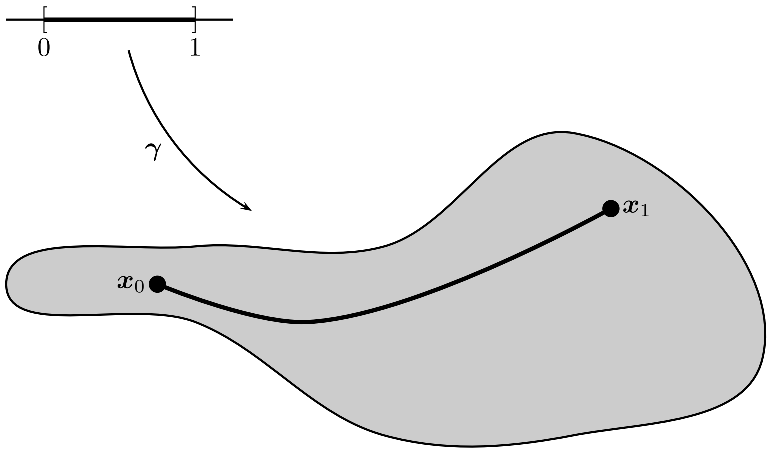

Figure 1.4 A depiction of a path-connected set

1.2.44 Definition (Connected subset of $\mathbb{R}^n$) Subsets $A, B \subseteq \mathbb{R}^n$ are ***separated*** if $A \cap \mathrm{cl}(B) = \emptyset$ and $\mathrm{cl}(A) \cap B = \emptyset$. A subset $S \subseteq \mathbb{R}^n$ is ***disconnected*** if $S = A \cup B$ for nonempty separated subsets A and B. A subset $S \subseteq \mathbb{R}^n$ is ***connected*** if it is not disconnected. •

For subsets of $\mathbb{R}$ (i.e., in the case when $n = 1$) we have the simple characterisation of connected sets from Theorem I-2.5.34. For subsets of $\mathbb{R}^n$ with $n > 1$ there is no such elementary characterisation. Indeed, as we shall see in Example 1.2.46 below, some connected sets can be pretty complicated, and not "obviously" connected.

But before we get to this, let us give the relationship between connectedness and path-connectedness.

1.2.45 Proposition (Path-connected sets are connected) *If $A \subseteq \mathbb{R}^n$ is path-connected then it is connected.*

Proof Suppose that A is not connected but is path-connected. Let $A = A_1 \cup A_2$ with A_1 and A_2 nonempty separated sets. Let $x_1 \in A_1$ and $x_2 \in A_2$ and let $\gamma\colon [0,1] \to \mathbb{R}^2$ be continuous, A-valued, and have the property that $\gamma(0) = x_1$ and $\gamma(1) = x_2$. Define $B_1 = \gamma^{-1}(A_1)$ and $B_2 = \gamma^{-1}(A_2)$. We claim that B_1 and B_2 are separated. Indeed, suppose that $B_1 \cap \mathrm{cl}(B_2)$ is nonempty and let $s_0 \in B_1 \cap \mathrm{cl}(B_2)$. Since $s_0 \in B_1$ we have $\gamma(s_0) \in A_1$. Note that $\mathrm{cl}(B_2)$ is closed and bounded, and so compact by the Heine–Borel Theorem. By Proposition 1.3.29 it follows that $\gamma(\mathrm{cl}(B_2))$ is compact, and so in particular closed. Since γ is continuous, since $\mathrm{cl}(B_2)$ is closed, and since $\gamma(\mathrm{cl}(B_2))$ is closed, it follows from Theorem 1.3.2 and Proposition 1.2.26 that $\gamma(s_0) \in \gamma(\mathrm{cl}(B_2))$. But this implies that $\gamma(s_0) \in \mathrm{cl}(A_2)$ and so this contradicts the connectedness of A. Thus A cannot be path-connected. ■

1.2.46 Example (A set that is connected but not path connected) Let us consider the subset S of $\mathbb{R}^2$ defined by

$$S = \{(x, y) \in \mathbb{R}^2 \mid y = \sin \tfrac{1}{x},\ x \neq 0\} \cup \{(0, y) \mid y \in [-1, 1]\}.$$

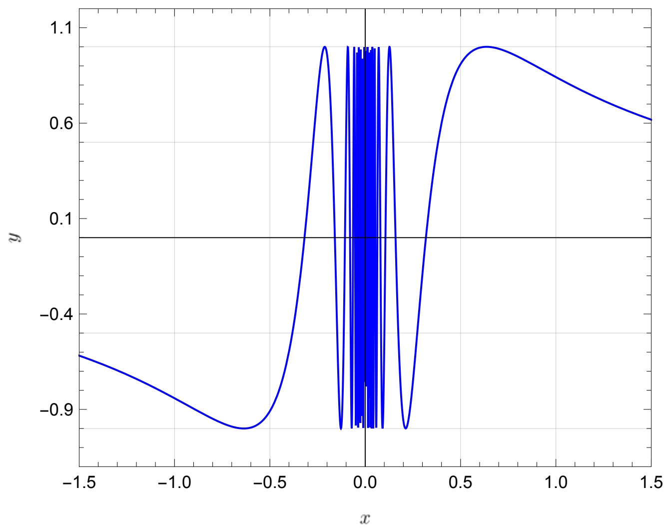

Figure 1.5 The topologist's sine curve

In Figure 1.5 we depict this subset which is sometimes called the ***topologist's sine curve***. (Actually, usually the first set in the definition of S is what is the topologist's sine curve, and the set S is its closure.)

We first claim that S is connected. Let us write $S = S_1 \cup S_2 \cup S_3$ with

$$\begin{aligned} S_1 &= \{(x,y) \in \mathbb{R}^2 \mid \ y = \sin\tfrac{1}{x},\ x > 0\}, \\ S_2 &= \{(x,y) \in \mathbb{R}^2 \mid \ y = \sin\tfrac{1}{x},\ x < 0\}, \\ S_3 &= \{(0,y) \mid \ y \in [-1,1]\}. \end{aligned}$$

It is evident that S_1, S_2, and S_2 are path-connected since they are images of intervals under continuous maps. Therefore, they are connected by Proposition 1.2.45. Thus none of S_1, S_2, or S_3 are the union of separated subsets. Moreover, since S is the closure of $S_1 \cup S_2$ (why?) it follows that S is connected by Exercise 1.2.6.

Next we claim that S is not path-connected. To see this, suppose that there exists a continuous map $\gamma\colon [0,1] \to \mathbb{R}^2$ taking values in S and such that $\gamma(0) = (\frac{1}{\pi}, 0)$ and $\gamma(1) = (0,0)$. Let

$$s_* = \inf\{s \in [0,1] \mid \ \gamma(s) \in \{0\} \times \mathbb{R}\}.$$

Such an s_* exists since $\gamma(1) = (0,0)$ and so $s_* \leq 1$. Therefore, $\gamma([0,s_*])$ intersects the y-axis at exactly one point. However, $S_3 \subseteq \mathrm{cl}(\gamma([0,s_*]))$ (why?) which implies that $\gamma([0,s_*])$ is not closed, and so not compact by the Heine–Borel Theorem. But this contradicts the continuity of γ by Proposition 1.3.29. •

An important class of subsets where connectedness and path-connectedness agree are open sets. Here one can connect points with particular paths called polygonal paths. The reader can get the precise definition from Definition 2.2.6, although the intuition is easy: a polygonal path is formed from a finite collection of line segments.

1.2.47 Theorem (Open connected sets are polygonally path connected) *If* $U \subseteq \mathbb{R}^n$ *is open and connected then, given* $\mathbf{x}_0, \mathbf{x}_1 \in U$*, there exists a polygonal path lying in* U *connecting* $\mathbf{x}_0$ *and* $\mathbf{x}_1$*.*

Proof Let $x_0 \in U$ and let $A_{x_0} \subseteq U$ be the set of points that can be connected to x_0 with a polygonal path lying in U. We claim that A_{x_0} is a nonempty open set. Since U is open there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\epsilon, x_0) \subseteq U$. If $v \in \mathbb{R}^n$ such that $\|v\|_{\mathbb{R}^n} = 1$ then

$$x_0 + sv \in \mathsf{B}^n(\epsilon, x_0) \subseteq U, \qquad s \in [0, \epsilon).$$

Thus A_{x_0} is not empty. Now let $x \in A_{x_0}$. Since $x \in U$ there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\epsilon, x) \subseteq U$. Again, for any vector $v \in \mathbb{R}^n$ such that $\|v\|_{\mathbb{R}^n} = 1$ we have

$$x + sv \in \mathsf{B}^n(\epsilon, x_0) \subseteq U, \qquad s \in [0, \epsilon).$$

Thus $\mathsf{B}^n(\epsilon, x) \subseteq A_{x_0}$ since x_0 can be connected to x by a polygonal path and every point in $\mathsf{B}^n(\epsilon, x)$ can be connected to x by a segment. This shows that A_{x_0} is open.

Next we claim that $\mathrm{bd}(A_{x_0}) \cap U = \emptyset$. Indeed, let $x \in \mathrm{bd}(A_{x_0}) \cap U$. Since $x \in U$ there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\epsilon, x) \subseteq U$. Since $x \in \mathrm{bd}(A_{x_0})$ and by Proposition 1.2.26, there exists $x' \in \mathsf{B}^n(\epsilon, x)$ such that $x' \in A_{x_0}$. But then x can be connected to x' by a segment (just as in the preceding parts of the proof) and x_0 can be connected to x' by a polygonal path, meaning that x_0 can be connected to x by a polygonal path. Thus $x \in A_{x_0} \cap \mathrm{bd}(A_{x_0})$, contradicting the openness of A_{x_0}.

Let $B_{x_0} = U \setminus A_{x_0}$. We claim that $B_{x_0} = \emptyset$. Suppose otherwise. First we claim that B_{x_0} is open. Let $x \in B_{x_0}$. Since $x \in U$ there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\epsilon, x) \subseteq U$. As above, x can be connected to any point in $\mathsf{B}^n(\epsilon, x)$ by a segment. This ensures that $\mathsf{B}^n(\epsilon, x) \cap A_{x_0} = \emptyset$ since otherwise this implies the existence of a polygonal path connecting x_0 to x. Thus $\mathsf{B}^n(\epsilon, x) \subseteq B_{x_0}$ and so B_{x_0} is indeed open.

We next claim that $\mathrm{bd}(B_{x_0}) \cap U = \emptyset$. Suppose otherwise and let $x \in \mathrm{bd}(B_{x_0}) \cap U$. Since $x \in U$ there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\epsilon, x) \subseteq U$. Since $x \in \mathrm{bd}(B_{x_0})$ and by Proposition 1.2.26, there exists $x' \in \mathsf{B}^n(\epsilon, x)$ such that $x' \in B_{x_0}$. This means, as we have seen several times now, that x can be connected to x' by a segment. This means $x \notin A_{x_0}$ since otherwise this would imply the existence of a polygonal path from x_0 to x'. Thus $x \in B_{x_0} \cap \mathrm{bd}(B_{x_0})$, contradicting the openness of B_{x_0}.

Since $\mathrm{cl}(B_{x_0}) = B_{x_0} \cup \mathrm{bd}(B_{x_0})$ and $\mathrm{cl}(A) = A_{x_0} \cup \mathrm{bd}(A_{x_0})$, since $\mathrm{bd}(A_{x_0}) \cap U = \emptyset$, and since $\mathrm{bd}(B_{x_0}) \cap U = \emptyset$, it follows that $\mathrm{cl}(A_{x_0}) \cap B_{x_0} = \emptyset$ and $A_{x_0} \cap \mathrm{cl}(B_{x_0}) = \emptyset$. Thus, by assuming that B_{x_0} we show that U is a disjoint union of separated sets, contradicting the connectedness of U. Thus we must have $B_{x_0} = \emptyset$ and so $U = A_{x_0}$, as desired. ■

The preceding proposition implies the following interesting result.

1.2.48 Corollary (Open connected sets are differentiably path connected) *If* $U \subseteq \mathbb{R}^n$ *is open and connected then, given* $\mathbf{x}_0, \mathbf{x}_1 \in U$*, there exists a differentiable path lying in* U *connecting* $\mathbf{x}_0$ *and* $\mathbf{x}_1$*.*

Proof From Theorem 1.2.47 let γ be a polygonal path connecting x_0 and x_1. Let $y_1, \dots, y_k \in U$ be the points at which γ is not differentiable, i.e., the "corner" points of the polygonal path. Now let $\epsilon \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^n(\epsilon, y_j) \subseteq U$ for each $j \in \{1, \dots, k\}$. By Theorem 2.2.8 (or more precisely, by following the idea of the proof of that theorem as depicted in Figure 2.4) there then exists a differentiable path γ_{diff} connecting x_0 with x_1 and that lies in U. ■

totally disconnected sets

1.2.8 Subsets and relative topology

We have thus far been discussing properties of subsets of $\mathbb{R}^n$. However, sometimes it is useful to discuss subsets of subsets, and the properties of the smaller subset relative to the larger subset, not relative to $\mathbb{R}^n$. We shall revisit this idea in a more general (and in some sense, more suitable) setting in Section III-1.4.1; one way to think of this section is that it gives s gentle introduction to the more general material to come. We shall in this section make occasional and casual use of the terminology "relative topology," although it will not be defined until Section III-1.4.1.

Relatively open and closed sets

The key is the following definition.

1.2.49 Definition (Relatively open and closed subsets) Let $S \subseteq \mathbb{R}^n$ and let $A \subseteq S$.

(i) The subset $A \subseteq S$ is ***relatively open*** in S if, for every $x \in A$ there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\epsilon, x) \cap S \subseteq A$.

(ii) The subset $A \subseteq S$ is ***relatively closed*** in S if $S \setminus A$ is relatively open in S. •

We shall often omit "in S" in "relatively open in S" when it is understood what set S is being used.

Let us characterise the notion of relatively open and relatively closed sets in a useful way.

1.2.50 Proposition (Characterisation of relatively open and closed subsets) *For* $S \subseteq \mathbb{R}^n$ *and for* $A \subseteq S$ *the following statements hold:*

(i) A *is relatively open in* S *if and only if there exists an open subset* $U \subseteq \mathbb{R}^n$ *such that* $A = S \cap U$*;*

(ii) A *is relatively closed in* S *if and only if there exists a closed subset* $C \subseteq \mathbb{R}^n$ *such that* $A = S \cap C$.

Proof (i) Suppose that A is relatively open and let $x \in A$. Let $\epsilon_x \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^n(\epsilon_x, x) \cap S \subseteq A$. Then $U = \cup_{x \in A} \mathsf{B}^n(\epsilon_x, x)$ is open and has the property that $A = S \cap U$.

Conversely, let $A = S \cap U$ for an open set U. Then, for $x \in A$ there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\epsilon, x) \subseteq U$. Therefore, $\mathsf{B}^n(\epsilon, x) \cap S \subseteq U \cap S = A$.

(ii) Suppose that A is relatively closed so that $S \setminus A$ is relatively open. By the previous part of the result, $S \setminus A = S \cap U$ for an open subset $U \subseteq \mathbb{R}^n$. Thus

$$A = S \setminus (S \setminus A) = S \setminus (S \cap (U \cap A)) = S \cap (S \setminus (U \cap A)) = S \cap (\mathbb{R}^n \setminus U),$$

using DeMorgan's Laws. Taking $C = \mathbb{R}^n \setminus U$ gives the result.

Conversely, suppose that $A = S \cap C$ for a closed set C. Then $S \setminus A = (\mathbb{R}^n \setminus C) \cap S$ so that $S \setminus A$ is relatively open by the previous part of the result. Thus A is relatively closed. ■

These ideas of relatively open and closed subsets seems simple, but some care must be exercised in using them. Some examples illustrate the possible pitfalls.

1.2.51 Examples (Relatively open and closed subsets)

1. For any subset $S \subseteq \mathbb{R}^n$, the subset $S \subseteq S$ is always both relatively open and relatively closed. It is also true that $\varnothing \subseteq S$ is also both open and closed.
2. Let $S = (0, 1)$. Then, as in the preceding general example, $S \subseteq S$ is closed. Note, however, that S is not a closed subset of $\mathbb{R}$.
3. Let $S = [0, 1]$. Then $S \subseteq S$ is open although S is not an open subset of $\mathbb{R}$.
4. Let us consider $S = \mathbb{Z}$ as a subset of $\mathbb{R}$. We claim every subset of S is open. Indeed, let $A \subseteq \mathbb{Z}$ and let $x \in A$. Then $\mathsf{B}^n(\frac{1}{2}, x) \cap S = \{x\} \subseteq A$, showing that A is indeed open. A subset where every subset is open is called a ***discrete*** subset, and agrees with the usual notion of a discrete subset; see Exercise 1.2.7.
5. Let us examine $\mathbb{Q} \subseteq \mathbb{R}$, and consider some of its open and closed sets.
   (a) We claim that every singleton $\{q\} \subseteq \mathbb{Q}$ is not relatively open but is relatively closed. Since $\{q\} = \{q\} \cap \mathbb{Q}$, $\{q\}$ is relatively closed by Proposition 1.2.50. By Proposition 1.2.50 it follows that a relatively open subset of $\mathbb{Q}$ containing q must be of the form $U \cap \mathbb{Q}$ where U is an open subset of $\mathbb{R}$ containing q. Since U is a disjoint union of open intervals by Proposition I-2.5.6, any relatively open subset of $\mathbb{Q}$ containing q will contain $(a, b)\mathbb{Q}$ for an open interval (a, b) containing q. However, *every* subset of $\mathbb{Q}$ of the form $(a, b) \cap \mathbb{Q}$ will contain infinitely many elements. Thus any relatively open subset of $\mathbb{Q}$ containing q will contain infinitely many elements. In particular, $\{q\}$ is not relatively open. Thus $\mathbb{Q}$ is not discrete.
   (b) We claim that for every $q \in \mathbb{Q}$ and for every $\epsilon \in \mathbb{R}_{>0}$ there exists a neighbourhood of q that is both open and closed and is contained in an interval of length at most ϵ. Indeed, let $r_1 \in (q - \frac{\epsilon}{2}, q)$ and $r_2 \in (q, q + \frac{\epsilon}{2})$ be irrational, this being possible by Proposition I-2.2.17. We claim that $(r_1, r_2) \cap \mathbb{Q}$ is both relatively open and relatively closed. It is relatively open by Proposition 1.2.50. Note that

$$\begin{aligned}\mathbb{Q} \setminus ((r_1, r_2) \cap \mathbb{Q}) &= ((-\infty, r_1] \cap \mathbb{Q}) \cup ([r_2, \infty) \cap \mathbb{Q}) \\ &= ((-\infty, r_1) \cap \mathbb{Q}) \cup ((r_2, \infty) \cap \mathbb{Q}),\end{aligned}$$

the latter inequality since r_1 and r_2 are irrational. This shows, by Proposition 1.2.50, that $\mathbb{Q} \setminus ((r_1, r_2) \cap \mathbb{Q})$ is open, and so $(r_1, r_2) \cap \mathbb{Q}$ is closed. •

One can, in the expected way, define the notion of a neighbourhood in this setup.

1.2.52 Definition (Relative neighbourhood) Let $S \subseteq \mathbb{R}^n$. A ***relative neighbourhood*** of $\boldsymbol{x} \in S$ is a relatively open subset $U \subseteq S$ for which $\boldsymbol{x} \in U$. More generally, a ***relative neighbourhood*** of $A \subseteq S$ is a relatively open set $U \subseteq S$ for which $A \subseteq U$. •

Many of the notions we have given above for subsets of $\mathbb{R}^n$ also apply to subsets of subsets of $\mathbb{R}^n$. For example...

1.2.53 Definition (Accumulation point, cluster point, limit point) For $S \subseteq \mathbb{R}^n$ and for $A \subseteq S$, a point $\boldsymbol{x} \in S$ is:

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

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

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

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

Relative interior, closure, and boundary

One can also define the notions of interior, closure, and boundary for subsets of subsets.

1.2.54 Definition (Relative interior, closure, and boundary) Let $S \subseteq \mathbb{R}^n$ and let $A \subseteq S$.

(i) The ***relative interior*** of A in S is the set

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

(ii) The ***relative closure*** of A in S is the set

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

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

The properties of the interior, closure, and boundary given in Propositions 1.2.26, 1.2.27, 1.2.28, and 1.2.29 are also valid for the relative interior, relative closure, and relative boundary. Indeed, they are valid in the far more general context of topological spaces (see Chapter III-1). Thus we do not present the results here, but we shall occasionally use them.

Let us give some examples to illustrate that these notions should be thought about carefully in examples.

1.2.55 Examples (Relative interior, closure, and boundary)

1. If $S = (0, 1)$ then $\mathrm{int}_S(S) = S$ and $\mathrm{cl}_S(S) = S$ since S is both open and closed. Note, however, that $\mathrm{cl}(S) = [0, 1]$. Also note that $\mathrm{bd}_S(S) = \emptyset$ while $\mathrm{bd}(S) = \{0, 1\}$.
2. If $S = [0, 1]$ then $\mathrm{int}_S(S) = S$ and $\mathrm{cl}_S(S) = S$ since S is both open and closed. Note, however, that $\mathrm{int}(S) = (0, 1)$. Also note that $\mathrm{bd}_S(S) = \emptyset$ while $\mathrm{bd}(S) = \{0, 1\}$. •

For subsets the notion of denseness carries over in an obvious way.

1.2.56 Definition (Dense subset) If $A \subseteq \mathbb{R}^n$ a subset $D \subseteq A$ is ***dense*** in A if $\mathrm{cl}_A(D) = A$. •

Some example illustrate the notion of dense subsets.

1.2.57 Examples (Dense subsets)

1. The set $\mathbb{Q} \cap [0, 1]$ is dense in $[0, 1]$.
2. The set $(0, 1)$ is dense in $[0, 1]$. •

Relatively compact sets

Let us now consider the matter of when a subset of a set is compact. The following definition is the obvious one.

1.2.58 Definition (Relatively compact) Let $A \subseteq \mathbb{R}^n$. A subset $K \subseteq A$ is ***relatively compact***[4] if, for every family $(U_i)_{i\in I}$ of relatively open subsets of A such that $K \subseteq \cup_{i\in I} U_i$, there exists $i_1, \dots, i_k \in I$ such that $K \subseteq \cup_{j=1}^k U_{i_j}$. •

It turns out that this definition of relative compactness is the same as compactness in the usual sense.

1.2.59 Proposition (Characterisation of relatively compact sets) *Let* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$. *A subset* $\mathrm{K} \subseteq \mathrm{A}$ *is relatively compact if and only if* K *is compact as a subset of* $\mathbb{R}^{\mathrm{n}}$.

Proof First suppose that K is a relatively compact subset of A. Let $(U_i')_{i\in I}$ be a family of open subsets of $\mathbb{R}^n$ such that $K \subseteq \cup_{i\in I} U_i'$. For each $i \in I$ define $U_i = U_i' \cap A$, noting that U_i is a relatively open subset of A by Proposition 1.2.50. Since $K \subseteq \cup_{i\in I} U_i$ and since K is relatively compact, there exists $i_1, \dots, i_k \in I$ such that $K \subseteq \cup_{j=1}^k U_{i_j}$. Evidently $K \subseteq \cup_{j=1}^k U_{i_j}'$ and so K is a compact subset of $\mathbb{R}^n$.

Next suppose that K is a compact subset of $\mathbb{R}^n$. Let $(U_i)_{i\in I}$ be a family of relatively open subsets of A such that $K \subseteq \cup_{i\in I} U_i$. By Proposition 1.2.50 let $(U_i')_{i\in I}$ be a family of open subsets of $\mathbb{R}^n$ such that $U_i = U_i' \cap A$ for every $i \in I$. Clearly $K \subseteq \cup_{i\in I} U_i'$. Since K is compact there exists $i_1, \dots, i_k \in I$ such that $K \subseteq \cup_{j=1}^k U_{i_j}'$. By Proposition I-1.1.7 we have

$$K \subseteq (\cup_{j=1}^k U_{i_j}') \cap A = \cup_{j=1}^k U_{i_j},$$

showing that K is relatively compact. ■

[4]This is not the usual meaning given to the words “relatively compact.” Most often, “relatively compact” is used to refer to what we call “precompact.” However, we think that the meaning we give to “relatively compact” here is far more natural.

Let us use the preceding result to characterise relatively compact subsets of $\mathbb{Q} \subseteq \mathbb{R}$.

1.2.60 Examples (Relatively compact subsets of $\mathbb{Q}$) Let us examine some properties of relatively compact subsets of $\mathbb{Q}$.

1. A finite subset $K \subseteq \mathbb{Q}$ is easily seen to be compact; see Exercise 1.2.9.
2. We claim that if $K \subseteq \mathbb{Q}$ is compact then K has an isolated point, i.e., there exists a point $q \in K$ and a neighbourhood U of q such that $U \cap K = \{q\}$. Indeed, suppose that K has no isolated points. Since finite subsets of $\mathbb{Q}$ are isolated and compact, we can consider the case when K is countable. Let us enumerate the points in K as $K = \{q_j\}_{j\in\mathbb{Z}_{>0}}$. Let us take $j_1 = 1$ and $p_1 = q_{j_1}$. As we saw in (1.2.51)–5, we can find a sufficiently small relatively closed relative neighbourhood U_1 of p_1 such that $K \not\subseteq U_1$. The subset $V_1 = K \setminus U_1$ is relatively open and relatively closed since U_1 is relatively open and relatively closed. Moreover, V_1 cannot be finite since K has no isolated points. Denote
$$j_2 = \min\{j \in \mathbb{Z}_{>0} \mid j > 1,\ q_j \notin U_1\}$$
and $p_2 = q_{j_2}$. Since $p_2 \in V_1$ and since V_1 is relatively open, by Proposition 1.2.50 we have that
$$\inf\{|p_2 - q| \mid q \in U_1\} > 0.$$
Therefore, again using the construction of (1.2.51)–5, there exists a sufficiently small relatively closed relative neighbourhood U_2 of p_2 such that $U_2 \cap U_1 = \emptyset$ and $V_1 \not\subseteq U_2$. Then define $V_2 = V_1 \setminus U_2$. Again, since K has no isolated points, V_2 is not finite. This process can be carried out to define a sequence $(j_k)_{j\in\mathbb{Z}_{>0}}$ of positive integers, a sequence $(p_k)_{k\in\mathbb{Z}_{>0}}$ of elements of K, and a sequence $(U_k)_{k\in\mathbb{Z}_{>0}}$ of pairwise disjoint subsets of K that are relatively open. We claim that $K \subseteq \cup_{k\in\mathbb{Z}_{>0}} U_k$. Indeed, suppose that $q_m \in K$ but $q_m \notin \cup_{k\in\mathbb{Z}_{>0}} U_k$ for some $m \in \mathbb{Z}_{>0}$. Denote
$$k_m = \min\{k \in \mathbb{Z}_{>0} \mid j_k > m\}.$$
Note that $q_m \notin \cup_{k=1}^{k_m-1} U_k$. However, the definition of k_m is that it is the smallest integer such that $q_{k_m} \notin \cup_{k=1}^{k_m-1} U_k$. Since $m < k_m$, we arrive at a contradiction. Thus the relatively open sets $(U_k)_{k\in\mathbb{Z}_{>0}}$ cover K, but clearly admit no finite subcover since they are pairwise disjoint. Thus subsets of $\mathbb{Q}$ with no isolated points cannot be compact.
3. The question raised by the previous two points is: "Are all relatively compact subsets of $\mathbb{Q}$ comprised only of isolated points, or, equivalently, are all relatively compact subsets of $\mathbb{Q}$ finite?" The answer is, "No." For example, the set
$$K = \{0\} \cup \{\tfrac{1}{k} \mid k \in \mathbb{Z}_{>0}\}$$
is relatively compact. To see this, by Proposition 1.2.59 and the Heine–Borel Theorem we need only show that it is closed and bounded as a subset of $\mathbb{R}$. It

is clearly bounded. By Proposition 1.2.26 we can easily that $\mathrm{cl}(K) = K$ and so K is closed. Thus this is an example of a relatively compact subset of $\mathbb{Q}$ with a nonisolated point, since 0 is not isolated.

4. Finally, let us show that there are relatively compact subsets of $\mathbb{Q}$ having infinitely many nonisolated points. Let us define

$$K = \{0\} \cup \{\tfrac{1}{k} \mid k \in \mathbb{Z}_{>0}\} \cup \{\tfrac{1}{j} + \tfrac{1}{k} \mid j,k \in \mathbb{Z}_{>0}\}.$$

Let us first identity the accumulation points and limit of K.

1 Lemma *The set of accumulation points of* K *is* $\{0\} \cup \{\frac{1}{k}\}_{k\in\mathbb{Z}_{>0}}$ *and the set of limit points of* K *is* K.

Proof Let the sequence $(\frac{1}{j_l} + \frac{1}{k_l})_{l\in\mathbb{Z}_{>0}}$ converge to $r \in \mathbb{R}$. The sequence $(\frac{1}{j_l})_{l\in\mathbb{Z}_{>0}}$ has a convergent subsequence $(\frac{1}{j_{l_m}})_{m\in\mathbb{Z}_{>0}}$ since it is bounded (see Proposition I-2.3.4). Since

$$\lim_{m\to\infty} \frac{1}{k_{l_m}} = r - \lim_{m\to\infty} \frac{1}{j_{l_m}},$$

the subsequence $(\frac{1}{k_{l_m}})_{m\in\mathbb{Z}_{>0}}$ also converges. By Proposition I-2.3.23 we have

$$\lim_{m\to\infty} \frac{1}{j_{l_m}} = \frac{1}{\lim_{m\to\infty} j_{l_m}}.$$

There are two possible cases.

1. $\lim_{m\to\infty} j_{l_m} = \infty$: In this case $\lim_{m\to\infty} \frac{1}{j_{l_m}} = 0$.
2. $\lim_{m\to\infty} j_{l_m} \neq \infty$: In this case there must be a positive integer j_0 such that $\lim_{m\to\infty j_{l_m}} = j_0$. Thus $\lim_{m\to\infty} \frac{1}{j_{l_m}} = \frac{1}{j_0}$.

Similarly, either $\lim_{m\to\infty} \frac{1}{k_{l_m}} = 0$ or there exists $k_0 \in \mathbb{Z}_{>0}$ such that $\lim_{m\to\infty} \frac{1}{k_{l_m}} = \frac{1}{k_0}$. Thus, in all cases,

$$\lim_{m\to\infty} \Big(\frac{1}{j_{l_m}} + \frac{1}{k_{l_m}}\Big) \in K$$

and we conclude that the set of limit points of K is K, as claimed. The accumulation points of K arise as limits of sequences that are not eventually constant. From the various cases presented above, the converging subsequences that are not eventually constant arise when one or both of the cases

$$\lim_{m\to\infty} j_{l_m} = \infty, \quad \lim_{m\to\infty} k_{l_m} = \infty$$

occur. In this case,

$$\lim_{m\to\infty} \Big(\frac{1}{j_{l_m}} + \frac{1}{k_{l_m}}\Big) \in \{0\} \cup \{\tfrac{1}{k}\}_{k\in\mathbb{Z}_{>0}},$$

as desired. ▼

The lemma allows us to conclude that the set of nonisolated points of K is exactly $\{0\} \cup \{\frac{1}{k}\}_{k\in\mathbb{Z}_{>0}}$. Thus the set of nonisolated points is infinite. Moreover, K is closed (because every point is a limit point) and bounded, and hence compact.

3. We claim that relatively compact subsets of $\mathbb{Q}$ have empty relative interior. If a subset of $\mathbb{Q}$ has an nonempty interior, it must contain a nonempty relatively open subset. This means that it must contain a subset of the form $I \cap \mathbb{Q}$ where I is an open interval.

   We claim that if $I \subseteq \mathbb{R}$ is an interval with a nonempty interior, then $I \cap \mathbb{Q}$ is not relatively compact in $\mathbb{Q}$. By the Bolzano–Weierstrass Theorem it suffices to show that there are sequences in $I \cap \mathbb{Q}$ that contain no subsequences converging in $I \cap \mathbb{Q}$. To exhibit such a sequence, let $r \in \mathrm{int}(I)$ be irrational and let $(q_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $I \cap \mathbb{Q}$ converging to r (by Proposition I-2.2.15). Any subsequence of this sequence also converges to $r \notin I \cap \mathbb{Q}$. •

While we are talking about compactness, let us characterise compact subsets of $\mathbb{R}^n$ using relatively open and closed sets.

1.2.61 Proposition (Characterisation of compactness in terms of relatively open sets) *A subset $K \subseteq \mathbb{R}^n$ is compact if and only if, for every collection $(U_a)_{a\in A}$ of relatively open subsets of K for which $K = \cup_{a\in A} U_a$, there exists $a_1, \dots, a_k \in A$ such that $K = \cup_{j=1}^k U_{a_j}$.*

Proof First suppose that K is compact. For a collection $(U_a)_{a\in A}$ of relatively open subsets of K that covers K, let $V_a \subseteq \mathbb{R}^n$ be open and such that $U_a = K \cap V_a$, $a \in A$, using Proposition 1.2.50. Thus $(V_a)_{a\in A}$ is an open cover of K. Since K is compact there exists $a_1, \dots, a_k \in A$ such that $K \subseteq \cup_{j=1}^k V_{a_j}$. Thus

$$K = \cup_{j=1}^k (V_{a_j} \cap K) = \cup_{j=1}^k U_{a_j},$$

as desired.

For the converse, let $(V_a)_{a\in A}$ be an open cover of K so that $(U_a = V_a \cap K)_{a\in A}$ is a cover of K by relatively open sets by Proposition 1.2.50. Thus there exists $a_1, \dots, a_k \in A$ such that $K = \cup_{j=1}^k U_{a_j}$ and so $K \subseteq \cup_{j=1}^k V_{a_j}$. That is, K is compact. ■

It is also possible to characterise compactness deftly in terms of relatively closed sets.

1.2.62 Definition (Finite intersection property) Let $A \subseteq \mathbb{R}^n$ and let $(B_j)_{j\in J}$ be a family of subset of A. The family has the ***finite intersection property*** if, for any finite subset $\{j_1, \dots, j_k\} \subseteq J$, the set $\cap_{m=1}^k B_{j_m} \neq \emptyset$. •

We then have the following characterisation of compact sets.

1.2.63 Proposition (Compactness and the finite intersection property) *A subset $K \subseteq \mathbb{R}^n$ is compact if and only if every family $(C_j)_{j\in J}$ of relatively closed subsets of K with the finite intersection property has the property that $\cap_{j\in J} C_j \neq \emptyset$.*

Proof Suppose that K is compact. Let $(C_j)_{j\in J}$ be a family of closed sets with the finite intersection property and suppose that $\cap_{j\in J} C_j = \varnothing$. Then we have

$$K = K \setminus (\cap_{j\in J} C_j) = \cup_{j\in J}(K \setminus C_j)$$

by DeMorgan's Laws. Then, since K is compact, there exists $j_1, \dots, j_k \in J$ such that $K = \cup_{m=1}^k (A \setminus C_{j_m})$. But this gives $K = K \setminus (\cap_{m=1}^k C_{j_m})$, again by DeMorgan's Laws. This means that $\cap_{m=1}^k C_{j_m} = \varnothing$, contradicting the finite intersection property of $(C_j)_{j\in J}$.

Conversely, suppose that $(U_j)_{j\in J}$ is an open cover of K and suppose that there is no finite subcover of this open cover. We claim that the family $(K \setminus U_j)_{j\in J}$ has the finite intersection property. Indeed, let $\{j_1, \dots, j_k\} \subseteq J$ so that

$$\cap_{m=1}^k (K \setminus U_{j_m}) = K \setminus (\cup_{m=1}^k U_{j_m}) \neq \varnothing$$

since $(U_j)_{j\in J}$ possesses no finite subcover. Now, for any finite subset $\{j_1, \dots, j_k\} \subseteq J$ we have

$$\varnothing \neq \cap_{j\in J}(K \setminus U_j) = K \setminus (\cup_{j\in J} U_j)$$

since $(K \setminus U_j)_{j\in J}$ has the finite intersection property. But this contradicts the fact that $(U_j)_{j\in J}$ covers K. ■

Connectedness using relative constructions

The use of relatively open and closed sets provides an elegant characterisation of connectedness. This characterisation will generalise to the notion of connectedness for general topological spaces in Section III-1.7.

1.2.64 Theorem (Characterisation of connectedness in terms of relative topology) *A subset $A \subseteq \mathbb{R}^n$ is connected if and only if the only subsets of A that are both relatively open and relatively closed in A are $\varnothing$ and A.*

Proof First suppose that A is disconnected so that $A = S \cup T$ for nonempty sets S and T with $\mathrm{cl}(S) \cap T = \varnothing$ and $S \cap \mathrm{cl}(T) = \varnothing$. Note that $S = A \cap \mathrm{cl}(S)$ since $S \subseteq \mathrm{cl}(S)$ and $\mathrm{cl}(S) \cap T = \varnothing$. By Proposition 1.2.50 this means that S is relative closed. In like manner T is relatively closed. Thus both S and T are also relatively open.

Now suppose that $S \subseteq A$ is relatively open and relatively closed, and that $S \neq A$ and $S \neq \varnothing$. Then $A = S \cup (A \setminus S)$ where S and $T \triangleq A \setminus S$ are both relatively open and relatively closed. We claim that $\mathrm{cl}(S) \cap T = \varnothing$. Indeed, if $\boldsymbol{x} \in T$ there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\epsilon, \boldsymbol{x}) \cap A \subseteq T$ since T is relatively open. Since $S \cap T = \varnothing$ this implies that $\mathsf{B}^n(\epsilon, \boldsymbol{x}) \cap S = \varnothing$. By the analogue of Proposition 1.2.26 for the relative closure this implies that $\boldsymbol{x} \notin \mathrm{cl}(S)$. Thus we indeed have $\mathrm{cl}(S) \cap T = \varnothing$. The same argument gives $S \cap \mathrm{cl}(T) = \varnothing$ and so A is disconnected. ■

1.2.9 Local compactness

In this section we introduce the important idea of local compactness. This property turns out to be exactly what is needed for certain constructions. Our investigation here will be rather elementary. In Section III-1.11 we give a deeper treatment of local compactness.

We begin with the definition.

1.2.65 Definition (Locally compact) A subset $A \subseteq \mathbb{R}^n$ is ***locally compact*** if, for every $\boldsymbol{x} \in A$, there exists a relative neighbourhood $U \subseteq A$ of $\boldsymbol{x}$ such that $\mathrm{cl}_A(U)$ is a relatively compact subset of A. •

Let us give some examples and counterexamples.

1.2.66 Examples (Locally compact subsets)

1. We claim that every open subset U of $\mathbb{R}^n$ is locally compact. Indeed, let $\boldsymbol{x} \in U$ and, since U is open, let $\epsilon \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^n(\epsilon, \boldsymbol{x}) \subseteq U$. Then $\mathsf{B}(\frac{\epsilon}{2}, \boldsymbol{x}) \subseteq U$ is a relative neighbourhood of $\boldsymbol{x}$ whose closure is a relatively compact subset of U.
2. We claim that every closed subset A of $\mathbb{R}^n$ is locally compact. Indeed, let $\boldsymbol{x} \in A$, let $\epsilon \in \mathbb{R}_{>0}$, and denote $U = \mathsf{B}^n(\epsilon, \boldsymbol{x}) \cap A$, noting that U is a relative neighbourhood of $\boldsymbol{x}$ by Proposition 1.2.50. We claim that
$$\mathrm{cl}_A(U) = \overline{\mathsf{B}}^n(\epsilon, \boldsymbol{x}) \cap A. \tag{1.8}$$
Note that
$$U \subseteq \overline{\mathsf{B}}^n(\epsilon, \boldsymbol{x}) \cap A \subseteq A,$$
the latter inclusion holding since A is closed. Thus
$$\overline{\mathsf{B}}^n(\epsilon, \boldsymbol{x}) \cap A \subseteq \mathrm{cl}_A(U)$$
by definition of $\mathrm{cl}_A(U)$. The opposite inclusion holds by Proposition 1.2.28. Thus we have (1.8). By Proposition 1.2.59 we have that $\mathrm{cl}_A(U)$ is relatively compact in A. This shows that U is a relative neighbourhood of $\boldsymbol{x}$ possessing a relatively compact closure.
3. We claim that the subset $\mathbb{Q} \subseteq \mathbb{R}$ is not locally compact. Indeed, we showed in Example 1.2.60–3 that all relatively compact subsets of $\mathbb{Q}$ have empty relative interior.
4. Let
$$A = \{(0,0)\} \cup \{(x,y) \in \mathbb{R}^2 \mid x \in \mathbb{R}_{>0}\} \subseteq \mathbb{R}^2$$
(see Figure 1.6). We claim that A is not locally compact. Indeed, let $U = U' \cap A$ (with $U' \subseteq \mathbb{R}^2$ a neighbourhood of $(0,0)$) be a relative neighbourhood of $(0,0)$. We claim that $\mathrm{cl}_A(U)$ is not compact.
Let $\epsilon \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^2(\epsilon, (0,0)) \subseteq U'$. For $j \in \mathbb{Z}_{>0}$ define an open subset $U'_j \subseteq \mathbb{R}^2$ by
$$U'_j = (\{(x,y) \in \mathbb{R}^2 \mid x > j^{-1}y,\ y \geq 0\} \\ \cup \{(x,y) \in \mathbb{R}^2 \mid x > -j^{-1}y,\ y \leq 0\}) \cap \mathsf{B}^2(3\epsilon, (0,0))$$
(see Figure 1.7 for a depiction). Also let
$$U'_0 = \mathsf{B}^2(\frac{\epsilon}{2}, (0,0)), \quad V' = \mathbb{R}^2 \setminus \overline{\mathsf{B}}^2(2\epsilon, (0,0)).$$

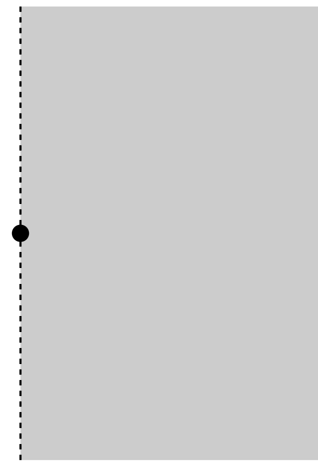

Figure 1.6 A subset of $\mathbb{R}^2$ that is not locally compact

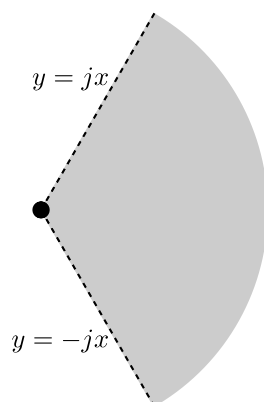

Figure 1.7 The open set U'_j

Note that

$$A \subseteq V' \cup U'_0 \cup_{j\in\mathbb{Z}_{>0}} U'_j.$$

Thus, if we define $U_j = U'_j \cap A$, $j \in \mathbb{Z}_{>0}$, $U_0 = U'_0 \cap A$, and $V = V' \cap A$, then

$$\mathrm{cl}_A(U) \subseteq V \cup U_0 \cup_{j\in\mathbb{Z}_{>0}} U_j.$$

We claim that there is no finite subset of the relatively open cover $\mathscr{O} = \{V\} \cup \{U_0\} \cup \{U_j\}_{j\in\mathbb{Z}_{>0}}$ that covers $\mathrm{cl}_A(U)$. Indeed, note that $\overline{\mathsf{B}}^2(\epsilon, (0,0)) \cap A \subseteq \mathrm{cl}_A(U)$. Therefore, any subset of $\mathscr{O}$ covering $\mathrm{cl}_A(U)$ must also cover $\overline{\mathsf{B}}^2(\epsilon, (0,0)) \cap A$. This, however, implies that all of the subsets U_j, $j \in \mathbb{Z}_{>0}$, must be contained in any subcover covering $\mathrm{cl}_A(U)$, and this ensures that $\mathrm{cl}_A(U)$ is not compact. •

1.2.10 Products of subsets

Next we consider subsets of Cartesian products of Euclidean spaces. Specifically, we consider sets of the form $A_1 \times \cdots \times A_k$ where $A_j \subseteq \mathbb{R}^{n_j}$, $j \in \{1, \dots, k\}$. For such subsets we shall give their properties in terms of properties of the subsets $A_1, \dots, A_k$. In studying these sets we make the natural identification of $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ with $\mathbb{R}^{n_1+\cdots++n_k}$ given by

$$\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k} \ni ((x_{1,1}, \dots, x_{1,n_1}), \dots, (x_{k,1}, \dots, x_{k,n_k})) \mapsto (x_{1,1}, \dots, x_{1,n_1}, x_{2,1}, \dots, x_{k-1,n_{k-1}}, x_{k,1}, \dots, x_{k,n_k}) \in \mathbb{R}^{n_1+\cdots+n_k}.$$

Thus, on $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ we shall use the Euclidean norm $\|\cdot\|_{\mathbb{R}^{n_1+\cdots+n_k}}$, and notions of openness, closedness, etc., will be derived from this. It is useful to relate this norm to the separate norms for $\mathbb{R}^{n_1}, \dots, \mathbb{R}^{n_k}$.

1.2.67 Lemma *For* $\mathbf{x}_j \in \mathbb{R}^{n_j}$, $j \in \{1, \dots, k\}$, *we have*

$$\|\mathbf{x}_1\|_{\mathbb{R}^{n_1}} + \cdots + \|\mathbf{x}_k\|_{\mathbb{R}^{n_k}} \leq \sqrt{k}\|(\mathbf{x}_1, \dots, \mathbf{x}_k)\|_{\mathbb{R}^{n_1+\cdots+n+k}},$$
$$\|(\mathbf{x}_1, \dots, \mathbf{x}_k)\|_{\mathbb{R}^{n_1+\cdots+n_k}} \leq \|\mathbf{x}_1\|_{\mathbb{R}^{n_1}} + \cdots + \|\mathbf{x}_k\|_{\mathbb{R}^{n_k}}.$$

Proof Define

$$\delta_j = \|\boldsymbol{x}_j\|_{\mathbb{R}^{n_j}} - \tfrac{1}{k}(\|\boldsymbol{x}_1\|_{\mathbb{R}^{n_1}} + \cdots + \|\boldsymbol{x}_k\|_{\mathbb{R}^{n_k}}), \qquad j \in \{1, \dots, k\},$$

noting that $\delta_1 + \cdots + \delta_k = 0$ and that

$$\|\boldsymbol{x}_j\|_{\mathbb{R}^{n_j}} = \tfrac{1}{k}(\|\boldsymbol{x}_1\|_{\mathbb{R}^{n_1}} + \cdots + \|\boldsymbol{x}_k\|_{\mathbb{R}^{n_k}}) + \delta_j, \qquad j \in \{1, \dots, k\}.$$

A computation then gives

$$\begin{aligned}\|(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_{\mathbb{R}^{n_1+\cdots+n_k}} &= (\|\boldsymbol{x}_1\|^2_{\mathbb{R}^{n_1}} + \cdots + \|\boldsymbol{x}_k\|^2_{\mathbb{R}^{n_k}})^{1/2} \\ &= \Big(\tfrac{1}{k}(\|\boldsymbol{x}_1\|_{\mathbb{R}^{n_1}} + \cdots + \|\boldsymbol{x}_k\|_{\mathbb{R}^{n_k}})^2 + \delta_1^2 + \cdots + \delta_k^2\Big)^{1/2}\end{aligned}$$

which gives

$$\|(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_{\mathbb{R}^{n_1+\cdots+n_k}} \geq \tfrac{1}{\sqrt{k}}(\|\boldsymbol{x}_1\|_{\mathbb{R}^{n_1}} + \cdots + \|\boldsymbol{x}_k\|_{\mathbb{R}^{n_k}}),$$

as desired.

For the second inequality we have

$$\begin{aligned}\|(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)\|_{\mathbb{R}^{n_1+\cdots+n_k}} &= \|(\boldsymbol{x}_1, \boldsymbol{0}, \dots, \boldsymbol{0}) + (\boldsymbol{0}, , \dots, \boldsymbol{0}, \boldsymbol{x}_k)\|_{\mathbb{R}^{n_1+\cdots+n_k}} \\ &\leq \|(\boldsymbol{x}_1, \boldsymbol{0}, \dots, \boldsymbol{0})\|_{\mathbb{R}^{n_1+\cdots+n_k}} + \|(\boldsymbol{0}, \dots, \boldsymbol{0}, \boldsymbol{x}_k)\|_{\mathbb{R}^{n_1+\cdots+n_k}} \\ &= \|\boldsymbol{x}_1\|_{\mathbb{R}^{n_1}} + \cdots + \|\boldsymbol{x}_k\|_{\mathbb{R}^{n_k}},\end{aligned}$$

as desired. ■

Using these inequalities one can directly check that

$$\begin{aligned}\mathsf{B}^{n_1}(\epsilon, \boldsymbol{x}_1) \times \cdots \times \mathsf{B}^{n_k}(\epsilon, \boldsymbol{x}_k) &\subseteq \mathsf{B}^{n_1+\cdots+n_k}(k\epsilon, (\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)), \\ \mathsf{B}^{n_1+\cdots+n_k}(\epsilon, (\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)) &\subseteq \mathsf{B}^{n_1}(\sqrt{k}\epsilon, \boldsymbol{x}_1) \times \cdots \times \mathsf{B}^{n_k}(\sqrt{k}\epsilon, \boldsymbol{x}_k).\end{aligned} \tag{1.9}$$

The following theorem states the results in which we are interested.

1.2.68 Theorem (Properties of products derived from properties of components) *If* $A_j \subseteq \mathbb{R}^{n_j}$, $j \in \{1, \dots, k\}$, *then the following statements hold:*

(i) $A_1 \times \cdots \times A_k$ *is open if and only if each of the sets* A_j, $j \in \{1, \dots, k\}$, *is open;*

(ii) $A_1 \times \cdots \times A_k$ *is closed if and only if each of the sets* A_j, $j \in \{1, \dots, k\}$, *is closed;*

(iii) $A_1 \times \cdots \times A_k$ *is compact if and only if each of the sets* A_j, $j \in \{1, \dots, k\}$, *is compact;*

(iv) $A_1 \times \cdots \times A_k$ *is connected if and only if each of the sets* A_j, $j \in \{1, \dots, k\}$, *is connected.*

Proof By an elementary induction argument in each case it suffices to prove the theorem in the case when $k = 2$. In this case, for simplicity of notation, we denote $n_1 = m$ and $n_2 = n$, and write a typical point in $\mathbb{R}^m \times \mathbb{R}^n$ as $(\boldsymbol{x}, \boldsymbol{y})$.

(i) Suppose that $A \times B$ is open and let $\boldsymbol{x}_0 \in A$ and $\boldsymbol{y}_0 \in B$. Since $A \times B$ is open there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^{m+n}(2\epsilon, (\boldsymbol{x}_0, \boldsymbol{y}_0)) \subseteq A \times B$. By (1.9) it follows that $\mathsf{B}^m(\epsilon, \boldsymbol{x}_0) \times \mathsf{B}^n(\epsilon, \boldsymbol{y}_0) \subseteq A \times B$ and so $\mathsf{B}^m(\epsilon, \boldsymbol{x}_0) \subseteq A$ and $\mathsf{B}^n(\epsilon, \boldsymbol{y}_0) \subseteq B$. Thus both A and B are open.

Now suppose that A and B are open and let $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in A \times B$. Let $\epsilon \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^m(\sqrt{2}\epsilon, \boldsymbol{x}_0) \subseteq A$ and $\mathsf{B}^n(\sqrt{2}\epsilon, \boldsymbol{y}_0) \subseteq B$. Then $\mathsf{B}^m(\sqrt{2}\epsilon, \boldsymbol{x}_0) \times \mathsf{B}^n(\sqrt{2}\epsilon, \boldsymbol{y}_0) \subseteq A \times B$. By (1.9) it follows that $\mathsf{B}^{m+n}(\epsilon, (\boldsymbol{x}_0, \boldsymbol{y}_0)) \subseteq A \times B$ and so $A \times B$ is open.

(ii) Suppose that $A \times B$ is closed and let $(\boldsymbol{x}_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in A that converges to some $\boldsymbol{x}_0 \in \mathbb{R}^m$. We will show that $\boldsymbol{x}_0 \in A$ which will show that A is closed by Proposition 1.2.26. Note that for $\boldsymbol{y}_0 \in B$ the sequence $((\boldsymbol{x}_j, \boldsymbol{y}_0))_{j \in \mathbb{Z}}$ is in $A \times B$. Moreover, since

$$\|(\boldsymbol{x}_j, \boldsymbol{y}_0) - (\boldsymbol{x}_0, \boldsymbol{y}_0)\|_{\mathbb{R}^{m+n}} = \|\boldsymbol{x}_j - \boldsymbol{x}_0\|_{\mathbb{R}^m},$$

the sequence converges to $(\boldsymbol{x}_0, \boldsymbol{y}_0)$. Since $A \times B$ is closed it follows that $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in A \times B$ and so $\boldsymbol{x}_0 \in B$, as desired.

Conversely, suppose that both A and B are closed. Then, by part (i), $(\mathbb{R}^m \setminus A) \times \mathbb{R}^n$ and $\mathbb{R}^m \times (\mathbb{R}^n \setminus B)$ are open and so too is their union. However,

$$(\mathbb{R}^m \times \mathbb{R}^n) \setminus (A \times B) = ((\mathbb{R}^m \setminus A) \times \mathbb{R}^n) \cup (\mathbb{R}^m \times (\mathbb{R}^n \setminus B))$$

and so $(\mathbb{R}^m \times \mathbb{R}^n) \setminus (A \times B)$ is open. Thus $A \times B$ is closed.

(iii) Suppose that $A \times B$ is compact, i.e., is closed and bounded by the Heine–Borel Theorem. Then A and B are closed by part (ii). Moreover, A and B are also bounded. Indeed, suppose that, say, A were unbounded and let $M \in \mathbb{R}_{>0}$. Then there exists $\boldsymbol{x}_1, \boldsymbol{x}_2 \in A$ such that $\|\boldsymbol{x}_1 + \boldsymbol{x}_2\|_{\mathbb{R}^m} \geq M$. Therefore, for $\boldsymbol{y} \in B$ we have

$$\|(\boldsymbol{x}_1, \boldsymbol{y}) - (\boldsymbol{x}_2, \boldsymbol{y})\|_{\mathbb{R}^{m+n}} \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathbb{R}^m} \geq M,$$

giving $A \times B$ as unbounded since $M \in \mathbb{R}_{>0}$ is arbitrary. Thus both A and B are closed and bounded, and so compact by the Heine–Borel Theorem.

Conversely, suppose that A and B are compact, i.e., closed and bounded by the Heine–Borel Theorem. Then $A \times B$ is closed by part (ii). To see that $A \times B$ is bounded, let $M \in \mathbb{R}_{>0}$ be such that

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathbb{R}^m} < \tfrac{M}{2}, \quad \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|_{\mathbb{R}^n} < \tfrac{M}{2}$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in A$ and $\boldsymbol{y}_1, \boldsymbol{y}_2 \in B$. Then

$$\|(\boldsymbol{x}_1, \boldsymbol{y}_1) - (\boldsymbol{x}_2, \boldsymbol{y}_2)\|_{\mathbb{R}^{m+n}} \leq \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathbb{R}^m} + \|\boldsymbol{y}_1 - \boldsymbol{y}_2\|_{\mathbb{R}^n} < M,$$

using Lemma 1.2.67.

(iv) Suppose that A is not connected. Then $A = S \cup T$ where S and T are nonempty sets satisfying $\mathrm{cl}(S) \cap T = \emptyset$ and $S \cap \mathrm{cl}(T) = \emptyset$. Then $A \times B = (S \times B) \cup (T \times B)$. We claim that $\mathrm{cl}(S \times B) \cap (T \times B) = \emptyset$. Let $((\boldsymbol{x}_j, \boldsymbol{y}_j))_{j\in\mathbb{Z}_{>0}}$ be a sequence in $S \times B$ converging to $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathrm{cl}(S \times B)$. It is evident that $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}} \subseteq S$ converges to $\boldsymbol{x}_0$ and so $\boldsymbol{x}_0 \in \mathrm{cl}(S)$. Therefore, $\boldsymbol{x}_0 \notin T$ and so $(\boldsymbol{x}_0, \boldsymbol{y}_0) \notin T \times B$. Thus $\mathrm{cl}(S \times B) \cap (T \times B) = \emptyset$, as claimed. We similarly show that $(S \times B) \cap \mathrm{cl}(T \times B) = \emptyset$. This shows that $A \times B$ is disconnected if A is disconnected. Similarly one shows that $A \times B$ is disconnected if B is disconnected.

Now suppose that A and B are connected but that $A \times B$ are disconnected. Thus we suppose that $A \times B = S \cup T$ for nonempty sets S and T such that $\mathrm{cl}(S) \cap T = \emptyset$ and $S \cap \mathrm{cl}(T) = \emptyset$. Let $(\boldsymbol{x}_1, \boldsymbol{y}_1) \in S$ and $(\boldsymbol{x}_2, \boldsymbol{y}_2) \in T$. We claim that $\{\boldsymbol{x}_1\} \times B$ and $A \times \{\boldsymbol{y}_2\}$ are connected. This is clear since if, for example, $\{\boldsymbol{x}_1\} \times B$ is disconnected then B is disconnected. Now note that $(\{\boldsymbol{x}_1\}\times B)\cap(A\times\{\boldsymbol{y}_2\}) \neq \emptyset$ since it contains the point $(\boldsymbol{x}_2, \boldsymbol{y}_1)$. By Exercise 1.2.5 it follows that $X = (\{\boldsymbol{x}_1\} \times B) \cup (A \times \{\boldsymbol{y}_2\})$ is connected. However, this is a contradiction since the disconnectedness of $A \times B$ implies that

$$X = (X \cap S) \cup (X \cap T)$$

where $\mathrm{cl}(X \cap S) \cap (X \cap T) = \emptyset$ and $(X \cap S) \cap \mathrm{cl}(X \cap T) = \emptyset$. Thus it must be that $A \times B$ is connected. ■

These characterisations of products allows us to prove the following result.

1.2.69 Proposition (Interior, closure, and boundary of products) *If* $A \subseteq \mathbb{R}^m$ *and* $B \subseteq \mathbb{R}^n$ *then*

(i) $\mathrm{int}(A \times B) = \mathrm{int}(A) \times \mathrm{int}(B)$,

(ii) $\mathrm{cl}(A \times B) = \mathrm{cl}(A) \times \mathrm{cl}(B)$, *and*

(iii) $\mathrm{bd}(A \times B) = (\mathrm{bd}(A) \times \mathrm{cl}(B)) \cup (\mathrm{cl}(A) \times \mathrm{bd}(B))$.

Proof (i) Since $\mathrm{int}(A) \times \mathrm{int}(B) \subseteq A \times B$ we have $\mathrm{int}(A) \times \mathrm{int}(B) \subseteq \mathrm{int}(A \times B)$ by the definition of interior. Now let $(\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{int}(A \times B)$. Then there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^{m+n}(2\epsilon, (\boldsymbol{x}, \boldsymbol{y})) \subseteq A \times B$. By (1.9) it then follows that

$$\mathsf{B}^m(\epsilon, \boldsymbol{x}) \times \mathsf{B}^n(\epsilon, \boldsymbol{y}) \subseteq A \times B,$$

and so $\mathsf{B}^m(\epsilon, \boldsymbol{x}) \subseteq A$ and $\mathsf{B}^m(\epsilon, \boldsymbol{y}) \subseteq B$. Thus $\boldsymbol{x} \in \mathrm{int}(A)$ and $\boldsymbol{y} \in \mathrm{int}(B)$.

(ii) Since $A \times B \subseteq \mathrm{cl}(A) \times \mathrm{cl}(B)$ and since $\mathrm{cl}(A) \times \mathrm{cl}(B)$ is closed by Theorem 1.2.68, it follows that $\mathrm{cl}(A \times B) \subseteq \mathrm{cl}(A) \times \mathrm{cl}(B)$. Now let $(\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{cl}(A) \times \mathrm{cl}(B)$. Then, for every $\epsilon \in \mathbb{R}_{>0}$ we have

$$\mathsf{B}^m(\tfrac{\epsilon}{2}, \boldsymbol{x}) \cap A \neq \emptyset, \quad \mathsf{B}^n(\tfrac{\epsilon}{2}, \boldsymbol{y}) \cap B \neq \emptyset \quad \Longrightarrow \quad (\mathsf{B}^m(\tfrac{\epsilon}{2}, \boldsymbol{x}) \times \mathsf{B}^n(\tfrac{\epsilon}{2}, \boldsymbol{y})) \cap (A \times B) \neq \emptyset.$$

Therefore, by (1.9) we have

$$\mathsf{B}^{m+n}(\epsilon, (\boldsymbol{x}, \boldsymbol{y})) \cap (A \times B) \neq \emptyset.$$

Thus $(\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{cl}(A \times B)$ since this holds for every $\epsilon \in \mathbb{R}_{>0}$.

(iii) Let $(x, y) \in \mathrm{bd}(A \times B)$. By Proposition 1.2.26 this means that for every $\epsilon \in \mathbb{R}_{>0}$

$$\mathsf{B}^{m+n}(\tfrac{\epsilon}{\sqrt{2}}, (x, y)) \cap (A \times B) \neq \emptyset, \quad \mathsf{B}^{m+n}(\tfrac{\epsilon}{\sqrt{2}}, (x, y)) \cap ((\mathbb{R}^n \times \mathbb{R}^m) \setminus (A \times B)) \neq \emptyset.$$

Therefore, by (1.9),

$$(\mathsf{B}^m(\epsilon, x) \times \mathsf{B}^n(\epsilon, y)) \cap (A \times B) \neq \emptyset, \quad (\mathsf{B}^m(\epsilon, x) \times \mathsf{B}^n(\epsilon, y)) \cap ((\mathbb{R}^n \times \mathbb{R}^m) \setminus (A \times B)) \neq \emptyset$$

for every $\epsilon \in \mathbb{R}_{>0}$. The condition

$$(\mathsf{B}^m(\epsilon, x) \times \mathsf{B}^n(\epsilon, y)) \cap (A \times B) \neq \emptyset$$

means that $x \in \mathrm{cl}(A)$ and $y \in \mathrm{cl}(B)$. Let us now these conditions along with the condition

$$(\mathsf{B}^m(\epsilon, x) \times \mathsf{B}^n(\epsilon, y)) \cap ((\mathbb{R}^n \times \mathbb{R}^m) \setminus (A \times B)) \neq \emptyset, \qquad \epsilon \in \mathbb{R}_{>0}.$$

This condition is exactly the condition that $(x, y) \in \mathrm{cl}((\mathbb{R}^n \times \mathbb{R}^m) \setminus (A \times B))$. We thus have the following possibilities.

1. $x \in \mathrm{cl}(A)$, $y \in \mathrm{cl}(B)$, $x \in A$, and $y \notin B$: In this case we must have $y \in \mathrm{bd}(\mathbb{R}^m \setminus B)$.
2. $x \in \mathrm{cl}(A)$, $y \in \mathrm{cl}(B)$, $x \in A$, and $y \in B$: In this case we cannot have $x \in \mathrm{int}(A)$ *and* $y \in \mathrm{int}(B)$ and so we must have either (a) $x \in \mathrm{bd}(A)$ and $y \in B$ or (b) $x \in B$ and $y \in \mathrm{bd}(B)$.
3. $x \in \mathrm{cl}(A)$, $y \in \mathrm{cl}(B)$, $x \notin A$, and $y \in A$: In this case we must have $x \in \mathrm{bd}(A)$.
4. $x \in \mathrm{cl}(A)$, $y \in \mathrm{cl}(B)$, $x \notin A$ and $y \notin B$: In this case we must have $x \in \mathrm{bd}(A)$ and $y \in \mathrm{bd}(B)$.

This means that we have either (1) $(x, y) \in \mathrm{bd}(A) \times \mathrm{cl}(B)$ or (2) $(x, y) \in \mathrm{cl}(A) \times \mathrm{bd}(B)$. Thus gives

$$\mathrm{bd}(A \times B) \subseteq (\mathrm{bd}(A) \times \mathrm{cl}(B)) \cup (\mathrm{cl}(A) \times \mathrm{bd}(B)).$$

Next suppose that $(x, y) \in \mathrm{bd}(A) \times \mathrm{cl}(B)$. This means that for every $\epsilon \in \mathbb{R}_{>0}$ the following sets are nonempty:

$$\mathsf{B}^m(\sqrt{2}\epsilon, x) \cap A, \quad \mathsf{B}^m(\sqrt{2}\epsilon, x) \cap (\mathbb{R}^n \setminus A), \quad \mathsf{B}^n(\sqrt{2}\epsilon, y) \cap B.$$

Thus take

$$x' \in \mathsf{B}^m(\sqrt{2}\epsilon, x) \cap A, \quad x'' \in \mathsf{B}^m(\sqrt{2}\epsilon, x) \cap (\mathbb{R}^n \setminus A), \quad y' \in \mathsf{B}^n(\sqrt{2}\epsilon, y) \cap B.$$

Then

$$\begin{aligned} & (x', y') \in (\mathsf{B}^m(\sqrt{2}\epsilon, x) \times \mathsf{B}^n(\sqrt{2}\epsilon, y)) \cap (A \times B) \\ \Longrightarrow \quad & (x', y') \in \mathsf{B}^{m+n}(\epsilon, (x, y)) \cap (A \times B). \end{aligned}$$

Also

$$\begin{aligned} & (x'', y') \in (\mathsf{B}^m(\sqrt{2}\epsilon, x) \times \mathsf{B}^n(\sqrt{2}\epsilon, y)) \cap ((\mathbb{R}^m \setminus A) \times B) \\ \Longrightarrow \quad & (x'', y') \in \mathsf{B}^{m+n}(\epsilon, (x, y)) \cap ((\mathbb{R}^m \setminus A) \times B) \\ \Longrightarrow \quad & (x'', y') \in \mathsf{B}^{m+n}(\epsilon, (x, y)) \cap ((\mathbb{R}^m \times \mathbb{R}^m) \setminus (A \times B)). \end{aligned}$$

In like manner one shows that if $(\boldsymbol{x}, \boldsymbol{y}) \in \mathrm{cl}(A) \times \mathrm{bd}(B)$ then

$$\mathsf{B}^{m+n}(\epsilon, (\boldsymbol{x}, \boldsymbol{y})) \cap (A \times B), \quad \mathsf{B}^{m+n}(\epsilon, (\boldsymbol{x}, \boldsymbol{y})) \cap ((\mathbb{R}^m \times \mathbb{R}^m) \setminus (A \times B))$$

are nonempty. That is, for every $\epsilon \in \mathbb{R}_{>0}$ the sets

$$\mathsf{B}^{m+n}(\epsilon, (\boldsymbol{x}, \boldsymbol{y})) \cap (A \times B), \quad \mathsf{B}^{m+n}(\epsilon, (\boldsymbol{x}, \boldsymbol{y})) \cap ((\mathbb{R}^n \times \mathbb{R}^m) \setminus (A \times B))$$

are nonempty. Thus $(\mathrm{bd}(A) \times \mathrm{cl}(B)) \cup (\mathrm{cl}(A) \times \mathrm{bd}(B)) \subseteq \mathrm{bd}(A \times B)$. ■

1.2.70 Remark (Finite Cartesian products) By an elementary induction argument, the first two statements in the preceding result carry over to finite Cartesian products of sets $A_1 \times \cdots \times A_k \subseteq \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$. The generalisation of the third statement is tedious, but straightforward, and left to the reader. •

1.2.11 Sets of measure zero

One can also talk about subsets of $\mathbb{R}^n$ which have measure zero. This is done in the obvious way, using balls instead of intervals to cover sets. While the "volume" (i.e., length) of an interval is obviously defined, the volume of a ball in $\mathbb{R}^n$ is not so easily deduced. Let us here just define this volume, saving for the calculations needed to verify the formula. Thus we denote by what?

$$\mathrm{vol}(\mathsf{B}^n(r, \mathbf{0})) = \frac{\pi^{n/2} r^n}{\Gamma(\frac{n}{2} + 1)} \tag{1.10}$$

volume of the ball of radius r, and we (reasonably) declare that the volume of a ball is independent of its centre. In the above formula, the function Γ (called, unsurprisingly, the **Γ-*function***) is defined by

$$\Gamma(x) = \int_0^\infty \mathrm{e}^{-y} y^{x-1}\,\mathrm{d}y.$$

This expression can be made more familiar by using property of the Γ-function that

$$\Gamma(\tfrac{k}{2} + 1) = \begin{cases} (\frac{k}{2})!, & k \text{ an even nonnegative integer,} \\ \frac{k!\pi^{1/2}}{2^k(\frac{k-1}{2})!}, & k \text{ an odd nonnegative integer.} \end{cases}$$

The reader is asked to explore some properties of the Γ-function in Exercise 1.2.16. In any case, we suppose that we know the volume of an n-dimensional ball.

With this we can make the following definition.

1.2.71 Definition (Set of measure zero) A subset $A \subseteq \mathbb{R}^n$ has ***measure zero*** if

$$\inf\Big\{ \sum_{j=1}^\infty \mathrm{vol}(\mathsf{B}^n(r_j, \boldsymbol{x}_j)) \,\Big|\, A \subseteq \bigcup_{j \in \mathbb{Z}_{>0}} \mathsf{B}^n(r_j, \boldsymbol{x}_j) \Big\} = 0. \quad \bullet$$

We refer the reader to Section I-2.5.6 for examples of sets of zero measure, some interesting and some not. Ideas concerning the generalisation to $\mathbb{R}^n$ of sets of measure zero are discussed in Section III-2.5.

1.2.12 Convergence in $\mathbb{R}^n$-nets and a second glimpse of Landau symbols

In Section I-2.3.7 we discussed convergence for generalisations of sequences where the index set is a subset of $\mathbb{R}$. In Section I-2.3.8 we used this general notion of convergence to define Landau symbols. In this section we make a further generalisation to the case of generalised sequences where the index set is a subset of $\mathbb{R}^n$.

We begin by defining the sorts of directed sets we consider. The definition we give is a generalisation of that given for $\mathbb{R}$ in Section I-2.3.7, but now we use the topology of $\mathbb{R}^n$ in a more fancy way.

1.2.72 Definition ($\mathbb{R}^n$-directed set) Let $A \subseteq \mathbb{R}^n$ and let $x_0 \in \mathbb{R}^n$.

(i) The ***$\mathbb{R}^n$-directed set in A at x_0*** is the family of subsets

$$D(A, x_0) = \{U \cap A \mid U \subseteq \mathbb{R}^n \text{ open}, \ x_0 \in U\}$$

with the partial order $\supseteq$.

(ii) The ***$\mathbb{R}$-directed set in A at ∞*** is the family of subsets

$$D(A, \infty) = \{U \cap A \mid U \subseteq \mathbb{R}^n \text{ open}, \ \mathbb{R}^n \setminus \overline{\mathsf{B}}^n(R, \mathbf{0}) \subseteq U \text{ for some } R \in \mathbb{R}_{>0}\}$$

with the partial order $\supseteq$. •

Let us verify that $\mathbb{R}^n$-directed sets are indeed directed sets.

1.2.73 Proposition ($\mathbb{R}^n$-directed sets are directed sets) *If $A \subseteq \mathbb{R}^n$ and if $x_0 \in \mathbb{R}^n$, then $(D(A, x_0), \supseteq)$ and $(D(A, \infty), \supseteq)$ are directed sets.*

Proof In the first case, let $U_1 \cap A, U_2 \cap A \in D(A, x_0)$ and note that, since $x_0 \in U_1 \cap A$ and $x_0 \in U_2 \cap A$, we have $U_1 \cap U_2$ is open and $x_0 \in (U_1 \cap U_2) \cap A$. Thus, $(U_1 \cap U_2) \cap A \in D(A, x_0)$ and

$$U_1 \cap A, U_2 \cap Q \supseteq (U_1 \cap U_2) \cap A.$$

In the second case, let $U_1 \cap A, U_2 \cap A \in D(A, \infty)$. Let $R_1, R_2 \in \mathbb{R}_{>0}$ be such that $\mathbb{R}^n \setminus \overline{\mathsf{B}}^n(R_1, \mathbf{0}) \subseteq U_1$ and $\mathbb{R}^n \setminus \overline{\mathsf{B}}^n(R_2, \mathbf{0}) \subseteq U_2$ and define $R = \max\{R_1, R_2\}$. Then $U_1 \cap U_2$ is open and $\mathbb{R}^n \setminus \overline{\mathsf{B}}^n(R, \mathbf{0}) \subseteq (U_1 \cap U_2) \cap A$. Thus $(U_1 \cap U_2) \cap A \in D(A, \infty)$ and

$$U_1 \cap A, U_2 \cap A \supseteq (U_1 \cap U_2) \cap A.$$ ■

Now we define the sort of nets we consider in this case.

1.2.74 Definition ($\mathbb{R}^n$-net, convergence in $\mathbb{R}^n$-nets) Let $A \subseteq \mathbb{R}^n$, let $x_0 \in \mathbb{R}^n$, and let $D \in \{D(A, x_0), D(A, \infty)\}$. A ***$\mathbb{R}^n$-net*** in D is a map $\phi \colon A \to \mathbb{R}^m$ for some $m \in \mathbb{Z}_{>0}$. A $\mathbb{R}^n$-net $\phi \colon A \to \mathbb{R}^m$ in the $\mathbb{R}^n$-directed set D

(i) ***converges to $s_0 \in \mathbb{R}^m$*** if, for any $\epsilon \in \mathbb{R}_{>0}$, there exists $U \cap A \in D$ such that, for any $V \cap A \in D$ for which $U \cap A \supseteq V \cap A$, $\|\phi(x) - s_0\|_{\mathbb{R}^m} < \epsilon$ for every $x \in V \cap A$;

(ii) has $s_0 \in \mathbb{R}^m$ as a ***limit*** if it converges to s_0, and we write $s_0 = \lim_D \phi$;

(iii) ***diverges*** if it does not converge,

(iv) has a limit that ***exists*** if $\lim_D \phi \in \mathbb{R}^m$, and

(v) is ***oscillatory*** if the limit of the $\mathbb{R}^n$-net does not exist, does not diverge to ∞, and does not diverge to $-\infty$. •

As with $\mathbb{R}$-nets, it is convenient to have some notation for $\mathbb{R}^n$-nets that allows us to understand more easily the sort of convergence that is taking place.

1.2.75 Notation (Limits of $\mathbb{R}^n$-nets) Let $A \subseteq \mathbb{R}^n$, let $x_0 \in \mathbb{R}^n$, let $D \in \{D(A, x_0), D(A, \infty)\}$, and let $\phi\colon A \to \mathbb{R}^m$ be a $\mathbb{R}^n$-net in D. Let us look at the two cases and give notation for each.

(i) $D = D(A, x_0)$: In this case we write $\lim_D \phi = \lim_{x \to_A x_0} \phi(x)$.

(ii) $D = D(A, \infty)$: In this case we write $\lim_D \phi = \lim_{x \to_A \infty} \phi(x)$. •

As with $\mathbb{R}$-nets, convergence in $\mathbb{R}^n$-nets can be characterised in terms of sequences in the case when x_0 is a limit of points in A.

1.2.76 Proposition (Convergence in $\mathbb{R}^n$-nets in terms of sequences) *Let* $\mathrm{A} \subseteq \mathbb{R}^n$, *let* $\mathbf{x}_0 \in \mathbb{R}^n$, *let* $\mathrm{D} \in \{\mathrm{D}(\mathrm{A}, \mathbf{x}_0), \mathrm{D}(\mathrm{A}, \infty)\}$, *and let* $\phi\colon \mathrm{A} \to \mathbb{R}^m$ *be a* $\mathbb{R}^n$*-net in* D. *Then, corresponding to the two cases in Notation 1.2.75, we have the following statements:*

(i) if $\mathbf{x}_0 \in \mathrm{cl}(\mathrm{A})$, *then the following statements are equivalent:*

(a) $\lim_{\mathbf{x} \to_A \mathbf{x}_0} \phi(\mathbf{x}) = \mathbf{s}_0$*;*

(b) $\lim_{j \to \infty} \phi(\mathbf{x}_j) = \mathbf{s}_0$ *for every sequence* $(\mathbf{x}_j)_{j \in \mathbb{Z}_{>0}}$ *in* A *converging to* $\mathbf{x}_0$*;*

(ii) if $\sup\{\|\mathbf{x}\|_{\mathbb{R}^n} \mid \mathbf{x} \in \mathrm{A}\} = \infty$, *then the following statements are equivalent:*

(a) $\lim_{\mathbf{x} \to_A \infty} \phi(\mathbf{x}) = \mathbf{s}_0$*;*

(b) $\lim_{j \to \infty} \|\phi(\mathbf{x}_j)\|_{\mathbb{R}^m} = \mathbf{s}_0$ *for every sequence* $(\mathbf{x}_j)_{j \in \mathbb{Z}_{>0}}$ *in* A *such that* $\lim_{j \to \infty} \|\mathbf{x}_j\|_{\mathbb{R}^n} = \infty$.

Proof For the first equivalence, suppose that $\lim_{x \to_A x_0} \phi(x) = s_0$ and let $(x_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in A converging to x_0. Let $\epsilon \in \mathbb{R}_{>0}$ and let $U \cap A \in D(A, x_0)$ be such that, for any $V \cap A \in D(A, x_0)$ for which $U \cap A \supseteq V \cap A$, we have $\|\phi(x) - s_0\|_{\mathbb{R}^m} < \epsilon$ for any $x \in V \cap A$. Let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $x_j \in U \cap A$ for every $j \geq N$, this being possible since $(x_j)_{j \in \mathbb{Z}_{>0}}$ converges to x_0. Now note that $\|\phi(x_j) - s_0\|_{\mathbb{R}^m} < \epsilon$ for every $j \geq N$ since $x_j \in U \cap A$ for every $j \geq N$. This gives $\lim_{j \to \infty} \phi(x_j) = s_0$, as desired.

For the converse, suppose that $\lim_{x \to_A x_0} \phi(x) \neq s_0$. Then there exists $\epsilon \in \mathbb{R}_{>0}$ such that, for any $U \cap A \in D(A, x_0)$, we have a $V \cap A \in D(A, x_0)$ with $U \cap A \supseteq V \cap A$ for which $\|\phi(x) - s_0\|_{\mathbb{R}^m} \geq \epsilon$ for some $x \in V \cap A$. Since $x_0 \in \mathrm{cl}(A)$ it follows that, for any $j \in \mathbb{Z}_{>0}$, there exists $x_j \in \mathsf{B}^n(\frac{1}{j}, x_0) \cap A$ such that $\|\phi(x_j) - s_0\|_{\mathbb{R}^m} \geq \epsilon$. Thus the sequence $(x_j)_{j \in \mathbb{Z}_{>0}}$ in A converging to x_0 has the property that $(\phi(x_j))_{j \in \mathbb{Z}_{>0}}$ does not converge to s_0.

For the second equivalence, suppose that $\lim_{x \to_A \infty} \phi(x) = s_0$ and let $(x_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in A such that $\lim_{j \to \infty} \|x_j\|_{\mathbb{R}^n} = \infty$. Let $M \in \mathbb{R}_{>0}$ and let $U \cap A \in D(A, \infty)$ be such that, for any $V \cap A \in D(A, \infty)$ for which $U \cap A \supseteq V \cap A$, we have $\|\phi(x) - s_0\|_{\mathbb{R}^m} < \epsilon$ for every $x \in V \cap A$. Let $R \in \mathbb{R}_{>0}$ be such that $\mathbb{R}^n \setminus \overline{\mathsf{B}}^n(R, \mathbf{0}) \subseteq U$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $\|x_j\|_{\mathbb{R}^n} > R$ for $j \geq N$. It then follows that $\|\phi(x_j) - s_0\|_{\mathbb{R}^m} < \epsilon$ for every $j \geq N$ since $x_j \in U \cap A$. Thus $\lim_{j \to \infty} \phi(x_j) = s_0$.

For the converse, suppose that $\lim_{x\to_A\infty}\phi(x)$. Then there exists $\epsilon\in\mathbb{R}_{>0}$ such that, for any $U\cap A\in D(A,\infty)$, we have a $V\cap A\in D(A,\infty)$ with $U\cap A\supseteq V\cap A$ for which $\|\boldsymbol{\phi}(\boldsymbol{x})-\boldsymbol{s}_0\|_{\mathbb{R}^m}\geq\epsilon$ for some $\boldsymbol{x}\in V\cap A$. By our assumption that A is unbounded, it follows that, for any $j\in\mathbb{Z}_{>0}$, there exists $\boldsymbol{x}_j\in(\mathbb{R}^n\setminus\overline{\mathsf{B}}^n(j,\boldsymbol{x}_0))\cap A$ such that $\|\boldsymbol{\phi}(\boldsymbol{x}_j)-\boldsymbol{s}_0\|_{\mathbb{R}^m}\geq\epsilon$. Thus the sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in A for which $(\|\boldsymbol{x}_j\|_{\mathbb{R}^n})_{j\in\mathbb{Z}_{>0}}$ diverges to ∞ has the property that $(\boldsymbol{\phi}(\boldsymbol{x}_j))_{j\in\mathbb{Z}_{>0}}$ does not converge to $\boldsymbol{s}_0$. ■

From the preceding result, we can easily establish the equivalence of convergence of $\mathbb{R}^n$-nets with $n=1$ with convergence of $\mathbb{R}$-nets from Section I-2.3.7.

Now let us give some examples to make the preceding construction concrete.

1.2.77 Examples (Convergence in $\mathbb{R}^n$-nets) In the examples below we will simply give "the answer," leaving to the reader the mundane details of verification.

1. Define $\phi\colon\mathbb{R}^n\to\mathbb{R}$ by
$$\phi(\boldsymbol{x})=\frac{1}{1+\|\boldsymbol{x}\|^2_{\mathbb{R}^n}}.$$
If we think of ϕ as a $\mathbb{R}^n$-net in $D=D(\mathbb{R}^n,\boldsymbol{0})$ then $\lim_D\phi=1$. If we think of ϕ as a $\mathbb{R}^n$-net in $D=D(\mathbb{R}^n,\infty)$ then $\lim_D\phi=0$.
2. Define $\phi\colon\mathbb{R}^n\to\mathbb{R}$ by $\phi(\boldsymbol{x})=\sin(\|\boldsymbol{x}\|_{\mathbb{R}^n})$. If we think of ϕ as a $\mathbb{R}^n$-net in $D=D(\mathbb{R}^n,\boldsymbol{0})$ then $\lim_D\phi=1$. If we think of ϕ as a $\mathbb{R}^n$-net in $D=D(\mathbb{R}^n,\infty)$ then $\lim_D\phi$ does not exist. •

There are also generalisations of lim sup and lim inf to $\mathbb{R}^n$-nets. We let $A\subseteq\mathbb{R}^n$, $\boldsymbol{x}_0\in\mathbb{R}^n$, $D=\in\{D(A,\boldsymbol{x}_0),D(A,\infty)\}$, and $\phi\colon A\to\mathbb{R}$. We denote by $\sup_D\phi,\inf_D\phi\colon A\to\mathbb{R}$ the $\mathbb{R}$-nets in D given by

$$\begin{aligned}\sup_D\phi(x)&=\sup\{\phi(y)\mid y\in U\cap A\text{ for all }U\cap A\in D\text{ for which }x\in U\cap A\},\\ \inf_D\phi(x)&=\inf\{\phi(y)\mid y\in U\cap A\text{ for all }U\cap A\in D\text{ for which }x\in U\cap A\}.\end{aligned}$$

Then we define

$$\limsup_D\phi=\lim_D\sup_D\phi,\quad\liminf_D\phi=\lim_D\inf_D\phi.$$

Let us now adapt our notion of Landau symbols from Section I-2.3.8 to $\mathbb{R}^n$-nets.

1.2.78 Definition (Landau symbols "O" and "o") Let $A\subseteq\mathbb{R}^n$, let $\boldsymbol{x}_0\in\mathbb{R}^n$, let $D\in\{D(A,\boldsymbol{x}_0),D(A,\infty)\}$ be a $\mathbb{R}^n$-directed set, and let $\phi\colon A\to\mathbb{R}$.

(i) Denote by $O_D(\phi)$ the functions $\boldsymbol{\psi}\colon A\to\mathbb{R}^m$ for which there exists $U\cap A\in D$ and $M\in\mathbb{R}_{>0}$ such that, for every $V\cap A\in D$ for which $U\cap A\supseteq V\cap A$, $\|\boldsymbol{\psi}(\boldsymbol{x})\|_{\mathbb{R}^m}\leq M|\phi(\boldsymbol{x})|$ for every $\boldsymbol{x}\in V\cap A$.

(ii) Denote by $o_D(\phi)$ the functions $\boldsymbol{\psi}\colon A\to\mathbb{R}$ such that, for any $\epsilon\in\mathbb{R}_{>0}$, there exists $U\cap A\in D$ such that $\|\boldsymbol{\psi}(\boldsymbol{x})\|_{\mathbb{R}^m}<\epsilon|\phi(\boldsymbol{x})|$ for $\boldsymbol{x}\in V\cap A$.

If $\psi \in O_D(\phi)$ (resp. $\psi \in o_D(\phi)$) then we say that ψ is ***big oh of ϕ*** (resp. ***little oh of ϕ***). •

It is often the case that the comparison function ϕ is positive on A. In such cases, one can give a somewhat more concrete characterisation of O_D and o_D.

1.2.79 Proposition (Alternative characterisation of Landau symbols) *Let $A \subseteq \mathbb{R}^n$, let $x_0 \in \mathbb{R}^n$, let $D \in \{D(A, x_0), D(A, \infty)\}$ be a $\mathbb{R}^n$-directed set, and let $\phi\colon A \to \mathbb{R}_{>0}$ and $\psi\colon A \to \mathbb{R}$. Then*

(i) $\psi \in O_D(\phi)$ if and only if $\limsup_D \frac{\|\psi\|_{\mathbb{R}^m}}{\phi} < \infty$ and

(ii) $\psi \in o_D(\phi)$ if and only if $\lim_D \frac{\|\psi\|_{\mathbb{R}^m}}{\phi} = 0$.

Proof We leave this as Exercise 1.2.15. ■

1.2.80 Examples (Landau symbols)

1. Generalising what we saw in Example I-2.3.34 for differentiability of $\mathbb{R}$-valued functions defined on intervals, let $U \subseteq \mathbb{R}^n$ be open, let $x_0 \in U$, and let $f\colon U \to \mathbb{R}^m$. Let $k \in \mathbb{Z}_{\geq 0}$ and for $A_j \in S^j(\mathbb{R}^n; \mathbb{R}^m)$, $j \in \{0, 1, \dots, k\}$, define $g_{f,x_0,A}\colon U \to \mathbb{R}^m$ by

$$g_{f,x_0,A}(x) = \frac{A_0}{0!} + \frac{A_1(x)}{1!} + \frac{A_2(x,x)}{2!} + \cdots + \frac{A_k(x,\dots,x)}{k!}.$$

Define a $\mathbb{R}^n$-net in $D = D(U, x_0)$ by $\phi_k(x) = \|x - x_0\|^k_{\mathbb{R}^n}$. Then one can verify (this is Taylor's Theorem) that f is k-times continuously differentiable at x_0 with $D^j f(x_0) = A_j$, $j \in \{0, 1, \dots, k\}$, if and only if $\|f - g_{f,x_0,A}\|_{\mathbb{R}^m} \in o_D(\phi_m)$.

Exercises

1.2.1 Show that

$$\Big\|\sum_{j=1}^m x_j\Big\|_{\mathbb{R}^n} \leq \sum_{j=1}^m \|x_j\|_{\mathbb{R}^n}$$

for any finite family $(x_1, \dots, x_m)$ in $\mathbb{R}^n$.

1.2.2 Let $A \subseteq \mathbb{R}^n$ be closed and let $(x_j)_{j \in \mathbb{Z}_{>0}}$ be a Cauchy sequence. Show that the sequence converges to a point in A.

1.2.3 Prove Proposition 1.2.19.

1.2.4 Show that a subset $C \subseteq \mathbb{R}^n$ is closed if and only if $C \cap K$ is closed for every compact subset K of $\mathbb{R}^n$.

1.2.5 Let $(A_i)_{i \in I}$ be a family of connected subsets of $\mathbb{R}^n$ and suppose that $\cap_{i \in I} A_i \neq \emptyset$. Show that $\cup_{i \in I} A_i$ is connected.

1.2.6 Show that the closure of a connected set is connected.

1.2.7 Show that for a subset $D \subseteq \mathbb{R}^n$ the following two statements are equivalent:

1. D is discrete, i.e., every subset of D is relatively open in D;

2. there exists $\epsilon \in \mathbb{R}_{>0}$ such that, for every $x \in D$, $\mathsf{B}^n(\epsilon, x) \cap D = \{x\}$.

1.2.8 Let $U \subseteq \mathbb{R}^n$ be open and $C \subseteq \mathbb{R}^n$ be closed.

(a) If $A \subseteq U$ show that $\operatorname{int}_U(A) = \operatorname{int}(A)$.

(b) If $A \subseteq C$ show that $\operatorname{cl}_C(A) = \operatorname{int}(A)$.

1.2.9 Show that finite subsets of $\mathbb{Q}$ are relatively compact.

1.2.10 Show that if $r, s \in \mathbb{R}_{>0}$, and $x_0 \in \mathbb{R}^m$ and $y_0 \in \mathbb{R}^n$, then $\mathsf{B}^m(r, x_0) \times \mathsf{B}^n(s, y_0)$ is an open subset of $\mathbb{R}^n \times \mathbb{R}^m$.

1.2.11 Let $A \subseteq \mathbb{R}^m$ and $B \subseteq \mathbb{R}^n$. Show that a sequence $((x_j, y_j))_{j \in \mathbb{Z}_{>0}}$ converges to (x_0, y_0) if and only if $(x_j)_{j \in \mathbb{Z}_{>0}}$ and $(y_j)_{j \in \mathbb{Z}_{>0}}$ converge to x_0 and y_0, respectively.

1.2.12 Let $(Z_j)_{j \in \mathbb{Z}_{>0}}$ be a family of subsets of $\mathbb{R}^n$ that each have measure zero. Show that $\cup_{j \in \mathbb{Z}_{>0}} Z_j$ also has measure zero.

1.2.13 If $\mathsf{V} \subseteq \mathbb{R}^n$ is a subspace of dimension at most $n-1$ show that V has measure zero.

1.2.14 Let $D \in \{D(A, x_0), D(A, \infty)\}$ be a $\mathbb{R}^n$-directed set and let $\phi \colon A \to \mathbb{R}^m$ be a $\mathbb{R}^n$-net in D. For $s_0 \in \mathbb{R}^m$ define the corresponding $\mathbb{R}^n$ net $\phi_{x_0, s_0} \colon A \to \mathbb{R}_{\geq 0}$ by $\phi_{x_0, s_0}(x) = \|\phi(x) - s_0\|_{\mathbb{R}^m}$. Show that $\lim_D \phi = s_0$ if and only if $\lim_D \phi_{x_0, s_0} = 0$.

1.2.15 Prove Proposition 1.2.79.

1.2.16

Section 1.3

Continuous functions of multiple variables

With the structure of $\mathbb{R}^n$ as given in Section 1.2 it is fairly easy to generalise the notion of continuity from the single-variable case to the multivariable case. Thus much of what we say in this section bears a strong resemblance to the material in Section I-3.1. We do, however, add more depth and detail in this section than we did in Section I-3.1. For example, we discuss the structure of linear maps, affine maps, isometries of $\mathbb{R}^n$, and homeomorphisms. Reading this section will be excellent preparation for understanding the general notion of a continuous map and its properties as presented in Section III-1.3.

Since this section does repeat some of the material from Section I-3.1, we omit reproducing the illustrative examples that we have already given, and only give examples that reveal something interesting about the multivariable case.

Do I need to read this section? If one is reading this chapter then one should read this section. Certain of the sections can be skipped, and these are clearly labelled. •

1.3.1 Definition and properties of continuous multivariable maps

First let us establish our notation for multivariable functions. If $A \subseteq \mathbb{R}^n$ we use a bold font, $\boldsymbol{f}\colon A \to \mathbb{R}^m$ to represent a multivariable function on A, reflecting the fact that we use a similar bold font to denote points in $\mathbb{R}^n$ for $n > 1$. In keeping with this convention, we will denote by $f\colon A \to \mathbb{R}$ a typical function taking values in $\mathbb{R}$, even though the domain is multi-dimensional. Note that, since $\boldsymbol{f}\colon A \to \mathbb{R}^m$ takes values in $\mathbb{R}^m$ we can write

$$\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \dots, f_m(\boldsymbol{x})),$$

where the functions $f_j\colon A \to \mathbb{R}$, $j \in \{1, \dots, m\}$, are the ***components*** of $\boldsymbol{f}$.

If a function $\boldsymbol{f}\colon A \to \mathbb{R}^m$ takes values in $B \subseteq \mathbb{R}^m$ we may write $\boldsymbol{f}\colon A \to B$.

The definition of continuity for $\mathbb{R}$-values functions on $\mathbb{R}$ is made using the absolute value function $|\cdot|$ on $\mathbb{R}$ in an essential way. Since the Euclidean norm $\|\cdot\|_{\mathbb{R}^n}$ provides a generalisation of the absolute value function, we shall use this to extend to multiple dimensions our definitions of continuity.

1.3.1 Definition (Continuous map) Let $n, m \in \mathbb{Z}_{>0}$ and let $A \subseteq \mathbb{R}^n$ be a subset. A map $\boldsymbol{f}\colon A \to \mathbb{R}^m$ is:

(i) ***continuous at*** $\mathbf{x_0} \in \mathbf{A}$ if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that $\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_0)\|_{\mathbb{R}^n} < \epsilon$ whenever $\boldsymbol{x} \in A$ satisfies $\|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^m} < \delta$;

(ii) ***continuous*** if it is continuous at each $\boldsymbol{x}_0 \in A$;

(iii) ***discontinuous at*** $\mathbf{x_0} \in \mathbf{A}$ if it is not continuous at x_0;

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

Note that if f takes values in $B \subseteq \mathbb{R}^m$ we shall say that $f\colon A \to B$ is continuous if it is continuous as a map into $\mathbb{R}^m$, i.e., if the map $i_B \circ f$ is continuous, where i_B is the inclusion of B into $\mathbb{R}^m$. •

Note that we define continuity for multivariable maps defined on *arbitrary* subsets of $\mathbb{R}^n$, whereas for the single-variable case we only considered functions defined on intervals. We do this principally because there is no really useful generalisation to higher-dimensions of the notion of an interval. We will mostly only use fairly well-behaved subsets of $\mathbb{R}^n$, e.g., open sets, or closures of open sets, although our definition allows rather degenerate domains for maps.

The following equivalent characterisations of continuity, except for the last, are just as they are in the case when $m = n = 1$, and, indeed, the proof also generalises the one-dimensional proof only by replacing open intervals by open balls. Here, for simplicity, we only consider maps whose domain is an open set (see Example III-1.1.5–3 for the definition of an open set in this case).

1.3.2 Theorem (Alternative characterisations of continuity) *For a map* $\mathbf{f}\colon \mathrm{A} \to \mathbb{R}$ *defined on a subset* $\mathrm{A} \subseteq \mathbb{R}^\mathrm{n}$ *and for* $\mathbf{x}_0 \in \mathrm{A}$*, the following statements are equivalent:*

(i) $\mathbf{f}$ *is continuous at* $\mathbf{x}_0$*;*

(ii) for every neighbourhood V *of* $\mathbf{f}(\mathbf{x}_0)$ *there exists a neighbourhood* U *of* $\mathbf{x}_0$ *such that* $\mathbf{f}(\mathrm{U} \cap \mathrm{A}) \subseteq \mathrm{V}$*;*

(iii) $\lim_{\mathbf{x} \to_{\mathrm{A}} \mathbf{x}_0} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{x}_0)$*;*

(iv) the components of $\mathbf{f}$ *are continuous at* $\mathbf{x}_0$*.*

Proof We shall show the equivalence of the first three statements, leaving the last as Exercise 1.3.2.

(i) $\Longrightarrow$ (ii) Let $V \subseteq \mathbb{R}^m$ be a neighbourhood of $f(x_0)$ and let $\epsilon \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^m(\epsilon, f(x_0)) \subseteq V$. Then, by continuity of f, let $\delta \in \mathbb{R}_{>0}$ be such that $\|f(x) - f(x_0)\|_{\mathbb{R}^m} < \epsilon$ if $x \in A$ satisfies $\|x - x_0\|_{\mathbb{R}^n} < \delta$. That is, if $U = \mathsf{B}^n(\delta, x_0)$ then $f(U \cap A) \subseteq \mathsf{B}^m(\epsilon, f(x_0)) \subseteq V$.

(ii) $\Longrightarrow$ (iii) Let $(x_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in A converging to x_0. For $\epsilon \in \mathbb{R}_{>0}$ let U be a neighbourhood of x_0 such that $f(U \cap A) \subseteq \mathsf{B}^m(\epsilon, f(x_0))$. Now let $\delta \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^n(\delta, x_0) \subseteq U$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $\|x_j - x_0\|_{\mathbb{R}^n} < \delta$ for $j \geq N$. Then, for $j \geq N$, $f(x_j) \subseteq \mathsf{B}^m(\epsilon, f(x_0))$, i.e., $\|f(x_j) - f(x_0)\|_{\mathbb{R}^m} < \epsilon$ for $j \geq N$. Thus $(f(x_0))_{j \in \mathbb{Z}_{>0}}$ converges to $f(x_0)$.

(iii) $\Longrightarrow$ (i) Suppose that f is not continuous at x_0. Then there exists $\epsilon \in \mathbb{R}_{>0}$ such that, for any $\delta \in \mathbb{R}_{>0}$, $f(\mathsf{B}^n(\delta, x_0) \cap A) \not\subset \mathsf{B}^m(\epsilon, f(x_0))$. For each $j \in \mathbb{Z}_{>0}$, therefore, let $x_j \in A$ satisfy $\|x_j - x_0\|_{\mathbb{R}^n} < \frac{1}{j}$ and $f(x_j) \notin \mathsf{B}^m(\epsilon, f(x_0))$. Then the sequence $(x_j)_{j \in \mathbb{Z}_{>0}}$ converges to x_0 but the sequence $(f(x_j))_{j \in \mathbb{Z}_{>0}}$ does not converge to $f(x_0)$. ■

Note that the last part of the preceding theorem says that "f is continuous if and only if its components are continuous." This is not to be confused with the incorrect statement that "f is continuous if and only if it is a continuous function of each component." The following example illustrates the distinction.

1.3.3 Example (A discontinuous function that is continuous in each of its variables) Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x_1, x_2) = \begin{cases} \frac{x_1x_2}{x_1^2+x_2^2}, & (x_1, x_2) \neq (0,0), \\ 0, & (0,0). \end{cases}$$

We first claim that this function is discontinuous at $(0,0)$. Indeed, consider points in $\mathbb{R}^2$ of the form (a,a) for $a \in \mathbb{R}^*$. At such points we have $f(a,a) = \frac{1}{2}$. Since $f(0,0) = 0$ and since every neighbourhood of $(0,0)$ contains a point of the form (a,a) for some $a \in \mathbb{R}^*$, it follows that f cannot be continuous at $(0,0)$.

We also claim that for fixed $x_{10} \in \mathbb{R}$ (resp. $x_{20} \in \mathbb{R}$) the function $x_2 \mapsto f(x_{10}, x_2)$ (resp. $x_1 \mapsto f(x_1, x_{20})$) is continuous. First fix $x_{10} \in \mathbb{R}^*$. Then the function $x_2 \mapsto \frac{x_{10}x_2}{x_{10}^2+x_2^2}$ is clearly continuous (since the denominator is nonzero and since sums, products, and quotients by nonzero functions preserve continuity). If $x_{10} = 0$ then we have $f(x_{10}, x_2) = 0$ for all $x_2 \in \mathbb{R}$, and this is obviously a continuous function. This shows that $x_2 \mapsto f(x_{10}, x_2)$ is continuous for every $x_{10} \in \mathbb{R}$. An entirely similar argument shows that $x_1 \mapsto f(x_1, x_{20})$ is continuous for all $x_{20} \in \mathbb{R}$. •

The previous theorem also has the following useful restatement which employs the relative topology discussed in Section 1.2.8.

1.3.4 Corollary (Characterisation of continuous maps) *For* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *and for* $\mathbf{f}\colon \mathrm{A} \to \mathbb{R}^{\mathrm{m}}$ *the following statements are equivalent:*

(i) $\mathbf{f}$ *is continuous;*

(ii) $\mathbf{f}^{-1}(\mathrm{V})$ *is relatively open in* A *for every open subset* V *of* $\mathbb{R}^{\mathrm{m}}$.

Proof First suppose that $\boldsymbol{f}$ is continuous and let $V \subseteq \mathbb{R}^m$ be open. Let $\boldsymbol{x}_0 \in \boldsymbol{f}^{-1}(V)$ so that $\boldsymbol{f}(\boldsymbol{x}_0) \in V$. Since V is open and so a neighbourhood of $\boldsymbol{f}(\boldsymbol{x}_0)$, by Theorem 1.3.2 there exists a neighbourhood U of $\boldsymbol{x}_0$ such that $f(U \cap A) \subseteq V$. Thus $U \cap A$ is a relative neighbourhood of $\boldsymbol{x}_0$ in $\boldsymbol{f}^{-1}(V)$ and so $\boldsymbol{f}^{-1}(V)$ is open.

Now suppose that $\boldsymbol{f}^{-1}(V)$ is relatively open in A for every open subset V of $\mathbb{R}^m$. Let $\boldsymbol{x}_0 \in A$ and let V be a neighbourhood of $\boldsymbol{f}(\boldsymbol{x}_0)$. Then $\boldsymbol{f}^{-1}(V)$ is a relative neighbourhood of $\boldsymbol{x}_0$ in A. By Proposition 1.2.50 there exists an open set U in $\mathbb{R}^n$ such that $\boldsymbol{f}^{-1}(V) = U \cap A$. Therefore, since $\boldsymbol{f}(\boldsymbol{f}^{-1}(V)) \subseteq V$ by Proposition I-1.3.5, it follows that $\boldsymbol{f}$ is continuous at $\boldsymbol{x}_0$ using Theorem 1.3.2. ■

The notion of uniform continuity can be extended to multivariable functions.

1.3.5 Definition (Uniform continuity) Let $A \subseteq \mathbb{R}^n$. A map $\boldsymbol{f}\colon A \to \mathbb{R}^m$ is ***uniformly continuous*** if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that $\|\boldsymbol{f}(\boldsymbol{x}_1) - \boldsymbol{f}(\boldsymbol{x}_2)\|_{\mathbb{R}^n} < \epsilon$ whenever $\boldsymbol{x}_1, \boldsymbol{x}_2 \in A$ satisfy $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathbb{R}^n} < \delta$. •

Obviously all uniformly continuous functions are continuous. We refer the reader to Example I-3.1.7 for an example of a continuous but not uniformly continuous function.

We close this section by initiating a discussion of the relationship between continuity, interior, closure, and boundary.

1.3.6 Proposition (Continuity and interior, closure, and boundary) *If* $A \subseteq \mathbb{R}^n$, *if* $S \subseteq A$, *if* $B \subseteq \mathbb{R}^m$, *and if* $\mathbf{f}\colon A \to \mathbb{R}^m$ *is continuous then the following statements hold:*

(i) $\mathrm{int}_B(\mathbf{f}(S))) \subseteq \mathbf{f}(\mathrm{int}_A(S))$;

(ii) $\mathbf{f}(\mathrm{cl}_S(A)) \subseteq \mathrm{cl}_B(\mathbf{f}(S))$;

(iii) $\mathbf{f}(\mathrm{bd}_S(A)) \subseteq \mathrm{bd}_B(\mathbf{f}(S))$.

Proof Let $\boldsymbol{y} \in \mathrm{int}_B(\boldsymbol{f}(S))$ then there exists a relative neighbourhood U of $\boldsymbol{y}$ in $\boldsymbol{f}(S)$ in B such that $U \subseteq \boldsymbol{f}(S)$. Then $\boldsymbol{f}^{-1}(U)$ is relatively open in A. That is, if $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ for $\boldsymbol{x} \in S$ then $\boldsymbol{x} \in \mathrm{int}_A(S)$. Thus $\boldsymbol{y} \in \boldsymbol{f}(\mathrm{int}_A(S))$.

Let $\boldsymbol{y} \in \boldsymbol{f}(\mathrm{cl}_A(S))$ with $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathrm{cl}_A(S)$. Then there exists a sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in S converging to $\boldsymbol{x}$. By Theorem 1.3.2 it follows that $(\boldsymbol{f}(\boldsymbol{x}_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\boldsymbol{y}$. Since $\boldsymbol{f}(\boldsymbol{x}_j) \in \boldsymbol{f}(S)$ it follows that $\boldsymbol{y} \in \mathrm{cl}_B(\boldsymbol{f}(S))$.

Let $\boldsymbol{y} \in \boldsymbol{f}(\mathrm{bd}_A(S))$ with $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ for $\boldsymbol{x} \in \mathrm{bd}_A(S)$. Then there exist sequences $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in S and $(\boldsymbol{x}'_j)_{j\in\mathbb{Z}_{>0}}$ in $A \setminus S$, both converging to $\boldsymbol{x}$. By continuity of $\boldsymbol{f}$ the sequences $(\boldsymbol{f}(\boldsymbol{x}_j))_{j\in\mathbb{Z}_{>0}}$ in $\boldsymbol{f}(S)$ and $(\boldsymbol{f}(\boldsymbol{x}'_j))_{j\in\mathbb{Z}_{>0}}$ in $\boldsymbol{f}(A \setminus S) = \boldsymbol{f}(A) \setminus \boldsymbol{f}(S)$ both converge to $\boldsymbol{y}$. Thus $\boldsymbol{y} \in \mathrm{bd}_{\boldsymbol{f}(S)}(\boldsymbol{f}(B))$. ■

In general, the converse inclusions of the preceding result are not true.

1.3.7 Examples (Continuity and interior, closure, and boundary)

1. Consider $A = S = [0, \pi] \subseteq \mathbb{R}$, $B = [0, 1] \subseteq \mathbb{R}$, and take $f\colon A \to B$ given by $f(x) = \sin(x)$. Note that $f(S) = [0, 1]$. Then $f(\frac{\pi}{2}) = 1$ and so $1 \in f(\mathrm{int}_A(S))$. However, $1 \notin \mathrm{int}_B(f(S))$.
2. Take $A = S = \mathbb{R} \subseteq \mathbb{R}$, $B = [-\frac{\pi}{2}, \frac{\pi}{2}]$, and let $f\colon A \to B$ be given by $f(x) = \tan^{-1}(x)$. Note that $f(S) = (-\frac{\pi}{2}, \frac{\pi}{2})$. Thus $\frac{\pi}{2} \in \mathrm{cl}_B(f(S))$ but $\frac{\pi}{2} \notin f(\mathrm{cl}_A(S))$ since S is closed.
3. The same example as the preceding works here since $\frac{\pi}{2} \in \mathrm{bd}_B(f(S))$ but $\frac{\pi}{2} \notin f(\mathrm{bd}_A(S))$ since $\mathrm{bd}_A(S) = \varnothing$. •

1.3.2 Discontinuous maps

This section is rather specialised and technical and so can be omitted until needed. However, the material is needed at certain points in the text.

Next we consider the discontinuities of multivariable functions. The discussion here is not much different from that in a single variable, so we keep things brief.

1.3.8 Definition (Types of discontinuity) Let $A \subseteq \mathbb{R}^n$ and suppose that $\boldsymbol{f}\colon A \to \mathbb{R}^m$ is discontinuous at $\boldsymbol{x}_0 \in A$. The point $\boldsymbol{x}_0$ is:

(i) a ***removable discontinuity*** if $\lim_{\boldsymbol{x}\to_A \boldsymbol{x}_0} \boldsymbol{f}(\boldsymbol{x})$ exists;

(ii) an ***essential discontinuity*** if the limit $\lim_{\boldsymbol{x}\to_A \boldsymbol{x}_0} \boldsymbol{f}(\boldsymbol{x})$ exists.

The set of all discontinuities of $\boldsymbol{f}$ is denoted by $D_{\boldsymbol{f}}$. •

Note that we are not quite able to give as refined a characterisation of a point of discontinuity as we did in the single-variable case. This is because the discontinuities of multiple-variable functions can be rather more general that those for single-variable functions. Let us explore this in the context of an example.

1.3.9 Example (Strangeness of discontinuities for multivariable functions) We again consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ considered in Example 1.3.3:

$$f(x_1, x_2) = \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2}, & (x_1, x_2) \neq (0,0), \\ 0, & (0,0). \end{cases}$$

In Example 1.3.3 we showed that this function was continuous when thought of separately as a function of x_1 and of x_2, but was actually discontinuous at $(0,0)$. Here we shall further explore the nature of the discontinuity at $(0,0)$. First let us consider how the function behaves as we approach the origin along lines. Thus consider the line

$$s \mapsto (0,0) + s(u_1, u_2), \qquad s \in \mathbb{R}$$

through $(0,0)$ in the direction (u_1, u_2). We easily compute

$$f((0,0) + s(u_1, u_2)) = \frac{u_1 u_2}{u_1^2 + u_2^2}.$$

If $u_1 = 0$ or $u_2 = 0$ then we have

$$\lim_{s \to 0} f((0,0) + s(u_1, u_2)) = 0.$$

For $u_1 \neq 0$ let us take $u_2 = a u_1$, i.e., the line has slope $a \in \mathbb{R}$. In this case we have

$$\lim_{s \to 0} f((0,0) + s(u_1, u_2)) = \frac{a}{1 + a^2}.$$

Similarly, if $u_2 \neq 0$ and $u_1 = b u_2$ then we have

$$\lim_{s \to 0} f((0,0) + s(u_1, u_2)) = \frac{b}{1 + b^2}.$$

Thus all of these limits are finite, but the value of the limit depends on the direction in which one approaches $(0,0)$. •

As in the single-variable case, we can use the oscillation to measure the discontinuity of a function.

1.3.10 Definition (Oscillation) Let $A \subseteq \mathbb{R}^n$ and let $\boldsymbol{f}\colon A \to \mathbb{R}^m$ be a map. The ***oscillation*** of $\boldsymbol{f}$ is the map $\omega_{\boldsymbol{f}}\colon A \to \mathbb{R}$ defined by

$$\omega_{\boldsymbol{f}}(\boldsymbol{x}) = \inf\{\sup\{\|\boldsymbol{f}(\boldsymbol{x}_1) - \boldsymbol{f}(\boldsymbol{x}_2)\|_{\mathbb{R}^m} \mid \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathsf{B}^n(\delta, \boldsymbol{x}) \cap A\} \mid \delta \in \mathbb{R}_{>0}\}.$$ •

Note that the definition makes sense since the function

$$\delta \mapsto \sup\{\|\boldsymbol{f}(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\|_{\mathbb{R}^m} \mid \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathsf{B}^n(\delta, \boldsymbol{x}) \cap A\}$$

is monotonically increasing. In particular, if $\boldsymbol{f}$ is bounded (see Definition 1.3.30 below) then $\omega_{\boldsymbol{f}}$ is also bounded. The following result indicates in what way $\omega_{\boldsymbol{f}}$ measures the continuity of $\boldsymbol{f}$.

1.3.11 Proposition (Oscillation measures discontinuity) *For a subset* $A \subseteq \mathbb{R}$ *and a map* $\mathbf{f}\colon A \to \mathbb{R}$, $\mathbf{f}$ *is continuous at* $\mathbf{x} \in A$ *if and only if* $\omega_{\mathbf{f}}(\mathbf{x}) = 0$.

Proof Suppose that f is continuous at x and let $\epsilon \in \mathbb{R}_{>0}$. Choose $\delta \in \mathbb{R}_{>0}$ such that if $y \in \mathsf{B}^n(\delta, x) \cap A$ then $\|f(y) - f(x)\|_{\mathbb{R}^m} < \frac{\epsilon}{2}$. Then, for $x_1, x_2 \in \mathsf{B}^n(\delta, x)$ we have

$$\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} \leq \|f(x_1) - f(x)\|_{\mathbb{R}^m} + \|f(x) - f(x_2)\|_{\mathbb{R}^m} < \epsilon.$$

Therefore,

$$\sup\{\|f(x_1) - f(x_2)\|_{\mathbb{R}^n} \mid x_1, x_2 \in \mathsf{B}^n(\delta, x) \cap A\} < \epsilon.$$

Since ϵ is arbitrary this gives

$$\inf\{\sup\{\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} \mid x_1, x_2 \in \mathsf{B}^n(\delta, x) \cap A\} \mid \delta \in \mathbb{R}_{>0}\} = 0,$$

meaning that $\omega_f(x) = 0$.

Now suppose that $\omega_f(x) = 0$. For $\epsilon \in \mathbb{R}_{>0}$ let $\delta \in \mathbb{R}_{>0}$ be chosen such that

$$\sup\{\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} \mid x_1, x_2 \in \mathsf{B}^n(\delta, x) \cap A\} < \epsilon.$$

In particular, $\|f(y) - f(x)\|_{\mathbb{R}^m} < \epsilon$ for all $y \in \mathsf{B}^n(\delta, x) \cap A$, giving continuity of f at x. ■

Let us consider an example where we can compute the oscillation.

1.3.12 Example (Oscillation for a discontinuous function) We again consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$

$$f(x_1, x_2) = \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2}, & (x_1, x_2) \neq (0, 0), \\ 0, & (0, 0) \end{cases}$$

that is discontinuous at $(0, 0)$. Let us determine $\omega_f(0, 0)$. As we saw in Example 1.3.9, the function is constant on lines through $(0, 0)$. Therefore, all values of the function in any neighbourhood of $(0, 0)$ are attained by considering the values of the function along lines through $(0, 0)$. Moreover, in Example 1.3.9 we did this computation and we recall that the results were as follows.

1. On the line $s \mapsto (s, 0)$, $f(s, 0) = 0$.
2. On the line $s \mapsto (0, s)$, $f(0, s) = 0$.
3. On the line $s \mapsto (s, as)$, $f(s, as) = \frac{a}{1+a^2}$.
4. On the line $s \mapsto (bs, s)$, $f(bs, s) = \frac{b}{1+a^2}$.

The bottom line is that the values of f in any neighbourhood of $(0, 0)$ are in 1–1 correspondence with the elements of the set $\{\frac{a}{1+a^2} \mid a \in \mathbb{R}\}$. Thus one should look at the graph of the function $g\colon a \mapsto \frac{a}{1+a^2}$ to determine its maxima and minima. Since g is differentiable and $\lim_{a \to \pm\infty} g(a) = 0$, by Theorem I-3.2.16 the maxima and minima occur where g' vanishes. We compute $g'(a) = \frac{1-a^2}{(a+1^2)^2}$ which means that maxima and minima must occur at $a \in \{-1, 1\}$. Also by Theorem I-3.2.16, minima occur when $g''(a) > 0$ and maxima occur when $g''(a) < 0$. We compute $g''(1) = -\frac{1}{2}$ and $g''(-1) = \frac{1}{2}$. That $a = 1$ is a maximum for g and $a = -1$ is a minimum. We compute $g(1) = \frac{1}{2}$ and $g(-1) = -\frac{1}{2}$. This then gives $\omega_f(0, 0) = 1$.

Normally it will be quite difficult to explicitly compute the oscillation of a function. •

Let us now describe the possible set of discontinuities of an arbitrary multivariable function. The key to this, just as in the single-variable case, is the following result.

1.3.13 Proposition (Closed preimages of the oscillation of a function) *Let* $A \subseteq \mathbb{R}^n$ *and let* $\mathbf{f}\colon I \to \mathbb{R}$ *be a function. Then, for every* $\alpha \geq 0$*, the set*

$$A_\alpha = \{\mathbf{x} \in A \mid \ \omega_{\mathbf{f}}(\mathbf{x}) \geq \alpha\}$$

is relatively closed in A*.*

Proof The result where $\alpha = 0$ is clear, so we assume that $\alpha \in \mathbb{R}_{>0}$. For $\delta \in \mathbb{R}_{>0}$ define

$$\omega_f(x, \delta) = \sup\{\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} \mid x_1, x_2 \in \mathsf{B}^n(\delta, x) \cap A\}$$

so that $\omega_f(x) = \lim_{\delta\to 0} \omega_f(x, \delta)$. Let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in A_α converging to $x \in \mathbb{R}^n$ and let $(\epsilon_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $(0, \alpha)$ converging to zero. Let $j \in \mathbb{Z}_{>0}$. We claim that there exists points $y_j, z_j \in \mathsf{B}^n(\epsilon_j, x_j) \cap A$ such that $\|f(y_j) - f(z_j)\|_{\mathbb{R}^m} \geq \alpha - \epsilon_j$. Suppose otherwise so that for every $y, z \in \mathsf{B}^n(\epsilon_j, x_j) \cap A$ we have $\|f(y) - f(z)\|_{\mathbb{R}^m} < \alpha - \epsilon_j$. It then follows that $\lim_{\delta\to 0} \omega_f(x_j, \delta) \leq \alpha - \epsilon_j < \alpha$, contradicting the fact that $x_j \in A_\alpha$. We claim that $(y_j)_{j\in\mathbb{Z}_{>0}}$ and $(z_j)_{j\in\mathbb{Z}_{>0}}$ converge to x. Indeed, let $\epsilon \in \mathbb{R}_{>0}$ and choose $N_1 \in \mathbb{Z}_{>0}$ sufficiently large that $\epsilon_j < \frac{\epsilon}{2}$ for $j \geq N_1$ and choose $N_2 \in \mathbb{Z}_{>0}$ such that $\|x_j - x\|_{\mathbb{R}^n} < \frac{\epsilon}{2}$ for $j \geq N_2$. Then, for $j \geq \max\{N_1, N_2\}$ we have

$$\|y_j - x\|_{\mathbb{R}^n} \leq \|y_j - x_j\|_{\mathbb{R}^n} + \|x_j - x\|_{\mathbb{R}^n} < \epsilon.$$

Thus $(y_j)_{j\in\mathbb{Z}_{>0}}$ converges to x, and the same argument, and therefore the same conclusion, also applies to $(z_j)_{j\in\mathbb{Z}_{>0}}$.

Thus we have sequences of points $(y_j)_{j\in\mathbb{Z}_{>0}}$ and $(z_j)_{j\in\mathbb{Z}_{>0}}$ in A converging to x and a sequence $(\epsilon_j)_{j\in\mathbb{Z}_{>0}}$ in $(0, \alpha)$ converging to zero for which $\|f(y_j) - f(z_j)\|_{\mathbb{R}^m} \geq \alpha - \epsilon_j$. We claim that this implies that $\omega_f(x) \geq \alpha$. Indeed, suppose that $\omega_f(x) < \alpha$. There exists $N \in \mathbb{Z}_{>0}$ such that $\alpha - \epsilon_j > \alpha - \omega_f(x)$ for every $j \geq N$. Therefore,

$$\|f(y_j) - f(z_j)\|_{\mathbb{R}^m} \geq \alpha - \epsilon_j > \alpha - \omega_f(x)$$

for every $j \geq N$. This contradicts the definition of $\omega_f(x)$ since the sequences $(y_j)_{j\in\mathbb{Z}_{>0}}$ and $(z_j)_{j\in\mathbb{Z}_{>0}}$ converge to x.

Now we claim that the sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ converges to x. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N_1 \in \mathbb{Z}_{>0}$ be large enough that $\|x - y_j\|_{\mathbb{R}^n} < \frac{\epsilon}{2}$ for $j \geq N_1$ and let $N_2 \in \mathbb{Z}_{>0}$ be large enough that $\epsilon_j < \frac{\epsilon}{2}$ for $j \geq N_2$. Then, for $j \geq \max\{N_1, N_2\}$ we have

$$\|x - x_j\|_{\mathbb{R}^n} \leq \|x - y_j\|_{\mathbb{R}^n} + \|y_j - x_j\|_{\mathbb{R}^n} < \epsilon,$$

as desired.

This shows that every sequence in A_α converges to a point in A_α. It follows from Exercise I-2.5.2 that A_α is closed. ■

For readers who like the fancy language, we comment that the preceding result means exactly that ω_f is upper semicontinuous, cf. Proposition 1.10.13.

The following corollary is somewhat remarkable, in that it shows that the set of discontinuities of a function cannot be arbitrary.

1.3.14 Corollary (Discontinuities are the countable union of closed sets) *Let* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *and let* $\mathbf{f}\colon \mathrm{A} \to \mathbb{R}^{\mathrm{m}}$ *be a function. Then the set*

$$\mathrm{D}_{\mathbf{f}} = \{\mathbf{x} \in \mathrm{A} \mid \mathbf{f} \textit{ is not continuous at } \mathbf{x}\}$$

is the countable union of closed sets.

Proof This follows immediately from Proposition 1.3.13 after we note that

$$D_{f} = \cup_{k \in \mathbb{Z}_{>0}} \{x \in A \mid \omega_{f}(x) \geq \tfrac{1}{k}\}. \qquad \blacksquare$$

1.3.3 Linear and affine maps

In this section we study a particularly simple, but as it turns out, very interesting class of continuous maps. While we studied linear maps in detail in Chapter I-5, let us redefine them here for fun, along with another, closely related type of map. The reader will recall that if $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ is an $m \times n$-matrix with real entries (see Definition I-5.1.1) then the product of A with $x \in \mathbb{R}^n$ is the element $Ax \in \mathbb{R}^m$ defined by

$$(Ax)_a = \sum_{j=1}^{n} A(a, j)x_j.$$

With this recollection we then make the following definition.

1.3.15 Definition (Linear map, affine map) A map $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is

(i) ***linear*** if there exists $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ such that $f(x) = Ax$ for every $x \in \mathbb{R}^n$ and is

(ii) ***affine*** if there exists $A \in \mathrm{Mat}_{m\times n}(\mathbb{R})$ and $b \in \mathbb{R}^m$ such that $f(x) = Ax + b$ for every $x \in \mathbb{R}^n$. •

Recall from Theorem I-5.1.13 that in the above definition we are establishing the natural identification of $\mathrm{Mat}_{m\times n}(\mathbb{R})$ with $\mathrm{Hom}_{\mathbb{R}}(\mathbb{R}^n; \mathbb{R}^m)$. Moreover, according to Proposition I-5.4.25 this identification is of a matrix with the matrix representative of the linear map with respect to the standard basis. In this chapter we shall unblinkingly use this identification, and use the words "matrix" and "linear map" interchangeably, keeping in mind the natural identifications we are making.

Let us give some of the elementary properties of linear and affine maps. Since linear maps are special cases of affine maps, we sometimes need only consider them.

1.3.16 Proposition (Affine maps are uniformly continuous) *For* $\mathbf{A} \in \mathrm{Mat}_{\mathrm{m}\times\mathrm{n}}(\mathbb{R})$ *and* $\mathbf{b} \in \mathbb{R}^{\mathrm{m}}$*, the affine map* $\mathbf{f}\colon \mathbf{x} \mapsto \mathbf{A}\mathbf{x} + \mathbf{b}$ *is uniformly continuous.*

Proof Note that the ath component of Ax is exactly $\langle r(A, a), x\rangle_{\mathbb{R}^n}$, where we recall from Definition I-5.1.4 that $r(A, a)$ denotes the ath row of A. Let

$$M = \max\{\|r(A, a)\| \mid a \in \{1, \dots, m\}\}.$$

For $\epsilon \in \mathbb{R}_{>0}$ let $\delta = \frac{\epsilon}{\sqrt{m}M}$ and compute

$$\begin{aligned}\|f(x) - f(y)\|_{\mathbb{R}^m} &= \Big(\sum_{a=1}^{m} \langle r(A,a), x - y\rangle_{\mathbb{R}^n}\Big)^{1/2} \\ &\leq \Big(\sum_{a=1}^{m} \|r(A,a)\|_{\mathbb{R}^n}^2 \|x - y\|_{\mathbb{R}^n}^2\Big)^{1/2} \\ &\leq \sqrt{m}M\|x - y\|_{\mathbb{R}^n}.\end{aligned}$$

Thus, if $\|x - y\|_{\mathbb{R}^n} < \delta$ then $\|f(x) - f(y)\|_{\mathbb{R}^m} < \epsilon$, giving uniform continuity as desired. ■

1.3.4 Isometries

There is a special class of maps on $\mathbb{R}^n$ which (as we shall see) are affine. Let us first define the desired property of such maps.

1.3.17 Definition (Isometry of $\mathbb{R}^n$) A map $f\colon \mathbb{R}^n \to \mathbb{R}^n$ is an ***isometry*** if

$$\|f(x_1) - f(x_2)\|_{\mathbb{R}^n} = \|x_1 - x_2\|_{\mathbb{R}^n}$$

for every $x_1, x_2 \in \mathbb{R}^n$. •

The idea of an isometry, then, is that it preserves the distance between points. It is not immediately obvious, but the set of isometries has a very simple structure. To get at this, we begin by considering linear isometries.

1.3.18 Theorem (Characterisation of linear isometries of $\mathbb{R}^n$) *For a matrix* $\mathbf{R} \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ *the following statements are equivalent:*

(i) $\mathbf{R}$ *is a linear isometry;*

(ii) $\|\mathbf{R}\mathbf{x}\|_{\mathbb{R}^n} = \|\mathbf{x}\|_{\mathbb{R}^n}$ *for all* $\mathbf{x} \in \mathbb{R}^n$*;*

(iii) $\langle \mathbf{R}\mathbf{x}, \mathbf{R}\mathbf{y}\rangle_{\mathbb{R}^n} = \langle \mathbf{x}, \mathbf{y}\rangle_{\mathbb{R}^n}$ *for all* $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$*;*

(iv) $\mathbf{R}\mathbf{R}^T = \mathbf{R}^T\mathbf{R} = \mathbf{I}_n$*;*

(v) $\mathbf{R}$ *is invertible and* $\mathbf{R}^{-1} = \mathbf{R}^T$*.*

Proof (i) $\Longrightarrow$ (ii) If R is a linear isometry then

$$\|Rx - R0\|_{\mathbb{R}^n} = \|x - 0\|_{\mathbb{R}^n}$$

or $\|Rx\|_{\mathbb{R}^n} = \|x\|_{\mathbb{R}^n}$, as desired.

(ii) $\Longrightarrow$ (iii) We are assuming that $\|Rx\|_{\mathbb{R}^n} = \|x\|_{\mathbb{R}^n}$ which implies that

$$\|Rx\|_{\mathbb{R}^n}^2 = \|x\|_{\mathbb{R}^n}^2 \quad\Longrightarrow\quad \langle Rx, Rx\rangle_{\mathbb{R}^n} = \langle x, x\rangle_{\mathbb{R}^n},$$

this holding for all $x \in \mathbb{R}^n$. Thus, for every $x, y \in \mathbb{R}^n$,

$$\begin{aligned}&\langle R(x+y), R(x+y)\rangle_{\mathbb{R}^n} = \langle x+y, x+y\rangle_{\mathbb{R}^n} \\ \Longrightarrow\quad &\langle Rx, Rx\rangle_{\mathbb{R}^n} + \langle Ry, Ry\rangle_{\mathbb{R}^n} + 2\langle Rx, Ry\rangle_{\mathbb{R}^n} = \langle x, x\rangle_{\mathbb{R}^n} + \langle y, y\rangle_{\mathbb{R}^n} + 2\langle x, y\rangle_{\mathbb{R}^n} \\ \Longrightarrow\quad &\langle Rx, Ry\rangle_{\mathbb{R}^n} = \langle x, y\rangle_{\mathbb{R}^n},\end{aligned}$$

as desired.

(iii) $\Longrightarrow$ (iv) Letting $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ be the standard basis for $\mathbb{R}^n$ we have

$$\langle \boldsymbol{R}\boldsymbol{e}_j, \boldsymbol{R}\boldsymbol{e}_k\rangle_{\mathbb{R}^n} = \langle \boldsymbol{e}_j, \boldsymbol{e}_k\rangle_{\mathbb{R}^n}, \qquad j,k \in \{1,\dots,n\}.$$

We have

$$\langle \boldsymbol{e}_j, \boldsymbol{e}_k\rangle_{\mathbb{R}^n} = \boldsymbol{I}_n(j,k) = \begin{cases} 1, & j = k, \\ 0, & j \neq k \end{cases}$$

and a direct calculation shows that

$$\langle \boldsymbol{R}\boldsymbol{e}_j, \boldsymbol{R}\boldsymbol{e}_k\rangle_{\mathbb{R}^n} = \sum_{i=1}^{n} \boldsymbol{R}(i,j)\boldsymbol{R}(i,k) = (\boldsymbol{R}^T\boldsymbol{R})(j,k).$$

Thus $\boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}_n$. From Theorem I-5.1.42 this means that $\boldsymbol{R}$ is invertible with inverse $\boldsymbol{R}^T$. This means that we also have $\boldsymbol{R}\boldsymbol{R}^T = \boldsymbol{I}_n$.

(iv) $\Longrightarrow$ (v) This was proved in the preceding part of the proof.

(v) $\Longrightarrow$ (i) We first note that a direct computation shows that

$$\langle \boldsymbol{A}\boldsymbol{x}, \boldsymbol{y}\rangle_{\mathbb{R}^n} = \langle \boldsymbol{x}, \boldsymbol{A}^T\boldsymbol{y}\rangle_{\mathbb{R}^n} \tag{1.11}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and $\boldsymbol{A} \in \mathrm{Mat}_{n\times n}(\mathbb{R})$; this idea will be revealed in a more general setting in . If $\boldsymbol{R}$ is invertible with inverse $\boldsymbol{R}^T$ we have

what?

$$\begin{aligned} & \boldsymbol{R}^T\boldsymbol{R} = \boldsymbol{I}_n \\ \Longrightarrow\quad & \boldsymbol{R}^T\boldsymbol{R}\boldsymbol{x} = \boldsymbol{x}, \qquad \boldsymbol{x} \in \mathbb{R}^n \\ \Longrightarrow\quad & \langle \boldsymbol{R}^T\boldsymbol{R}\boldsymbol{x}, \boldsymbol{x}\rangle_{\mathbb{R}^n} = \langle \boldsymbol{x}, \boldsymbol{x}\rangle_{\mathbb{R}^n}, \qquad \boldsymbol{x} \in \mathbb{R}^n \\ \Longrightarrow\quad & \langle \boldsymbol{R}\boldsymbol{x}, \boldsymbol{R}\boldsymbol{x}\rangle_{\mathbb{R}^n} = \langle \boldsymbol{x}, \boldsymbol{x}\rangle_{\mathbb{R}^n}, \qquad \boldsymbol{x} \in \mathbb{R}^n, \end{aligned}$$

using (1.11). Thus $\|\boldsymbol{R}\boldsymbol{x}\|_{\mathbb{R}^n} = \|\boldsymbol{x}\|_{\mathbb{R}^n}$ for every $\boldsymbol{x} \in \mathbb{R}^n$. Therefore,

$$\|\boldsymbol{R}\boldsymbol{x}_1 - \boldsymbol{R}\boldsymbol{x}_2\|_{\mathbb{R}^n} = \|\boldsymbol{R}(\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_{\mathbb{R}^n} = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_{\mathbb{R}^n}$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^n$, meaning that $\boldsymbol{R}$ is an isometry. ■

Clearly linear isometries are very special. They are also very important, although we will not engage in a general investigation of these until . For now we just make a definition.

when?

1.3.19 Definition (Orthogonal matrix) A matrix $\boldsymbol{R} \in \mathrm{Mat}_{n\times n}(\mathbb{R})$ is ***orthogonal*** if it is a linear isometry. The set of orthogonal $n \times n$ matrices is denoted by $\mathsf{O}(n)$ and is called the ***orthogonal group*** in n-dimensions. •

Since we call $\mathsf{O}(n)$ the orthogonal group, it ought to be a group. The reader can verify that this is the case in Exercise 1.3.9.

With an understanding of linear isometries, it is possible to understand the structure of a general isometry. The following result gives the characterisation.

1.3.20 Theorem (Characterisation of isometries of $\mathbb{R}^n$) *A map* $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^n$ *is an isometry if and only if there exists* $\mathbf{R} \in \mathsf{O}(n)$ *and* $\mathbf{r} \in \mathbb{R}^n$ *such that*

$$\mathbf{f}(\mathbf{x}) = \mathbf{R}\mathbf{x} + \mathbf{r}, \qquad \mathbf{x} \in \mathbb{R}^n.$$

Proof First let us verify that the map $x \mapsto Rx + r$ is an isometry. We compute

$$\|(Rx_1 + r) - (Rx_2 + r)\|_{\mathbb{R}^n} = \|R(x_1 - x_2)\|_{\mathbb{R}^n} = \|x_1 - x_2\|_{\mathbb{R}^n},$$

using Theorem 1.3.18. Thus maps of the form given in the theorem statement are isometries.

Now suppose that f is an isometry. First suppose that f fixes $\mathbf{0} \in \mathbb{R}^n$: $f(\mathbf{0}) = \mathbf{0}$. We shall use the fact (see Exercise 1.3.1) that the Euclidean norm space satisfies the parallelogram law:

$$\|x + y\|^2_{\mathbb{R}^n} + \|x - y\|^2_{\mathbb{R}^n} = 2\Big(\|x\|^2_{\mathbb{R}^n} + \|y\|^2_{\mathbb{R}^n}\Big).$$

Using this equality, and the fact that f is an isometry fixing $\mathbf{0}$, we compute

$$\begin{aligned}\|f(x) + f(y)\|^2_{\mathbb{R}^n} &= 2\|f(x)\|^2_{\mathbb{R}^n} + 2\|f(y)\|^2_{\mathbb{R}^n} - \|f(x) - f(y)\|^2_{\mathbb{R}^n}\\ &= 2\|f(x) - f(\mathbf{0})\|^2_{\mathbb{R}^n} + 2\|f(y) - f(\mathbf{0})\|^2_{\mathbb{R}^n} - \|f(x) - f(y)\|^2_{\mathbb{R}^n}\\ &= 2\|x\|^2_{\mathbb{R}^n} + 2\|y\|^2_{\mathbb{R}^n} - \|x - y\|^2_{\mathbb{R}^n} = \|x + y\|^2_{\mathbb{R}^n}.\end{aligned} \tag{1.12}$$

By the polarization identity, see Exercise 1.3.1, we obtain

$$\langle x, y\rangle_{\mathbb{R}^n} = \tfrac{1}{2}\Big(\|x + y\|^2_{\mathbb{R}^n} - \|x\|^2_{\mathbb{R}^n} - \|y\|^2_{\mathbb{R}^n}\Big)$$

for every $x, y \in \mathbb{R}^n$. In particular, using (1.12) and the fact that f is an isometry fixing $\mathbf{0}$, we compute

$$\begin{aligned}\langle f(x), f(y)\rangle_{\mathbb{R}^n} &= \tfrac{1}{2}\Big(\|f(x) + f(y)\|^2_{\mathbb{R}^n} - \|f(x)\|^2_{\mathbb{R}^n} - \|f(y)\|^2_{\mathbb{R}^n}\Big)\\ &= \tfrac{1}{2}\Big(\|f(x) + f(y)\|^2_{\mathbb{R}^n} - \|f(x) - f(\mathbf{0})\|^2_{\mathbb{R}^n} - \|f(y) - f(\mathbf{0})\|^2_{\mathbb{R}^n}\Big)\\ &= \tfrac{1}{2}\Big(\|x + y\|^2_{\mathbb{R}^n} - \|x\|^2_{\mathbb{R}^n} - \|y\|^2_{\mathbb{R}^n}\Big) = \langle x, y\rangle_{\mathbb{R}^n}.\end{aligned} \tag{1.13}$$

We now claim that this implies that f is a linear map. Indeed, let $\{e_1, \dots, e_n\}$ be the standard basis for $\mathbb{R}^n$ and let $(x_1, \dots, x_n)$ be the components of $x \in \mathbb{R}^n$ in this basis (thus $x_i = \langle x, e_i\rangle_{\mathbb{R}^n}$, $i \in \{1, \dots, n\}$). Since

$$\langle f(e_i), f(e_j)\rangle_{\mathbb{R}^n} = \langle e_i, e_j\rangle_{\mathbb{R}^n}, \qquad i, j \in \{1, \dots, n\},$$

the vectors $\{f(e_1), \dots, f(e_n)\}$ form an orthonormal basis for $\mathbb{R}^n$ (see for the notion of an orthonormal basis). The components of $f(x)$ in this basis are given by $\langle f(x, f(e_i)\rangle_{\mathbb{R}^n}$, $i \in \{1, \dots, n\}$. By (1.13) this means that the components of $f(x)$ are precisely $(x_1, \dots, x_n)$. That is, what

$$f\Big(\sum_{i=1}^n x_i e_i\Big) = \sum_{i=1}^n x_i f(e_i).$$

Therefore, if f fixes $\mathbf{0} \in \mathbb{R}^n$ then f is linear and so, by Theorem 1.3.18, there exists $\boldsymbol{R} \in \mathsf{O}(n)$ such that $f(\boldsymbol{x}) = \boldsymbol{R}\boldsymbol{x}$. Thus the theorem holds when f fixes $\mathbf{0}$.

Now, suppose that f fixes not $\mathbf{0}$, but some other point $\boldsymbol{x}_0 \in \mathbb{R}^n$: $f(\boldsymbol{x}_0) = \boldsymbol{x}_0$. Then define $f_{\boldsymbol{x}_0} \colon \mathbb{R}^n \to \mathbb{R}^n$ by

$$f_{\boldsymbol{x}_0}(\boldsymbol{x}) = f(\boldsymbol{x} + \boldsymbol{x}_0) - \boldsymbol{x}_0,$$

and note that $f_{\boldsymbol{x}_0}(\mathbf{0}) = \mathbf{0}$. Thus $f_{\boldsymbol{x}_0}(\boldsymbol{x}) = \boldsymbol{R}(\boldsymbol{x})$ for some $\boldsymbol{R} \in \mathsf{O}(n)$. Therefore,

$$f(\boldsymbol{x}) = f_{\boldsymbol{x}_0}(\boldsymbol{x} - \boldsymbol{x}_0) + \boldsymbol{x}_0 = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{x}_0 - \boldsymbol{R}\boldsymbol{x}_0.$$

Thus the theorem holds when f fixes a general point in $\mathbb{R}^n$.

Finally, suppose that f maps $\boldsymbol{x}_1 \in \mathbb{R}^n$ to $\boldsymbol{x}_2 \in \mathbb{R}^n$. In this most general case define $f_{\boldsymbol{x}_1,\boldsymbol{x}_2} \colon \mathbb{R}^n \to \mathbb{R}^n$ by

$$f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(\boldsymbol{x}) = f(\boldsymbol{x}) - (\boldsymbol{x}_2 - \boldsymbol{x}_1),$$

noting that $f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(\boldsymbol{x}_1) = \boldsymbol{x}_1$. Therefore, by the previous part of the proof,

$$f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(\boldsymbol{x}) = \boldsymbol{R}\boldsymbol{x} + \boldsymbol{r}'$$

for some $\boldsymbol{R} \in \mathsf{O}(n)$ and some $\boldsymbol{r}' \in \mathbb{R}^n$. Thus we get the theorem by taking $\boldsymbol{r} = \boldsymbol{r}' + (\boldsymbol{x}_2 - \boldsymbol{x}_1)$. ■

Now that we have described the set of isometries, let us name them.

1.3.21 Definition (Euclidean group) The ***Euclidean group*** in n-dimensions is the set of isometries of $\mathbb{R}^n$ and is denoted by $\mathsf{E}(n)$. •

Of course, the Euclidean group is a group, as the reader may verify in Exercise 1.3.12.

Note that there are two fundamental sorts of isometries. The first are ***translations*** which are of the form $\boldsymbol{x} \mapsto \boldsymbol{x} + \boldsymbol{r}$ for some $\boldsymbol{r} \in \mathbb{R}^n$. The second fundamental sort of isometry are those that are linear: $\boldsymbol{x} \mapsto \boldsymbol{R}\boldsymbol{x}$ for $\boldsymbol{R} \in \mathsf{O}(n)$. These are called ***rotations***. Theorem 1.3.20 tells us that a general isometry is a rotation followed by a translation.

1.3.5 Continuity and operations on functions

In this section we prove the hoped for properties of continuous functions with respect to the algebraic and topological properties of Euclidean space. First of all let us note that if $A \subseteq \mathbb{R}^n$ then the set of $\mathbb{R}^m$-valued maps on A is a $\mathbb{R}$-vector space. Indeed, the operations of vector addition and scalar multiplication are defined by

$$(f + g)(\boldsymbol{x}) = f(\boldsymbol{x}) + g(\boldsymbol{x}), \qquad (af)(\boldsymbol{x}) = a(f(\boldsymbol{x})),$$

where $f, g \colon A \to \mathbb{R}^m$ and where $a \in \mathbb{R}$. These operations respect continuous functions.

1.3.22 Proposition (Continuity, and addition and scalar multiplication) *If* $A \subseteq \mathbb{R}^n$*, if* $\mathbf{f}, \mathbf{g}\colon A \to \mathbb{R}^m$ *are continuous, and if* $a \in \mathbb{R}$ *then* $\mathbf{f} + \mathbf{g}$ *and* $a\mathbf{f}$ *are continuous.*

Proof The proof differs from the relevant parts of the proof of Proposition I-3.1.15 only by change of notation so we omit it here. ■

1.3.23 Proposition (Continuity and composition) *Let* $A \subseteq \mathbb{R}^m$*,* $B \subseteq \mathbb{R}^m$ *and let* $\mathbf{f}\colon A \to \mathbb{R}^m$ *and* $\mathbf{g}\colon B \to \mathbb{R}^k$ *have the properties that* $\mathrm{image}(A) \subseteq B$ *and that* $\mathbf{f}$ *is continuous at* $\mathbf{x}_0 \in A$ *and* $\mathbf{g}$ *is continuous at* $\mathbf{f}(\mathbf{x}_0)$*. Then* $\mathbf{g} \circ \mathbf{f}$ *is continuous at* $\mathbf{x}_0$*.*

Proof This is proved in the same manner as Proposition I-3.1.16. ■

1.3.24 Proposition (Continuity and restriction) *If* $A \subseteq \mathbb{R}^n$*, if* $B \subseteq A$*, and if* $\mathbf{f}\colon A \to \mathbb{R}^m$ *be continuous at* $\mathbf{x}_0 \in B$*, then* $\mathbf{f}|B$ *is continuous at* $\mathbf{x}_0$*.*

Proof The manner of proof here is like that in Proposition I-3.1.17. ■

Note that the converse of the previous result is not generally true.

1.3.25 Example (Continuity and restriction) Define $f\colon \mathbb{R} \to \mathbb{R}$ by

$$f(x) = \begin{cases} 1, & x \in \mathbb{Z}, \\ 0, & x \notin \mathbb{Z}. \end{cases}$$

Then $f|\mathbb{Z}$ is continuous (it is constant), but f is not continuous at points in $\mathbb{Z}$. •

Let us also indicate how continuity interacts with products.

1.3.26 Proposition (Continuity and products) *The following statements hold:*

(i) *if* $A_j \subseteq \mathbb{R}^{n_j}$*,* $j \in \{1, \dots, k\}$*, and if* $\mathbf{f}\colon A_1 \times \cdots \times A_k \to \mathbb{R}^k$ *is continuous at* $(\mathbf{x}_{10}, \dots, \mathbf{x}_{k0})$*, then the maps*

$$\mathbf{x}_j \mapsto \mathbf{f}(\mathbf{x}_{10}, \dots, \mathbf{x}_j, \dots, \mathbf{x}_{k0}), \qquad j \in \{1, \dots, k\},$$

are continuous at $\mathbf{x}_{j0}$*;*

(ii) *if* $C \subseteq \mathbb{R}^k$ *and if* $\mathbf{g}\colon C \to \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ *is given by* $\mathbf{g}(\mathbf{z}) = (\mathbf{g}_1(\mathbf{z}), \dots, \mathbf{g}_k(\mathbf{z}))$ *for* $\mathbf{g}_j\colon C \to \mathbb{R}^{n_j}$*,* $j \in \{1, \dots, k\}$*, then* $\mathbf{g}$ *is continuous at* $\mathbf{z}_0 \in C$ *if and only if each of the maps* $\mathbf{g}_j$*,* $j \in \{1, \dots, k\}$*, are continuous at* $\mathbf{z}_0$*.*

Proof By induction it suffices to prove the result for $k = 2$. We denote $n_1 = m$, $n_2 = n$, and write a typical point in $\mathbb{R}^m \times \mathbb{R}^n$ as $(\boldsymbol{x}, \boldsymbol{y})$.

(i) Suppose that f is continuous at $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ and let $(\boldsymbol{x}_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence converging to $\boldsymbol{x}_0$. Then the sequence $((\boldsymbol{x}_j, \boldsymbol{y}_0))_{j \in \mathbb{Z}_{>0}}$ is easily verified to converge to $(\boldsymbol{x}, \boldsymbol{y}_0)$. Continuity of f and Theorem 1.3.2 ensures that

$$\lim_{j \to \infty} f(\boldsymbol{x}_j, \boldsymbol{y}_0) = f(\boldsymbol{x}_0, \boldsymbol{y}_0),$$

which in turn gives continuity of $\boldsymbol{x} \mapsto f(\boldsymbol{x}, \boldsymbol{y}_0)$ at $\boldsymbol{x}_0$ by Theorem 1.3.2. An entirely similar argument gives continuity of $\boldsymbol{y} \mapsto f(\boldsymbol{x}_0, \boldsymbol{y})$ at $\boldsymbol{y}_0$.

(ii) First suppose that g is continuous at z_0. Then, for a sequence $(z_j)_{j\in\mathbb{Z}_{>0}}$ in C converging to z_0, the sequence $((g_1(z_j), g_2(z_j)))_{j\in\mathbb{Z}_{>0}}$ converges to $(g_1(z_0), g_2(z_0))$ by Theorem 1.3.2. From Exercise 1.2.11 we know that the sequences $(g_1(z_j))_{j\in\mathbb{Z}_{>0}}$ and $(g_2(z_j))_{j\in\mathbb{Z}_{>0}}$ converge to $g_1(z_0)$ and $g_2(z_0)$, respectively. By Theorem 1.3.2 it follows that g_1 and g_2, respectively.

The argument can be reversed, using Exercise 1.2.11 and Theorem 1.3.2, to show that g is continuous at (x_0, y_0) if g_1 is continuous at x_0 and g_2 is continuous at y_0. ■

The reader will notice that an implication is missing from the preceding result. This is not an oversight.

1.3.27 Example (Discontinuous function continuous in both variables) Define $f\colon \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x_1, x_2) = \begin{cases} \frac{x_1^2 x_2}{x_1^4 + x_2^2}, & (x_1, x_2) \neq (0,0), \\ 0, & (x_1, x_2) = (0,0). \end{cases}$$

We claim that f is not continuous at $(0,0)$. Consider a point in $\mathbb{R}^2$ of the form (a, a^2) for $a \in \mathbb{R}$. At such points we have $f(a, a^2) = \frac{1}{2}$. Since $f(0,0) = 0$ and since any neighbourhood of $(0,0)$ contains a point of the form (a, a^2) for some $a \in \mathbb{R}^*$, it follows that f cannot be continuous at $(0,0)$.

However, the two functions

$$x_1 \mapsto f(x_1, 0) = 0, \quad x_2 \mapsto f(0, x_2) = 0$$

are obviously continuous. •

Let us finally consider the behaviour of continuity with respect to the operations of selection of maximums and minimums.

1.3.28 Proposition (Continuity and min and max) *If* $A \subseteq \mathbb{R}^n$ *and if* $f, g\colon I \to \mathbb{R}$ *are continuous functions, then the functions*

$$A \ni \mathbf{x} \mapsto \min\{f(\mathbf{x}), g(\mathbf{x})\} \in \mathbb{R}, \qquad A \ni \mathbf{x} \mapsto \max\{f(\mathbf{x}), g(\mathbf{x})\} \in \mathbb{R}$$

are continuous.

Proof Let $x_0 \in A$ and let $\epsilon \in \mathbb{R}_{>0}$. Let us first assume that $f(x_0) > g(x_0)$. That is to say, assume that $(f - g)(x_0) \in \mathbb{R}_{>0}$. Continuity of f and g ensures that there exists $\delta_1 \in \mathbb{R}_{>0}$ such that if $x \in \mathsf{B}^n(\delta_1, x_0) \cap A$ then $(f - g)(x) \in \mathbb{R}_{>0}$. That is, if $x \in \mathsf{B}^n(\delta_1, x_0) \cap A$ then

$$\min\{f(x), g(x)\} = g(x), \quad \max\{f(x), g(x)\} = f(x).$$

Continuity of f ensures that there exists $\delta_2 \in \mathbb{R}_{>0}$ such that if $x \in \mathsf{B}^n(\delta_2, x_0) \cap A$ then $|f(x) - f(x_0)| < \epsilon$. Similarly, continuity of f ensures that there exists $\delta_3 \in \mathbb{R}_{>0}$ such that if $x \in \mathsf{B}^n(\delta_3, x_0) \cap A$ then $|g(x) - g(x_0)| < \epsilon$. Let $\delta_4 = \min\{\delta_1, \delta_2\}$. If $x \in \mathsf{B}(\delta_4, x_0) \cap A$ then

$$|\min\{f(x), g(x)\} - \min\{f(x_0), g(x_0)\}| = |g(x) - g(x_0)| < \epsilon$$

and

$$|\max\{f(x), g(x)\} - \max\{f(x_0), g(x_0)\}| = |f(x) - f(x_0)| < \epsilon.$$

This gives continuity of the two functions in this case. Similarly, swapping the rôle of f and g, if $f(x_0) < g(x_0)$ one can arrive at the same conclusion. Thus we need only consider the case when $f(x_0) = g(x_0)$. In this case, by continuity of f and g, choose $\delta \in \mathbb{R}_{>0}$ such that $|f(x) - f(x_0)| < \epsilon$ and $|g(x) - g(x_0)| < \epsilon$ for $x \in \mathsf{B}(\delta, x_0) \cap A$. Then let $x \in \mathsf{B}(\delta, x_0) \cap A$. If $f(x) \geq g(x)$ then we have

$$|\min\{f(x), g(x)\} - \min\{f(x_0), g(x_0)\}| = |g(x) - g(x_0)| < \epsilon$$

and

$$|\max\{f(x), g(x)\} - \max\{f(x_0), g(x_0)\}| = |f(x) - f(x_0)| < \epsilon.$$

This gives the result in this case, and one similarly gets the result when $f(x) < g(x)$. ■

1.3.6 Continuity, and compactness and connectedness

As we saw in Section I-3.1.4 for single-variable functions, continuity acts nicely with respect to certain topological notions including compactness and connectedness. We give these results here in the multivariable case, noting that there is a great deal in common with the single-variable case. Thus we will go through this fairly quickly.

1.3.29 Proposition (The continuous image of a compact set is compact) *If* $A \subseteq \mathbb{R}^n$ *is compact and if* $\mathbf{f}\colon A \to \mathbb{R}^m$ *is continuous, then* $\mathrm{image}(\mathbf{f})$ *is compact.*

Proof Let $(U_i)_{i\in I}$ be an open cover of $\mathrm{image}(\boldsymbol{f})$. Then $(\boldsymbol{f}^{-1}(U_i))_{i\in I}$ is an open cover of A, and so there exists a finite subset $(i_1, \dots, i_k) \subseteq I$ such that $\cup_{j=1}^k \boldsymbol{f}^{-1}(U_{i_k}) = A$. It is then clear that $(\boldsymbol{f}(\boldsymbol{f}^{-1}(U_{i_1})), \dots, \boldsymbol{f}(\boldsymbol{f}^{-1}(U_{i_k})))$ covers $\mathrm{image}(\boldsymbol{f})$. Moreover, by Proposition I-1.3.5, $\boldsymbol{f}(\boldsymbol{f}^{-1}(U_{i_j})) \subseteq U_{i_j}$, $j \in \{1, \dots, k\}$. Thus $(U_{i_1}, \dots, U_{i_k})$ is a finite subcover of $(U_i)_{i\in I}$. ■

The following properties of functions interact well with compactness.

1.3.30 Definition (Bounded map) For an subset $A \subseteq \mathbb{R}^n$, a map $\boldsymbol{f}\colon A \to \mathbb{R}^m$ is:

(i) ***bounded*** if there exists $M \in \mathbb{R}_{>0}$ such that $\mathrm{image}(\boldsymbol{f}) \subseteq \overline{\mathsf{B}}^n(M, \mathbf{0})$;

(ii) ***locally bounded*** if $\boldsymbol{f}|K$ is bounded for every compact subset $K \subseteq A$;

(iii) ***unbounded*** if it is not bounded. •

1.3.31 Theorem (Continuous functions on compact sets are bounded) *If* $A \subseteq \mathbb{R}^n$ *is compact, then a continuous function* $\mathbf{f}\colon A \to \mathbb{R}^m$ *is bounded.*

Proof Let $x \in A$. As $\boldsymbol{f}$ is continuous, there exists $\delta \in \mathbb{R}_{>0}$ so that $\|\boldsymbol{f}(y) - \boldsymbol{f}(x)\|_{\mathbb{R}^m} < 1$ provided that $\|y - x\|_{\mathbb{R}^n} < \delta$. In particular, if $x \in A$, there is a neighbourhood U_x of x such that $\|\boldsymbol{f}(y)\|_{\mathbb{R}^n} \leq \|\boldsymbol{f}(x)\|_{\mathbb{R}^m} + 1$ for all $x \in U_x \cap A$. Thus $\boldsymbol{f}$ is bounded on $U_x \cap A$. This can be done for each $x \in A$, so defining a family of open sets $(U_x)_{x\in A}$. Clearly $A \subseteq \cup_{x\in A} U_x$, and so, by Theorem 1.2.35, there exists a finite collection of points $x_1, \dots, x_k \in A$ such that $A \subseteq \cup_{j=1}^k U_{x_j}$. Obviously for any $x \in A$,

$$\|\boldsymbol{f}(x)\|_{\mathbb{R}^m} \leq 1 + \max\{\boldsymbol{f}(x_1), \dots, \boldsymbol{f}(x_k)\},$$

thus showing that f is bounded. ■

1.3.32 Theorem (Continuous functions on compact sets achieve their extreme values) *If* $A \subseteq \mathbb{R}^n$ *is a compact interval and if* $f\colon A \to \mathbb{R}$ *is continuous, then there exist points* $\mathbf{x}_{\min}, \mathbf{x}_{\max} \in A$ *such that*

$$f(\mathbf{x}_{\min}) = \inf\{f(\mathbf{x}) \mid \mathbf{x} \in A\}, \quad f(\mathbf{x}_{\max}) = \sup\{f(\mathbf{x}) \mid \mathbf{x} \in A\}.$$

Proof It suffices to show that f achieves its maximum on A since if f achieves its maximum, then $-f$ will achieve its minimum. So let $M = \sup\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in A\}$, and suppose that there is no point $\boldsymbol{x}_{\max} \in A$ for which $f(\boldsymbol{x}_{\max}) = M$. Then $f(\boldsymbol{x}) < M$ for each $\boldsymbol{x} \in A$. For a given $\boldsymbol{x} \in A$ we have

$$f(\boldsymbol{x}) = \tfrac{1}{2}(f(\boldsymbol{x}) + f(\boldsymbol{x})) < \tfrac{1}{2}(f(\boldsymbol{x}) + M).$$

Continuity of f ensures that there is an open set $U_{\boldsymbol{x}}$ containing $\boldsymbol{x}$ such that, for each $\boldsymbol{y} \in U_{\boldsymbol{x}} \cap A$, $f(\boldsymbol{y}) < \frac{1}{2}(f(\boldsymbol{x}) + M)$. Since $A \subseteq \cup_{\boldsymbol{x}\in A} U_{\boldsymbol{x}}$, by the Heine–Borel theorem, there exists a finite number of points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ such that $A \subseteq \cup_{j=1}^k U_{\boldsymbol{x}_j}$. Let $m = \max\{f(\boldsymbol{x}_1), \dots, f(\boldsymbol{x}_k)\}$ so that, for each $\boldsymbol{y} \in I_{\boldsymbol{x}_j}$, and for each $j \in \{1, \dots, k\}$, we have

$$f(\boldsymbol{y}) < \tfrac{1}{2}(f(\boldsymbol{x}_j) + M) < \tfrac{1}{2}(m + M),$$

which shows that $\frac{1}{2}(m + M)$ is an upper bound for f. However, since f attains the value m on A, we have $m < M$ and so $\frac{1}{2}(m + M) < M$, contradicting the fact that M is the least upper bound. Thus our assumption that f cannot attain the value M on A is false. ■

As in the single-variable case we saw that continuity and compactness conspire to give uniform continuity. This is true in the multivariable case as well, and serves to further establish the connection between "compactness" and "uniformly."

1.3.33 Theorem (Heine–Cantor Theorem) *Let* $A \subseteq \mathbb{R}^n$ *be compact. If* $\mathbf{f}\colon A \to \mathbb{R}^m$ *is continuous, then it is uniformly continuous.*

Proof Let $\boldsymbol{x} \in A$ and let $\epsilon \in \mathbb{R}_{>0}$. Since $\boldsymbol{f}$ is continuous, then there exists $\delta_{\boldsymbol{x}} \in \mathbb{R}_{>0}$ such that, if $\boldsymbol{y} \in \mathsf{B}^n(\delta_x, \boldsymbol{x}) \cap A$ then $\boldsymbol{f}(\boldsymbol{y}) \in \mathsf{B}^m(\frac{\epsilon}{2}, \boldsymbol{f}(\boldsymbol{x}))$. Note that $A \subseteq \cup_{\boldsymbol{x}\in A} \mathsf{B}^n(\frac{\delta_x}{2}, \boldsymbol{x})$, so that the open sets $(\mathsf{B}^n(\frac{\delta_x}{2}, \boldsymbol{x}))_{\boldsymbol{x}\in A}$ cover A. By definition of compactness, there then exists a finite number of these open sets that cover A. Denote this finite family by $(\mathsf{B}^n(\frac{\delta_{x_1}}{2}, \boldsymbol{x}_1), \dots, \mathsf{B}^n(\frac{\delta_{x_k}}{2}, \boldsymbol{x}_k))$ for some $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in A$. Take $\delta = \frac{1}{2}\min\{\delta_{\boldsymbol{x}_1}, \dots, \delta_{\boldsymbol{x}_k}\}$. Now let $\boldsymbol{x}, \boldsymbol{y} \in A$ satisfy $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} < \delta$. Then there exists $j \in \{1, \dots, k\}$ such that $\boldsymbol{x} \in \mathsf{B}^n(\frac{\delta_{x_k}}{2}, \boldsymbol{x}_j)$. We also have

$$\|\boldsymbol{y} - \boldsymbol{x}_j\|_{\mathbb{R}^n} \le \|\boldsymbol{y} - \boldsymbol{x}\|_{\mathbb{R}^n} + \|\boldsymbol{x} - \boldsymbol{x}_j\|_{\mathbb{R}^n} < \delta_{\boldsymbol{x}_j},$$

using the triangle inequality. Therefore,

$$\|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^m} \le \|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{x}_j)\|_{\mathbb{R}^m} + \|\boldsymbol{f}(\boldsymbol{x}_j) - \boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^m} < \epsilon,$$

again using the triangle inequality. Since this holds for *any* $\boldsymbol{x} \in A$, it follows that $\boldsymbol{f}$ is uniformly continuous. ■

Now let us turn to connectedness and its relation to continuity.

1.3.34 Proposition (The continuous image of a (path) connected set is (path) connected) *If* $A \subseteq \mathbb{R}^n$ *is (path) connected and if* $\mathbf{f}: A \to \mathbb{R}^m$ *is continuous, then* $\mathbf{f}(A)$ *is (path) connected.*

Proof Suppose that $f(A)$ is not connected. Then there exist nonempty separated sets S and T such that $f(A) = S \cup T$. Let $S' = f^{-1}(S)$ and $T' = f^{-1}(T)$ so that $A = S' \cup T'$. By Propositions 1.2.28 and I-1.3.5, and since $f^{-1}(\mathrm{cl}(S))$ is closed, we have

$$\mathrm{cl}(S') = \mathrm{cl}(f^{-1}(S)) \subseteq \mathrm{cl}(f^{-1}(\mathrm{cl}(S)) = f^{-1}(\mathrm{cl}(S)).$$

Therefore, by Proposition I-1.3.5,

$$\mathrm{cl}(S') \cap T' \subseteq f^{-1}(\mathrm{cl}(S)) \cap f^{-1}(T) = f^{-1}(\mathrm{cl}(S) \cap T) = \varnothing.$$

We also similarly have $S' \cap \mathrm{cl}(T') = \varnothing$. Thus A is not connected, which gives the result for connectedness.

Now suppose that A is path connected and let $y_1, y_2 \in \mathrm{image}(f)$. Thus $y_1 = f(x_1)$ and $y_2 = f(x_2)$. Since A is path connected there exists a continuous path $\gamma: [a, b] \to A$ such that $\gamma(a) = x_1$ and $x_2 = \gamma(b)$. The path $f \circ \gamma$ in image(f) is continuous by Proposition 1.3.23 and has the property that $f \circ \gamma(a) = y_1$ and $f \circ \gamma(b) = y_2$. Thus image(f) is path connected. ■

In multiple variables, the Intermediate Value Theorem is actually significantly more revealing than it is in the single-variable case. Indeed, it illustrates that it is connectivity that is the crucial ingredient in the theorem.

1.3.35 Theorem (Intermediate Value Theorem) *Let* $A \subseteq \mathbb{R}^n$ *be connected and let* $f: A \to \mathbb{R}$ *be continuous. If* $\mathbf{x}_1, \mathbf{x}_2 \in A$ *then, for any* $y \in [f(\mathbf{x}_1), f(\mathbf{x}_2)]$, *there exists* $\mathbf{x} \in A$ *such that* $f(\mathbf{x}) = y$.

Proof From Proposition 1.3.34 we know that image(f) is connected and so is an interval by virtue of Theorem I-2.5.34. The points $f(x_1)$ and $f(x_2)$ lie in this interval, and so too, therefore, does every point between $f(x_1)$ and $f(x_1)$. ■

1.3.7 Homeomorphisms

As we become more mature, we become more able to digest advanced concepts. In this section introduce the idea of a homeomorphism. The idea of a homeomorphism is an important one; it plays the rôle played by isomorphism for algebraic objects. That is, a homeomorphism gives the backdrop for understanding those things that are "continuous invariants," meaning that they are invariant under continuous maps. Obviously, not just any continuous map will do. Upon reflection, the following sort of continuous map is the reasonable one to generate the notion of "continuous invariants."

1.3.36 Definition (Homeomorphism, homeomorphic) If $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$, a ***homeomorphism*** from A to B is a continuous bijection $f: A \to B$ whose inverse $f^{-1}: B \to A$ is also continuous. If $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ have the property that there exists a homeomorphism $f: A \to B$, then A and B are ***homeomorphic***. •

The following result is obvious, but is worth recording so it is out in the open.

1.3.37 Proposition ("Homeomorphic" is an equivalence relation) *If* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$, $\mathrm{B} \subseteq \mathbb{R}^{\mathrm{m}}$, *and* $\mathrm{C} \subseteq \mathbb{R}^{\mathrm{k}}$ *then the following statements hold:*

(i) A *is homeomorphic to* A*;*

(ii) *if* A *is homeomorphic to* B *then* B *is homeomorphic to* A*;*

(iii) *if* A *and* B *are homeomorphic and if* B *and* C *are homeomorphic, then* A *and* C *are homeomorphic.*

In other words, the relation "$\mathrm{A} \sim \mathrm{B}$ *if* A *and* B *are homeomorphic" between subsets of Euclidean spaces is an equivalence relation.*

Let us give some examples so that we develop some feeling for what a homeomorphism is and is not.

1.3.38 Examples (Homeomorphisms)

1. For any subset $A \subseteq \mathbb{R}^n$ the identity map $\mathrm{id}_A\colon A \to A$ is a homeomorphism. This is easy to check.
2. Let $\mathsf{V} \subseteq \mathbb{R}^n$ be a subspace and let $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$ be a basis for V. We claim that the map $\mathsf{L}\colon \mathbb{R}^k \to \mathsf{V}$ defined by

$$\mathsf{L}(x_1, \dots, x_k) = x_1\boldsymbol{v}_1 + \dots + x_k\boldsymbol{v}_k$$

is a homeomorphism. Certainly it is bijective (if you do not immediately see this, this means you need to read up on linear independence in Section I-4.5.3). To see that it is continuous, denote

$$M = \max\{\|\boldsymbol{v}_1\|_{\mathbb{R}^k}, \dots, \|\boldsymbol{v}_k\|_{\mathbb{R}^k}\}$$

and, for $\epsilon \in \mathbb{R}_{>0}$, choose $\delta = \frac{\epsilon}{kM}$. If $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^k} < \delta$ then $|x_j - y_j| < \delta$ for every $j \in \{1, \dots, k\}$. Thus we have, for $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^k} < \delta$,

$$\begin{aligned}\|\mathsf{L}(\boldsymbol{x}) - \mathsf{L}(\boldsymbol{y})\|_{\mathbb{R}^n} &= \|(x_1 - y_1)\boldsymbol{v}_1 + \dots + (x_k - y_k)\boldsymbol{v}_k\|_{\mathbb{R}^m} \\ &\le |x_1 - y_1|\|\boldsymbol{v}_1\|_{\mathbb{R}^m} + \dots + |x_k - y_k|\|\boldsymbol{v}_k\|_{\mathbb{R}^m} \\ &< kM\delta = \epsilon.\end{aligned}$$

This shows that L is continuous, indeed uniformly continuous, consistent with Proposition 1.3.16.

Now let us show that L^{-1} is continuous. By Theorem I-4.5.26 we take vectors $\boldsymbol{v}_{k+1}, \dots, \boldsymbol{v}_n \in \mathbb{R}^n$ such that $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_n\}$ is a basis for $\mathbb{R}^n$. Then define a linear map $\hat{\mathsf{L}}\colon \mathbb{R}^n \to \mathbb{R}^k$ by asking that

$$\hat{\mathsf{L}}(\boldsymbol{v}_j) = \begin{cases} \boldsymbol{e}_j, & j \in \{1, \dots, k\}, \\ \boldsymbol{0}, & j \in \{k+1, \dots, n\}, \end{cases}$$

cf. Theorem I-4.5.24. By Proposition 1.3.16 we know that $\hat{\mathsf{L}}$ is continuous and by Proposition 1.3.24 we know that, as a result, $\mathsf{L} = \hat{\mathsf{L}}|\mathsf{V}$ is continuous.

3. Let $A = (0, \infty)$ and let $B = \mathbb{R}$. Define $f: A \to B$ by $f(x) = \log(x)$. By Proposition I-3.8.6 f is a homeomorphism. Since every open unbounded interval that is a strict subset of $\mathbb{R}$ is of the form (a, ∞) or $(-\infty, b)$, one can easily modify our construction to show that all such intervals homeomorphic to $\mathbb{R}$; see Exercise 1.3.13.
4. Let $A = (0, 1) \subseteq \mathbb{R}$ and let $B = \mathbb{R}$. The map $f: A \to B$ given by $f(x) = \tan^{-1}(\pi(x - \frac{1}{2}))$ is a homeomorphism, this following from Proposition I-3.8.20. It is possible to modify this example to show that every bounded open interval is homeomorphic to $\mathbb{R}$; see Exercise 1.3.13.
5. Let $A = (-\pi, \pi] \subseteq \mathbb{R}$ and let

$$B = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1\}.$$

Thus B is the unit circle in $\mathbb{R}^2$. Any point in $(x_1, x_2) \in B$ is expressed in the form $(x_1, x_2) = (\cos(x), \sin(x))$ for some $x \in \mathbb{R}$; see Proposition I-3.8.19(iii). Moreover, if we ask that $x \in (-\pi, \pi]$ then there exists a unique such point such that $(x_1, x_2) = (\cos(x), \sin(x))$. That is, the map $f: A \to B$ defined by $f(x) = (\cos(x), \sin(x))$ is a bijection. We claim that f is continuous. This follows directly from the continuity of cos and sin; see Proposition I-3.8.19(i). We also claim that f^{-1} is discontinuous at $(-1, 0)$. To see why this is so, note that $f^{-1}(-1, 0) = \pi$. Now let $(x_1, x_2) \in B$ satisfy $x_1, x_2 < 0$. Then $f^{-1}(x_1, x_2) \in (-\pi, -\frac{\pi}{2})$. Thus, for all such points we have

$$|f^{-1}(x_1, x_2) - f^{-1}(-1, 0)| > \tfrac{\pi}{2}.$$

However, for any $\delta \in \mathbb{R}_{>0}$ there exists a point $(x_1, x_2) \in B$ with $(x_1, x_2) < 0$ such that $\|(x_1, x_2) - (-1, 0)\|_{\mathbb{R}^2} < \delta$. Thus $f^{-1}(\mathsf{B}^2(\delta, (-1, 0))) \not\subseteq \mathsf{B}^1(1, \pi)$, giving discontinuity of f^{-1} at $(-1, 0)$.

The point is that a continuous bijection need not be a homeomorphism. •

The second of the preceding examples is worth expounding on a little.

1.3.39 Remark (The topology of a subspace) If one has two bases $\{v_1, \dots, v_k\}$ and $\{v_1', \dots, v_k'\}$ for a subspace $\mathsf{V} \subseteq \mathbb{R}^n$, these induce as in Example 2 two homeomorphisms $\mathsf{L}, \mathsf{L}': \mathbb{R}^k \to \mathsf{V}$. Thus, by Proposition 1.3.37, the subspace V is homeomorphic to $\mathbb{R}^k$ in a manner not depending in the use of a basis to establish the homeomorphism. In other words, a k-dimensional subspace inherits in a natural way the topological structure of $\mathbb{R}^k$. We shall use this fact in the sequel to, without loss of generality, work with all of $\mathbb{R}^n$ rather than a subspace of $\mathbb{R}^n$. This is a special case of the general principle that it is sometimes convenient to work with a set homeomorphic to the one in a given problem. •

As mentioned in the preparatory comments of this section, the notion of a homeomorphism has the intent of allowing us to consider properties that are "continuous invariants." The reader may understand this idea by comparing it to a statement from linear algebra; Proposition I-4.5.30 says that the dimension of a vector space is an isomorphism invariant (indeed, it is actually the only isomorphism invariant). We are interested in properties of subsets of Euclidean space that are homeomorphism invariant. Let us make an actual definition so we know what we are talking about.

1.3.40 Definition (Topological invariant) A property P is a ***topological invariant*** if, whenever $A \subseteq \mathbb{R}^n$ has property P then every subset $B \subseteq \mathbb{R}^m$ that is homeomorphic to A also has property P. •

Unlike the comparatively simple situation in linear algebra where the only isomorphism invariant is dimension, an exhaustive list of topological invariants (okay, well, "simple" topological invariants) seems not to be practical. However, let us list some topological invariants that we have already encountered, as well as some concepts that are not topological invariants.

1.3.41 Theorem (Some topological invariants) *The following properties are topological invariants:*

(i) *compactness;*

(ii) *connectedness;*

(iii) *path-connectedness;*

(iv) *existence of a continuous map into given subset* $\mathrm{S} \subseteq \mathbb{R}^{\mathrm{n}}$*;*

(v) *existence of a continuous map from a given subset* $\mathrm{S} \subseteq \mathbb{R}^{\mathrm{n}}$*.*

The following properties are not topological invariants:

(vi) *openness;*

(vii) *closedness;*

(viii) *boundedness;*

(ix) *total boundedness.*

Proof Suppose that $A \subseteq \mathbb{R}^n$ is a compact (resp. connected, path connected) and let $f\colon A \to B \subseteq \mathbb{R}^m$ be a homeomorphism. Then B is compact (resp. connected, path connected) by Proposition 1.3.29 (resp. Proposition 1.3.34). This gives the first three properties as being topological invariants.

That the last two properties asserted as being topological invariants are, in fact, topological invariants is a consequence of the composition of continuous maps being continuous, i.e., of Proposition 1.3.23. For example, if A is homeomorphic to B with a homeomorphism $h\colon A \to B$ and if $f\colon S \to A$ is continuous, then $h \circ f$ is a continuous map of S into B.

To show that a property is not a topological invariant it suffices to give an example, and this is what we do for the last four parts of the theorem.

Note that $A = \mathbb{R}$ is open and is homeomorphic to the set

$$B = \{(x_1, x_2) \in \mathbb{R}^2 \mid x_2 = 0\}$$

which is not open. Also, B is closed and homeomorphic to $(0, 1)$ (cf. Example 1.3.38–4) which is not closed.

The same example will suffice in each of the last two statements. Indeed, let $A = (0,1)$ which is both bounded and totally bounded. However, B is homeomorphic to $\mathbb{R}$ by Example 1.3.38–4, and $\mathbb{R}$ is neither bounded nor totally bounded. ■

1.3.42 Remark ("Intrinsic" versus "extrinsic" properties) It is interesting to note that the three topological invariants we give in the preceding theorem differ in a fundamental way from the four properties that are not topological invariants. Indeed, note that the four properties that are not topological invariants have to do, not with the set itself, but with its properties as a subset of the Euclidean space in which it resides. The three properties that *are* topological invariants, however, have to do with the set itself, not how it sits in Euclidean space. There is something in this observation. •

Note that Example 1.3.38–5 shows that for a map to be a homeomorphism it is not sufficient for it to be a continuous bijection. Let us now turn to cases where it is possible to make this inference.

1.3.43 Theorem (Continuous bijections on compact sets are homeomorphisms) *If* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *is compact and if* $\mathbf{f}\colon \mathrm{A} \to \mathbb{R}^{\mathrm{m}}$ *is a continuous injection then* $\mathbf{f}$ *is a homeomorphism of* A *with* $\mathrm{image}(\mathbf{f})$*.*

Proof Let us denote $B = \mathrm{image}(f)$ and $f^{-1}\colon B \to A$ the inverse. By Proposition 1.3.29 it follows that B is compact. We claim that the image of a relatively closed subset of A is relatively closed in B. Thus let $C \subseteq A$ be relatively closed so that, by Corollary 1.2.36, C is compact. Then $f(C)$ is a compact subset of B and so relatively closed, again by Corollary 1.2.36. Therefore, f maps relatively closed sets to relatively closed sets, and so also maps relatively open sets to relatively open sets by virtue of f being a bijection. Thus f^{-1} is continuous. ■

In our proof of the topological invariance of the property of openness in Proposition 1.3.41 we showed that the open subset $\mathbb{R} \subseteq \mathbb{R}$ is homeomorphic to the non-open subset of $\mathbb{R}^2$ consisting of the x_1-axis. The reader might protest that this is unfair, and that to make the statement interesting we should produce an open subset of $\mathbb{R}^n$ that is homeomorphic to a subset of $\mathbb{R}^n$ (the same "n," note) that is not open. It turns out, however, that such an example does not exist. This is nontrivial, but we will give the proof here anyway. The following theorem which gives the desired conclusion is an extremely important one, and is difficult to prove by "elementary" methods; the result is most naturally viewed from the point of view of either dimension theory or algebraic topology (see Section 1.10.10 for references). Our long but elementary proof relies crucially on Theorem 1.10.51, which itself relies on the Weierstrass Approximation Theorem (Theorem 1.7.4), the Tietze Extension Theorem (Theorem 1.10.43), and the Brouwer Fixed Point Theorem (Theorem 1.11.6).

1.3.44 Theorem (Domain Invariance Theorem) *If* U *is an open subset of* $\mathbb{R}^{\mathrm{n}}$ *and if* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$ *is an injective continuous map, then* $\mathrm{image}(\mathbf{f})$ *is open and* $\mathbf{f}$ *is a homeomorphism between* U *and* $\mathrm{image}(\mathbf{f})$*.*

Proof We begin with a couple of lemmata that contain the crux of the proof. We note that

$$\mathbb{S}^{n-1} = \{x \in \mathbb{R}^n \mid \|x\|_{\mathbb{R}^n} = 1\}$$

denotes the unit sphere in $\mathbb{R}^n$.

1 **Lemma** *If* $C \subseteq \mathbb{R}^n$ *is closed then the following two statements regarding* $x \in C$ *are equivalent:*

(i) $x \in \mathrm{bd}(C)$*;*

(ii) for any relative neighbourhood V *of* x *in* C *there exists a relative neighbourhood* U *of* x *in* C *having the properties that*

(a) $U \subseteq V$ *and*

(b) if $g\colon C \setminus U \to \mathbb{S}^{n-1}$ *is continuous then there exists a continuous map* $\hat{g}\colon C \to \mathbb{S}^{n-1}$ *such that* $g = \hat{g}|(C \setminus U)$.

Proof (i) $\Longrightarrow$ (ii) Suppose that $x_0 \in \mathrm{bd}(C)$ and let V be a relative neighbourhood of x_0 in C. By Proposition 1.2.50 there exists an open subset V' in $\mathbb{R}^n$ such that $V = C \cap V'$. Then let $\epsilon \in \mathbb{R}_{>0}$ be sufficiently small that $\mathsf{B}^n(\epsilon, x_0) \subseteq V'$ and take $U = C \cap \mathsf{B}^n(\epsilon, x_0)$. Let

$$\mathbb{S}^{n-1}(\epsilon, x_0) = \{x \in \mathbb{R}^n \mid \|x - x_0\|_{\mathbb{R}^n} = \epsilon\}$$

be the sphere of radius ϵ centred at x_0, i.e., $\mathbb{S}^{n-1}(\epsilon, x_0) = \mathrm{bd}(\mathsf{B}^n(\epsilon, x_0))$. Define

$$C_0 = C \cap \overline{\mathsf{B}}^n(\epsilon, x_0), \quad C_1 = C \setminus \mathsf{B}^n(\epsilon, x_0),$$

noting that $C = C_0 \cup C_1$, that $C_0 \cap C_1 \subseteq \mathbb{S}^{n-1}(\epsilon, x_0)$ and that

$$C_0 \cap \mathbb{S}^{n-1}(\epsilon, x_0) = C_1 \cap \mathbb{S}^{n-1}(\epsilon, x_0).$$

Now let $g\colon C_1 \to \mathbb{S}^{n-1}$ be continuous. We shall define the extension $\hat{g}\colon C \to \mathbb{S}^{n-1}$ by defining it on C_0 and then showing that the resulting map is consistently defined on $C_0 \cap C_1$.

The first observation to make is that $\mathbb{S}^{n-1}$ is homeomorphic to $\mathbb{S}^{n-1}(\epsilon, x_0)$ (see Exercise 1.3.16) and so any homeomorphism $\iota\colon \mathbb{S}^{n-1} \to \mathbb{S}^{n-1}(\epsilon, x_0)$ of these two sets will give a continuous map $h = \iota \circ g\colon C_1 \to \mathbb{S}^{n-1}(\epsilon, x_0)$. We shall define a map $\hat{h}\colon C \to \mathbb{S}^{n-1}(\epsilon, x_0)$ which extends h, and the desired map $\hat{g}$ is then given by $\hat{g} = \iota^{-1} \circ \hat{h}$.

Next note that by Corollary 1.10.52 there exists a continuous map $h'\colon \mathbb{S}^{n-1}(\epsilon, x_0) \to \mathbb{S}^{n-1}(\epsilon, x_0)$ that agrees with h on $C_1 \cap \mathbb{S}^{n-1}(\epsilon, x_0)$.

To define $\hat{h}$ on C_0 we note that, since $x_0 \in \mathrm{bd}(C)$, there exists a point $x_1 \in \mathsf{B}^n(\epsilon, x) - C$. If $x \in C_0 \subseteq \overline{\mathsf{B}}^n(\epsilon, x_0)$ define

$$y_x = x_1 + \frac{\|x - x_1\|^2_{\mathbb{R}^n} - \|x - x_0\|^2_{\mathbb{R}^n} + \epsilon\|x - x_1\|_{\mathbb{R}^n}}{\|x - x_1\|^2_{\mathbb{R}^n}}(x - x_1). \tag{1.14}$$

Note that y_x is the point on the sphere $\mathbb{S}^{n-1}(\epsilon, x_0)$ obtained as the intersection of the sphere with the ray from x_1 passing through x. The essential feature of y_x is that it is a continuous function of x. We take $\hat{h}(x) = h'(y_x)$. Since $y_x = x$ for $x \in C_0 \cap \mathbb{S}^{n-1}(\epsilon, x)$ we have $\hat{h}(x) = h(x)$ for $x \in C_0 \cap C_1$.

Thus we can take $\hat{\boldsymbol{h}}(\boldsymbol{x}) = \boldsymbol{h}(\boldsymbol{x})$ for $\boldsymbol{x} \in C_1$ and the result will be a consistently defined continuous $\mathbb{S}^{n-1}(\epsilon, \boldsymbol{x}_0)$-valued map on C.

(ii) $\implies$ (i) Now suppose that $\boldsymbol{x}_0 \in \operatorname{int}(C)$. Then there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\epsilon, \boldsymbol{x}_0) \subseteq C$. Now let U be a relatively open neighbourhood of $\boldsymbol{x}_0$ in C with the property that $U \subseteq \mathsf{B}^n(\epsilon, \boldsymbol{x}_0)$. Now define $\boldsymbol{h}\colon C \setminus U \to \mathbb{S}^{n-1}(\epsilon, \boldsymbol{x}_0)$ by

$$\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{x}_0 + \epsilon \frac{\boldsymbol{x} - \boldsymbol{x}_0}{\|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n}}.$$

Note that $\boldsymbol{h}(\boldsymbol{x})$ is the point on the sphere $\mathbb{S}^{n-1}(\epsilon, \boldsymbol{x}_0)$ which is the intersection of the sphere with the ray from $\boldsymbol{x}_0$ passing through $\boldsymbol{x}$. Now suppose that there exists $\hat{\boldsymbol{h}}\colon C \to \mathbb{S}^{n-1}(\epsilon, \boldsymbol{x}_0)$ which extends $\boldsymbol{h}$. Since $\mathbb{S}^{n-1}(\epsilon, \boldsymbol{x}_0) \subseteq C \setminus U$ and since $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{x}$ for $\boldsymbol{x} \in \mathbb{S}^{n-1}(\epsilon, \boldsymbol{x}_0)$, it follows that $\hat{\boldsymbol{h}}|\overline{\mathsf{B}}^n(\epsilon, \boldsymbol{x}_0)$ is a retraction of $\overline{\mathsf{B}}^n(\epsilon, \boldsymbol{x}_0)$ onto $\mathbb{S}^{n-1}(\epsilon, \boldsymbol{x}_0)$. This is not possible by Proposition 1.11.9, after recalling, as above, that $\overline{\mathsf{B}}^n(\epsilon, \boldsymbol{x}_0)$ is homeomorphic to $\mathbb{D}^n$ (see Exercise 1.3.15). ▼

2 Lemma *If $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$ are closed and if $\boldsymbol{f}\colon A \to B$ is a homeomorphism then $\boldsymbol{f}(\operatorname{bd}(A)) = \operatorname{bd}(B)$.*

Proof By Proposition 1.3.6 we have $\boldsymbol{f}(\operatorname{bd}(A)) \subseteq \operatorname{bd}(B)$. Let $\boldsymbol{y} \in \operatorname{bd}(B)$ so that $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ for some $\boldsymbol{x} \in A$. Let V be a relative neighbourhood of $\boldsymbol{x}$ in A. Then continuity of $\boldsymbol{f}^{-1}$ gives $V' = \boldsymbol{f}(V)$ as a relative neighbourhood of $\boldsymbol{y}$ in B. By Lemma 1 there exists a relative neighbourhood U' of $\boldsymbol{y}$ in B such that

1. $U' \subseteq V'$ and
2. if $\boldsymbol{g}'\colon B \setminus U' \to \mathbb{S}^{n-1}$ is continuous then there exists a continuous map $\hat{\boldsymbol{g}}'\colon B \to \mathbb{S}^{n-1}$ such that $\boldsymbol{g} = \hat{\boldsymbol{g}}|(B \setminus U)$.

Then define $U = \boldsymbol{f}^{-1}(U')$ which, by continuity of $\boldsymbol{f}$, is a relative neighbourhood of $\boldsymbol{x}$. Moreover, $U \subseteq V$. Now let $\boldsymbol{g}\colon A \setminus U \to \mathbb{S}^{n-1}$ be continuous. Then $\boldsymbol{g}' \triangleq \boldsymbol{g} \circ \boldsymbol{f}^{-1}$ is a continuous map from $B \setminus U'$ to $\mathbb{S}^{n-1}$. There that exists $\hat{\boldsymbol{g}}'\colon B \to \mathbb{S}^{n-1}$ extending $\boldsymbol{g}'$. Now define $\hat{\boldsymbol{g}}\colon A \to \mathbb{S}^{n-1}$ by $\hat{\boldsymbol{g}} = \hat{\boldsymbol{g}}' \circ \boldsymbol{f}$. The continuity of $\hat{\boldsymbol{g}}$ allows us to conclude that $\boldsymbol{x} \in \operatorname{bd}(A)$ and so $\boldsymbol{y} \in \boldsymbol{f}(\operatorname{bd}(A))$. ▼

Proceeding with the proof, if $U' \subseteq U$ is open we claim that $\boldsymbol{f}(U')$ is open. Let us denote $V' = \boldsymbol{f}(U')$ and let $\boldsymbol{y} \in V'$. Thus $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$ for some $\boldsymbol{x} \in U'$. Let $r \in \mathbb{R}_{>0}$ be such that $\overline{\mathsf{B}}^n(r, \boldsymbol{x}) \subseteq U'$. Then $\boldsymbol{f}|\mathsf{B}^n(r, \boldsymbol{x})$ is a homeomorphism onto its image by Theorem 1.3.43. Therefore, $\boldsymbol{f}(\boldsymbol{x}) \in \operatorname{int}(\boldsymbol{f}(U'))$ by Lemma 2. This shows that every point in V' is an interior point and so V' is open. In other words, if $V = \boldsymbol{f}(U)$ then $\boldsymbol{f}^{-1}\colon V \to U$ is continuous, as desired. ■

As we have said, the Domain Invariance Theorem is very important. Let us explore interpretations of it and some important consequences of it. First of all, the following result follows directly, and gives a useful topological invariance property.

1.3.45 Corollary (Openness in $\mathbb{R}^n$ is a topological invariant) *Let $n \in \mathbb{Z}_{>0}$. Then the property "A is an open subset of $\mathbb{R}^n$" is a topological invariant.*

Proof Suppose that $A \subseteq \mathbb{R}^n$ is open and that $B \subseteq \mathbb{R}^n$ is homeomorphic to A. Then there exists a homeomorphism $\boldsymbol{f}\colon A \to B$. The map $\boldsymbol{f} \circ i_B\colon A \to \mathbb{R}^n$ is then injective and continuous. Thus, by Theorem 1.3.44, its image is open. But its image is B. ■

Now let us attempt to understand the Domain Invariance Theorem by trying to gain some appreciation for why it is nontrivial. Let us see if we can do this for $n = 1$. Thus we consider an open subset $U \subseteq \mathbb{R}$ and a continuous injective map $f\colon U \to \mathbb{R}$. Since U is open, it is a union of intervals by Proposition I-2.5.6. Thus we may as well restrict our attention to the case when U is an interval. In this case a continuous function will be strictly monotonically increasing or strictly monotonically decreasing; this is Exercise I-3.3.1. In the case when f is differentiable with positive or negative derivative the Domain Invariance Theorem is more or less obvious since, in this case, f is approximately linear with a positive or negative slope. So the real content of the Domain Invariance Theorem in this case occurs at points where f is either not differentiable, or has derivative zero. Let us then give an example which illustrates some facets of the Domain Invariance Theorem.

1.3.46 Example (A continuous, strictly monotonically increasing function that is not differentiable on a dense set) We give another peculiar sort of function to illustrate a rather subtle point. We define a sequence of functions $(f_k)_{k\in\mathbb{Z}_{\geq 0}}$ on $[0,1]$ as follows. We take $f_0(x) = x$. To define f_1 take

$$f_1(0) = f_0(0) = 0,\ f_1(1) = f_0(1) = 1,$$
$$f_1(\tfrac{1}{2}) = (1-\alpha)f_0(0) + \alpha f_0(1) = \alpha,$$

where $\alpha \in (0,1)$. We then define f_1 on $(0,\frac{1}{2})$ and $(\frac{1}{2},1)$ by asking that it be continuous and linear on these intervals. Now suppose that we have defined $f_0, f_1, \dots, f_k$ and define f_{k+1} as follows. We require that

$$f_{k+1}(\tfrac{j}{2^k}) = f_k(\tfrac{j}{2^k}), \qquad j \in \{0,1,\dots,2^k\},$$
$$f_{k+1}(\tfrac{2j+1}{2^{k+1}}) = (1-\alpha)f_k(\tfrac{j}{2^k}) + \alpha f_k(\tfrac{j+1}{2^k}), \qquad j \in \{0,1,\dots,2^k-1\}.$$

We then define f_{k+1} on all of $[0,1]$ by asking that it be linear on each of the subintervals $[\frac{j}{2^{k+1}}, \frac{j+1}{2^{k+1}}]$, $j \in \{0,1,\dots,2^{k+1}-1\}$. We then define $f_\alpha\colon [0,1] \to \mathbb{R}$ by

$$f_\alpha(x) = \lim_{k\to\infty} f_k(x), \qquad x \in [0,1].$$

In Figure 1.8 we show the first step in this construction for various α. The idea is that this construction is applied recursively to each on the subintervals on which the function is linear.

Now we record some of the features of this function by proving a series of lemmata. First let us show that the definition of f_α makes sense.

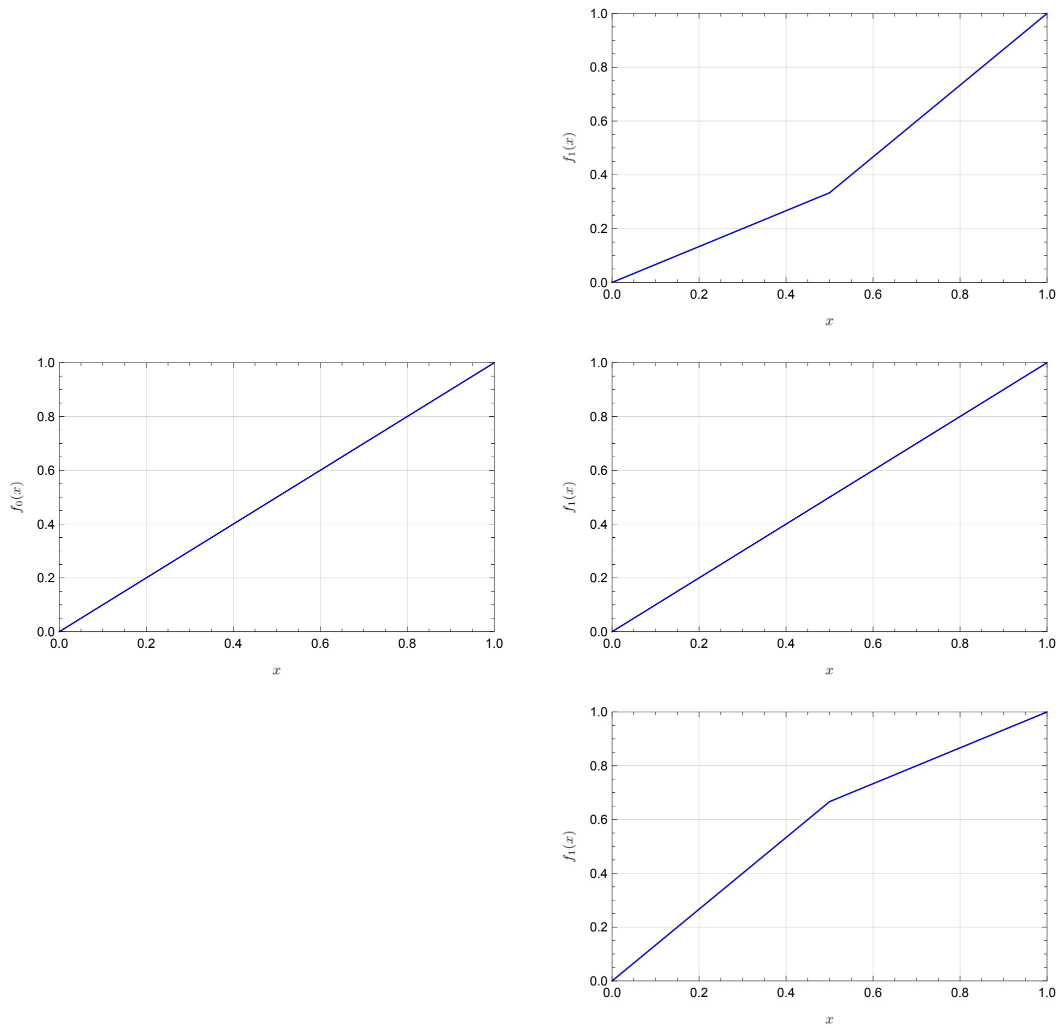

Figure 1.8 The first step in constructing the function f_α for $\alpha < \frac{1}{2}$ (top), $\alpha = \frac{1}{2}$ (middle), and $\alpha > \frac{1}{2}$ (bottom)

1 Lemma *For each* $\mathrm{x} \in [0,1]$ *and* $\alpha \in (0,1)$ *the limit* $\lim_{\mathrm{k}\to\infty} \mathrm{f}_\mathrm{k}(\mathrm{x})$ *exists.*

Proof Using the linearity of f_k between the endpoints of the intervals used to define it, we compute

$$\begin{aligned} f_{k+1}(\tfrac{2j+1}{2^{k+1}}) - f_k(\tfrac{2j+1}{2^{k+1}}) &= (1-\alpha)f_k(\tfrac{j}{2^k}) + \alpha f_k(\tfrac{j+1}{2^k}) - \tfrac{1}{2}(f_k(\tfrac{j}{2^k}) + f_k(\tfrac{j+1}{2^k})) \\ &= (\alpha - \tfrac{1}{2})(f_k(\tfrac{j+1}{2^k}) - f_k(\tfrac{j}{2^k})), \end{aligned}$$

for $k \in \mathbb{Z}_{\geq 0}$ and $j \in \{0,1,\dots,2^k-1\}$. Thus we have three cases.

1. When $\alpha = \frac{1}{2}$ we have $f_{k+1}(\frac{2j+1}{2^{k+1}}) = f_k(\frac{2j+1}{2^{k+1}})$, giving $f_{k+1} = f_k$.
2. When $\alpha < \frac{1}{2}$ then the sequence $(f_k(\frac{2j+1}{2^{k+1}}))_{k\in\mathbb{Z}_{\geq 0}}$ is strictly monotonically decreasing and bounded below by zero. Thus it converges.

3. When $\alpha > \frac{1}{2}$ then the sequence $(f_k(\frac{2j+1}{2^{k+1}}))_{k\in\mathbb{Z}_{\geq 0}}$ is strictly monotonically increasing and bounded above by zero. Thus it converges. ▼

2 Lemma *The function* f_α *is strictly monotonically increasing for* $\alpha \in (0,1)$.

Proof We shall first show that each of the functions f_k, $k \in \mathbb{Z}_{\geq 0}$, are strictly monotonically increasing. We show this by induction. It is clear that f_0 is strictly monotonically increasing. Now suppose that f_k is strictly monotonically increasing. We have

$$\begin{aligned} f_{k+1}(\tfrac{j}{2^k}) - f_{k+1}(\tfrac{2j+1}{2^{k+1}}) &= f_k(\tfrac{j}{2^k}) - f_{k+1}(\tfrac{2j+1}{2^{k+1}}) \\ &= f_k(\tfrac{j}{2^k}) - (1-\alpha)f_k(\tfrac{j}{2^k}) - \alpha f_k(\tfrac{j+1}{2^k}) \\ &= \alpha(f_k(\tfrac{j}{2^k}) - f_k(\tfrac{j+1}{2^k})) < 0 \end{aligned}$$

and

$$\begin{aligned} f_{k+1}(\tfrac{j+1}{2^k}) - f_{k+1}(\tfrac{2j+1}{2^{k+1}}) &= f_k(\tfrac{j+1}{2^k}) - f_{k+1}(\tfrac{2j+1}{2^{k+1}}) \\ &= f_k(\tfrac{j+1}{2^k}) - (1-\alpha)f_k(\tfrac{j}{2^k}) - \alpha f_k(\tfrac{j+1}{2^k}) \\ &= (1-\alpha)(f_k(\tfrac{j+1}{2^k}) - f_k(\tfrac{j}{2^k})) > 0. \end{aligned}$$

Thus we have

$$f_{k+1}(\tfrac{j}{2^k}) < f_{k+1}(\tfrac{2j+1}{2^{k+1}}) < f_{k+1}(\tfrac{j+1}{2^k}), \qquad j \in \{0,1,\dots,2^k-1\}.$$

Since f_{k+1} is defined to be linear on the subintervals $[\frac{j}{2^{k+1}}, \frac{j+1}{2^{k+1}}]$, $j \in \{0,1,\dots,2^{k+1}-1\}$, it follows that f_{k+1} is strictly monotonically increasing. It therefore follows that f_α is nondecreasing. To show that f_α is, in fact, strictly monotonically increasing, let $x_1, x_2 \in [0,1]$ satisfy $x_1 < x_2$. By Exercise I-2.1.5 let $j,k \in \mathbb{Z}_{>0}$ satisfy $\frac{j}{2^k} \in (x_1,x_2)$. We consider three cases.

1. In the case when $\alpha = \frac{1}{2}$ it follows easily that f_α is strictly monotonically increasing since, as we showed in Lemma 1, $f_{1/2}(x) = x$.
2. If $\alpha > \frac{1}{2}$ we have
$$f_\alpha(x_1) \leq f_\alpha(\tfrac{j}{2^k}) = f_k(\tfrac{j}{2^k}) \leq f_k(x_2) \leq f_\alpha(x_2).$$
3. When $\alpha < \frac{1}{2}$ we have
$$f_\alpha(x_1) \leq f_k(x_1) < f_k(\tfrac{j}{2^k}) = f_\alpha(\tfrac{j}{2^k}) \leq f_\alpha(x_2).$$ ▼

3 Lemma *The function* f_α *is continuous for* $\alpha \in (0,1)$.

Proof Let us first make a preliminary construction. We call a sequence $([a_k, b_k])_{k\in\mathbb{Z}_{\geq 0}}$ of subintervals of $[0,1]$ ***binary*** if $a_0 = 1$ and $b_0 = 1$, and if, for each $k \in \mathbb{Z}$, either

1. $a_{k+1} = a_k$ and $b_{k+1} = b_k - \frac{1}{2^{k+1}}$ or
2. $a_{k+1} = a_k + \frac{1}{2^{k+1}}$ and $b_{k+1} = b_k$.

Thus, for example, either $[a_1, b_1] = [0, \frac{1}{2}]$ or $[a_1, b_2] = [\frac{1}{2}, 1]$. If $([a_k, b_k])_{k\in\mathbb{Z}_{\geq 0}}$ is a binary sequence, if $k \in \mathbb{Z}_{\geq 0}$, and if $a_{k+1} = a_k$, then we compute

$$\begin{aligned} f_\alpha(b_{k+1}) - f_\alpha(a_{k+1}) &= f_{k+1}(b_{k+1}) - f_{k+1}(a_{k+1}) \\ &= (1-\alpha) f_k(a_j) + \alpha f_k(b_j) - f_k(a_k) \\ &= \alpha(f_k(b_j) - f_k(a_k)). \end{aligned}$$

In the case when $b_{k+1} = b_k$ we similarly compute

$$\begin{aligned} f_\alpha(b_{k+1}) - f_\alpha(a_{k+1}) &= f_{k+1}(b_{k+1}) - f_{k+1}(a_{k+1}) \\ &= f_k(b_k) - ((1-\alpha) f_k(a_k) + \alpha f_k(b_k)) \\ &= (1-\alpha)(f_k(b_k) - f_k(a_k)). \end{aligned}$$

Therefore, using $f_0(b_0) - f_0(a_0) = 1$, a trivial inductive argument gives

$$f_\alpha(b_k) - f_\alpha(a_k) = \prod_{j=1}^{k} \sigma_j,$$

where $\sigma_j \in \{\alpha, 1-\alpha\}$, depending on whether $a_j = a_{j-1}$ or $b_j = b_{j=1}$. In any case, the above computations show that

$$f_\alpha(b_k) - f_\alpha(a_k) \leq \begin{cases} (1-\alpha)^k, & \alpha \leq \frac{1}{2}, \\ \alpha^k, & \alpha > \frac{1}{2}. \end{cases}$$

Now we show the continuity of f_α. Let $\epsilon \in \mathbb{R}_{>0}$ and let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $(1-\alpha)^N < \frac{\epsilon}{2}$ if $\alpha \leq \frac{1}{2}$ or $\alpha^N < \frac{\epsilon}{2}$ if $\alpha > \frac{1}{2}$. Let $x_0 \in (0,1)$ and let $([a_k, b_k])_{k\in\mathbb{Z}_{\geq 0}}$ and $([a'_k, b'_k])_{k\in\mathbb{Z}_{\geq 0}}$ be binary intervals such that $a_N < x_0$, $x_0 < b'_N$, and $b_N = a'_N$. (By choosing N large enough we can ensure that $a_N > 0$ and $b'_N < 1$.) Then let $\delta \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^1(\delta, x_0) \subseteq [a_N, b'_N]$. Then we have

$$f_\alpha(b'_N) - f_\alpha(a_N) < \epsilon \quad \Longrightarrow \quad |f_\alpha(x) - f_\alpha(x_0)| < \epsilon, \ x \in \mathsf{B}^1(\delta, x_0),$$

by monotonicity of f_α. Continuity of f_α at 0 and 1 is shown in a similar manner, so we forgo the routine details. ▼

4 Lemma *Suppose that* $x \in [0,1]$ *has a binary expansion* $x = \sum_{j=1}^{\infty} \frac{x_j}{2^j}$ *with* $x_j \in \{0,1\}$, $j \in \mathbb{Z}_{>0}$, *and suppose that the sets*

$$\{j \in \mathbb{Z}_{>0} \mid x_j = 0\}, \quad \{j \in \mathbb{Z}_{>0} \mid x_j = 1\}$$

are infinite, i.e., suppose that x *is irrational in base* 2*. Then* $f'_\alpha(x) = 0$*. In particular,* f_α *is differentiable with zero derivative on a subset of* $[0,1]$ *that has full measure.*

Proof Since x is irrational in base 2 it follows that for each $k \in \mathbb{Z}$ there exists a unique $j \in \mathbb{Z}_{\geq 0}$ such that $x \in (\frac{j}{2^k}, \frac{j+1}{2^k})$ (the binary irrationality of x ensures that the endpoints are not included in the interval $(\frac{j}{2^k}, \frac{j+1}{2^k})$). Moreover, if we write

$$\frac{j}{2^k} = \frac{y_1}{2} + \cdots + \frac{y_n}{2^n}$$

as the binary decimal expansion, then we have

$$a_k \triangleq \frac{l}{2^k} = \frac{y_1}{2} + \cdots + \frac{y_k}{2^k} < x < \frac{y_1}{2} + \cdots + \frac{y_k}{2^k} + \frac{1}{2^k} = \frac{l+1}{2^k} \triangleq b_k,$$

which implies that $y_j = x_j$, $j \in \{1, \dots, k\}$. Therefore, if $x_k = 0$ then

$$a_k = a_{k-1}, \; b_k = a_{k-1} + \tfrac{1}{2^k} = \tfrac{a_{k-1}+b_{k-1}}{2},$$

and if $x_k = 1$ then

$$a_k = a_{k-1} + \tfrac{1}{2^k} = \tfrac{a_{k-1}+b_{k-1}}{2}, \; b_k = b_{k-1}.$$

Therefore, if $x_k = 0$ then

$$\begin{aligned}\frac{f_\alpha(b_k) - f_\alpha(a_k)}{\frac{1}{2^k}} &= 2^k\Big((1-\alpha)f_{k-1}(a_{k-1}) + \alpha f_{k-1}(b_{k-1}) - f_{k-1}(a_{k-1})\Big) \\ &= 2^k\alpha(f_{k-1}(b_{k-1}) - f_{k-1}(a_{k-1}))\end{aligned}$$

and if $x_k = 1$ then

$$\begin{aligned}\frac{f_\alpha(b_k) - f_\alpha(a_k)}{\frac{1}{2^k}} &= 2^k\Big(f_{k-1}(b_{k-1}) - (1-\alpha)f_{k-1}(a_{k-1}) - \alpha f_{k-1}(b_{k-1})\Big) \\ &= 2^k(1-\alpha)(f_{k-1}(b_k) - f_{k-1}(a_k)).\end{aligned}$$

In either case, we have

$$\frac{f_\alpha(b_k) - f_\alpha(a_k)}{\frac{1}{2^k}} = 2(x_k + (-1)^{x_k}\alpha)2^{k-1}(f_{k-1}(b_k) - f_{k-1}(a_k)),$$

and so a simple induction gives

$$\frac{f_\alpha(b_k) - f_\alpha(a_k)}{\frac{1}{2^k}} = \prod_{j=1}^{k} 2(x_j + (-1)^{x_j}\alpha).$$

Thus

$$f'(x) = \lim_{k\to\infty} \frac{f_\alpha(b_k) - f_\alpha(a_k)}{\frac{1}{2^k}} = 0$$

since $\alpha, (1-\alpha) < 1$.

The final assertion follows since the irrational numbers in base 2 have measure 1. This can be proved in exactly the same way as it is proved in base 10; see Exercise I-2.1.4. ▼

In Figure 1.9 we show the graph of f_α for a few α's. Since this function is

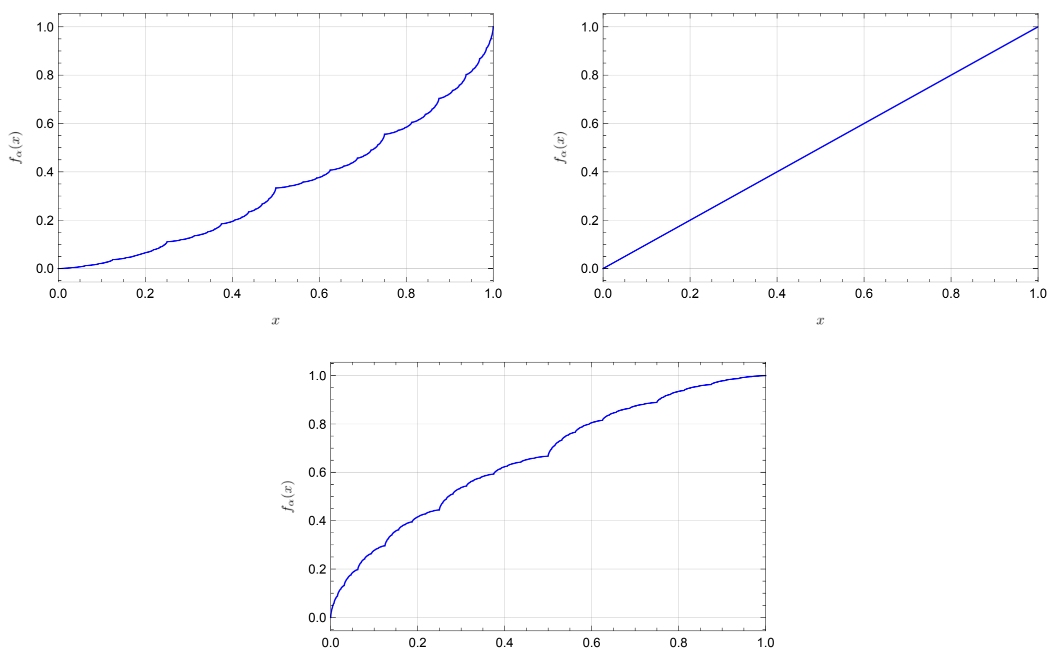

Figure 1.9 The function f_α for $\alpha = \frac{1}{3}$ (top left), $\alpha = \frac{1}{2}$ (top left), and $\alpha = \frac{2}{3}$ (bottom)

continuous and monotonically increasing it is injective by Exercise I-3.3.1. Therefore, by the Domain Invariance Theorem, $f|(0,1)$ is a homeomorphism onto $(0,1)$. In particular, the Domain Invariance Theorem allows us to conclude that f^{-1} is continuous. This may not be perfectly clear from the construction.

Interestingly, there are a number of places where the function f_α comes up in applications. The most common of these is in the "bold play" strategy in probability. The situation is this. A gambler possesses a fraction $x \in [0,1]$ of what she wants, and wishes to play a game at even money (i.e., the same amount is either paid out on a loss or collected on a win) until the desired goal is achieved or the gambler is bankrupt. The probability of winning a game is the quantity $\alpha \in (0,1)$. It then turns out that the probability of eventual success is $f_\alpha(x)$. Note that if $\alpha < \frac{1}{2}$ (i.e., the game is biased against the gambler) then the gambler must start with a fraction $x > \frac{1}{2}$ of the desired goal in order to have a greater that 50% chance of winning. This makes sense, I guess. •

Let us give another consequence of the Domain Invariance Theorem. One expects, and it is true, that two Euclidean spaces are homeomorphic if and only if they have the same dimension. Perhaps this seems "obvious," but it becomes less so the more one gets to know about the possible complex behaviour of continuous maps between Euclidean spaces and their subsets. Indeed, the following theorem is intimately and essentially connected to the Domain Invariance Theorem.

1.3.47 Theorem (Dimension Invariance Theorem) *The sets* $\mathbb{R}^n$ *and* $\mathbb{R}^m$ *are homeomorphic if and only if* $m = n$.

Proof Since "homeomorphic" is an equivalence relation, we suppose without loss of generality that $m \leq n$. Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is a homeomorphism. Consider the m-dimensional subspace V of $\mathbb{R}^n$ defined by

$$\mathsf{V} = \{x \in \mathbb{R}^n \mid x_{m+1} = \cdots = x_n = 0\}.$$

By Example 1.3.38–2 we know that V is homeomorphic to $\mathbb{R}^m$. That there exists a homeomorphism $g\colon \mathbb{R}^m \to \mathsf{V}$. Therefore, the composition of homeomorphisms being a homeomorphism, $g \circ f\colon \mathbb{R}^n \to \mathsf{V}$ is a homeomorphism. By the Domain Invariance Theorem this means that V is open in $\mathbb{R}^n$, and this is the case if and only if $m = n$. ■

1.3.8 Notes

Theorem 1.3.44 on "invariance of domain" is due to Brouwer [1912]. For a "basic" result, it is rather difficult to prove, and its proof properly belongs to the domains of dimension theory ([Hurewicz and Wallman 1941] is the classical reference here) and algebraic topology (Munkres [1984] has a good treatment).

Exercises

1.3.1 Answer the following questions:

(a) Verify that the Euclidean inner product satisfies the ***parallelogram law***:

$$\|x_1 + x_2\|^2_{\mathbb{R}^n} + \|x_1 - x_2\|^2_{\mathbb{R}^n} = 2\Big(\|x_1\|^2_{\mathbb{R}^n} + \|x_2\|^2_{\mathbb{R}^n}\Big).$$

(b) Give an interpretation of the parallelogram law in $\mathbb{R}^2$.

(c) Verify that the Euclidean inner product satisfies the ***polarisation identity***:

$$4\langle x_1, x_2\rangle_{\mathbb{R}^n} = \langle x_1 + x_2, x_1 + x_2\rangle_{\mathbb{R}^n} - \langle x_1 - x_2, x_1 - x_2\rangle_{\mathbb{R}^n}.$$

1.3.2 Let $A \subseteq \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}^m$ be a map. Show that f is continuous at $x_0 \in A$ if and only if the components of f are continuous at x_0.

1.3.3 For $A \subseteq \mathbb{R}^n$, show that $f\colon A \to \mathbb{R}^m$ is continuous if and only if $f^{-1}(B)$ is relatively closed in A for every closed subset B of $\mathbb{R}^m$.

1.3.4 Let $U \subseteq \mathbb{R}^n$ and $V \subseteq \mathbb{R}^m$ be open and let $f\colon U \times V \to \mathbb{R}$ be uniformly continuous and bounded. Define

$$\underline{f}\colon U \to \mathbb{R} \qquad\qquad \overline{f}\colon U \to \mathbb{R}$$
$$\boldsymbol{x} \mapsto \sup\{f(\boldsymbol{x},\boldsymbol{y}) \mid \boldsymbol{y} \in V,\} \qquad \boldsymbol{x} \mapsto \sup\{f(\boldsymbol{x},\boldsymbol{y}) \mid \boldsymbol{y} \in V.\}$$

Show that $\underline{f}$ and $\overline{f}$ are continuous.

1.3.5 Is the preimage of a (path) connected set under a continuous map (path) connected?

1.3.6 Consider the subset

$$S = \{(x_1, 0) \in \mathbb{R}^2 \mid x_1 \in \mathbb{R}\} \cup \{(0, x_2) \in \mathbb{R}^2 \mid x_2 > 0\}$$

of $\mathbb{R}^2$ and the subset $A = \{(x_1, 0) \mid x_1 \in \mathbb{R}\}$ of S.

(a) Is A relatively open in S?
(b) Is A relatively closed in S?
(c) Determine $\mathrm{int}_S(A)$, $\mathrm{cl}_S(A)$, and $\mathrm{bd}_S(A)$.

1.3.7 Show that the image of an affine map is an affine subspace.

1.3.8 Let $\boldsymbol{R} \in \mathsf{O}(3)$.

(a) Show that $\boldsymbol{R}$ has at least one real eigenvalue and that its magnitude must 1.

Let $\boldsymbol{v}$ be an eigenvector for the real eigenvalue ± 1 and let $\boldsymbol{v}^\perp$ be the subspace orthogonal to $\boldsymbol{v}$.

(a) Show that $\boldsymbol{R}(\boldsymbol{v}^\perp) \subseteq \boldsymbol{v}^\perp$.
(b) Argue that if $\boldsymbol{R} \neq \boldsymbol{I}_3$ then $\boldsymbol{R}$ has no eigenvectors that are not collinear with $\boldsymbol{v}$?
Hint: *Use the fact that* $\mathbf{v}^\perp$ *is two-dimensional.*
(c) Which of the preceding parts of the exercise fail if $\boldsymbol{R} \in \mathsf{O}(n)$ for $n \neq 3$?

1.3.9 Answer the following questions.

(a) Show that $\mathsf{O}(n)$ is a group with the group operation given by matrix multiplication.
(b) Is $\mathsf{O}(n)$ a subspace of the $\mathbb{R}$-vector space $\mathrm{Mat}_{n\times n}(\mathbb{R})$?

1.3.10 Show that if $\boldsymbol{R} \in \mathsf{O}(n)$ then $\det \boldsymbol{R} \in \{-1, 1\}$.

1.3.11 Show that if $\boldsymbol{R} \in \mathsf{O}(n)$ and if $\lambda \in \mathbb{C}$ is an eigenvalue for the complexification $\boldsymbol{R}_{\mathbb{C}}$, then $|\lambda| = 1$.

1.3.12 Show that $\mathsf{E}(n)$ is a group with the group operation of map composition. Be sure to explicitly given the formulae for the product of two elements and the inverse of an element.

1.3.13 Let $I \subseteq \mathbb{R}$ be an open interval. Explicitly construct a homeomorphism from I to $\mathbb{R}$.

1.3.14 Show that $\mathsf{B}^n(1, \mathbf{0})$, the open ball of radius 1 centred at the origin in $\mathbb{R}^n$, is homeomorphic to $\mathbb{R}^n$.

1.3.15 Show that the following sets are homeomorphic:

1. $\mathbb{D}^n = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_{\mathbb{R}^n} \leq 1\}$;
2. $\mathbb{D}^n(r, \boldsymbol{x}_0) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n} \leq \epsilon\}$ where $r \in \mathbb{R}_{>0}$ and $\boldsymbol{x}_0 \in \mathbb{R}^n$;
3. a fat compact rectangle R;
4. $\mathbb{S}^n_+ = \{\boldsymbol{x} \in \mathbb{S}^n \subseteq \mathbb{R}^{n+1} \mid x_{n+1} \geq 0\}$.

1.3.16 Show that the following sets are homeomorphic:

1. $\mathbb{S}^n = \{\boldsymbol{x} \in \mathbb{R}^{n+1} \mid \|\boldsymbol{x}\|_{\mathbb{R}^{n+1}} = 1\}$;
2. $\mathbb{S}^n(r, \boldsymbol{x}_0) = \{\boldsymbol{x} \in \mathbb{R}^{n+1} \mid \|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^{n+1}} = \epsilon\}$ where $r \in \mathbb{R}_{>0}$ and $\boldsymbol{x}_0 \in \mathbb{R}^{n+1}$;
3. $\mathrm{bd}(R)$ where $R \subseteq \mathbb{R}^{n+1}$ is a fat compact rectangle.

Section 1.4

Differentiable multivariable functions

Unlike our discussion of continuity, the notion of differentiability for maps involving multiple variables is not so much a straightforward generalisation of the single-variable case. For example, we shall see that the appropriate way to think about the derivative in the multivariable case (and therefore, by specialisation, the single-variable case) is as a linear map. This turns out to be an important conceptual idea in understanding just what the derivative "is."

Some of the ideas in this section can be illustrated using single-variable examples, and we refer to Section I-3.2 for these. However, there are phenomenon in the multivariable case that do not arise in the single-variable case, and we give particular examples to exhibit these phenomenon.

Do I need to read this section? If you want to understand differentiability of multivariable functions, and you do not already, then you need to read this section. It is true that we do not make a great deal of use of the material in this section, but it does come up on occasion. •

local diffeos are open, mean value theorem using only differentiability and not C1, continuity of Jacobian derivative

1.4.1 Definition and basic properties of the derivative

The definition of what it means for a map to be differentiable immediately emphasises the linear algebraic character that is essential to the picture in higher-dimensions. The definition we give for the derivative in this case should be thought of as the generalisation of Proposition I-3.2.4; let us therefore present a result along these lines that will ensure that our definition of derivative makes sense.

1.4.1 **Proposition (Uniqueness of linear approximation)** *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *be an open set and let* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$ *be a map. For* $\mathbf{x}_0 \in \mathrm{U}$*, there exists at most one* $\mathsf{L} \in \mathrm{L}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{m}})$ *such that*

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0) - \mathsf{L}(\mathbf{x}-\mathbf{x}_0)\|_{\mathbb{R}^{\mathrm{m}}}}{\|\mathbf{x}-\mathbf{x}_0\|_{\mathbb{R}^{\mathrm{n}}}} = 0. \tag{1.15}$$

Proof Suppose there are two such maps L_1 and L_2. For any $x \in U$, we may write $x = x_0 + av$ for some $a \in \mathbb{R}_{>0}$ and $v \in \mathbb{R}^n$ such that $\|v\|_{\mathbb{R}^n} = 1$. We compute

$$\begin{aligned}\|\mathsf{L}_1(v) - \mathsf{L}_2(v)\|_{\mathbb{R}^m} &= \frac{\|\mathsf{L}_1(x-x_0) - \mathsf{L}_2(x-x_0)\|_{\mathbb{R}^m}}{\|x-x_0\|_{\mathbb{R}^n}} \\ &= \frac{\|-f(x) + f(x_0) + \mathsf{L}_1(x-x_0) + f(x) - f(x_0) - \mathsf{L}_2(x-x_0)\|_{\mathbb{R}^m}}{\|x-x_0\|_{\mathbb{R}^n}} \\ &\le \frac{\|f(x) - f(x_0) - \mathsf{L}_1(x-x_0)\|_{\mathbb{R}^m}}{\|x-x_0\|_{\mathbb{R}^n}} + \frac{\|f(x) - f(x_0) - \mathsf{L}_2(x-x_0)\|_{\mathbb{R}^m}}{\|x-x_0\|_{\mathbb{R}^n}}.\end{aligned}$$

Since L_1 and L_2 both satisfy (1.15), as we let $x \to x_0$ the right-hand side goes to zero showing that $\|\mathsf{L}_1(v) - \mathsf{L}_2(v)\|_{\mathbb{R}^m} = \|(\mathsf{L}_1 - \mathsf{L}_2)(v)\|_{\mathbb{R}^m} = 0$ for every v with $\|v\|_{\mathbb{R}^n} = 1$. Thus $\mathsf{L}_1 - \mathsf{L}_2$ is the trivial map sending any vector to zero, or equivalently $\mathsf{L}_1 = \mathsf{L}_2$. ■

We can now state the definition of the derivative for multivariable maps.

1.4.2 Definition (Derivative and differentiable map) Let $U \subseteq \mathbb{R}^n$ be an open subset and let $f: U \to \mathbb{R}^m$ be a map.

(i) The map f is ***differentiable at*** $\mathbf{x_0} \in \mathbf{U}$ if there exists a linear map $\mathsf{L}_{f,x_0}: \mathbb{R}^n \to \mathbb{R}^m$ such that
$$\lim_{x \to x_0} \frac{\|f(x) - f(x_0) - \mathsf{L}_{f,x_0}(x - x_0)\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}} = 0.$$

(ii) If f is differentiable at x_0, then the linear map L_{f,x_0} is denoted by $\boldsymbol{D}f(x_0)$ and is called the ***derivative*** of f at x_0.

(iii) If f is differentiable at each point $x \in U$, then f is ***differentiable***.

(iv) If f is differentiable and if the map $x \mapsto \boldsymbol{D}f(x)$ is continuous (using any norm one wishes on $\mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$) then f is ***continuously differentiable***, or of ***class*** $\mathbf{C^1}$. •

Sometimes the derivative is called the ***total derivative*** or the ***Fréchet derivative***. Similarly, differentiability in the sense of the preceding definition is sometimes called ***Fréchet differentiability***. The reason for this is that the existence of this derivative implies the existence of other derivatives, such as the directional derivative which we discuss in Section 1.4.3.

1.4.3 Notation (Evaluation of the derivative) Since $\boldsymbol{D}f(x_0) \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$, we can write $\boldsymbol{D}f(x_0)(v)$ as the image of $v \in \mathbb{R}^n$ under the derivative thought of as a linear map. To avoid the somewhat cumbersome looking double parentheses, we shall often write $\boldsymbol{D}f(x_0) \cdot v$ instead of $\boldsymbol{D}f(x_0)(v)$. •

With the derivative defined, it is now possible to talk about higher-order derivatives in a systematic way. We let $U \subseteq \mathbb{R}^n$ be open and let $f: U \to \mathbb{R}^m$ be continuously differentiable. The derivative is then a map $U \ni x \mapsto \boldsymbol{D}f(x) \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$. Given that from Section 1.1.3 we have a norm on $\mathrm{L}(\mathbb{R}^n; \mathbb{R}^m)$, this map is a candidate for having its derivative defined. The derivative of $\boldsymbol{D}f$ at $x_0 \in U$, if it exists, is the linear map $\boldsymbol{D}^2 f \in \mathrm{L}(\mathbb{R}^n; \mathrm{L}(\mathbb{R}^n; \mathbb{R}^m))(x_0)$ satisfying

$$\lim_{x \to x_0} \frac{\|\boldsymbol{D}f(x) - \boldsymbol{D}f(x_0) - \boldsymbol{D}^2 f(x_0) \cdot (x - x_0)\|_{\mathbb{R}^n, \mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}} = 0.$$

By Proposition I-5.6.7 we implicitly think of $\boldsymbol{D}^2 f$ as being an element of $\mathrm{L}^2(\mathbb{R}^n; \mathbb{R}^m)$. Now we can carry on this process recursively to define derivatives of arbitrary order.

1.4.4 Definition (Higher-order derivatives) Let $U \subseteq \mathbb{R}^n$ be open, let $f\colon U \to \mathbb{R}^m$ be a function, let $r \in \mathbb{Z}_{>0}$, and suppose that f is $(r-1)$ times differentiable with $G\colon U \to \mathrm{L}^{r-1}(\mathbb{R}^n; \mathbb{R}^m)$ denoting the $(r-1)$st derivative.

(i) The map f is ***r times continuously differentiable at $x_0 \in U$*** if there exists $DG(x_0) \in \mathrm{L}(\mathbb{R}^n; \mathrm{L}^{r-1}(\mathbb{R}^n; \mathbb{R}^m))$ such that

$$\lim_{x \to x_0} \frac{\|G(x) - G(x_0) - DG(x_0) \cdot (x - x_0)\|_{\mathbb{R}^n, \mathrm{L}^{r-1}(\mathbb{R}^n;\mathbb{R}^m)}}{\|x - x_0\|_{\mathbb{R}^n}} = 0. \tag{1.16}$$

(ii) If (1.16) holds then the map $DG(x_0)$ is identified, using Proposition I-5.6.7, with the multilinear map $D^r f(x_0) \in \mathrm{L}^r(\mathbb{R}^n; \mathbb{R}^m)$ and called the ***r*th derivative** of f at x_0.

(iii) If f is r times differentiable at each point $x \in U$, then f is ***r times differentiable***.

(iv) If f is r times differentiable and if the function $x \mapsto D^r f(x)$ is continuous, then f is ***r times continuously differentiable***, or of ***class C^r***.

If f is of class C^r for each $r \in \mathbb{Z}_{>0}$, then f is ***infinitely differentiable***, or of ***class C^∞***. •

The following result gives an important property of higher-order derivatives. Parts of the proof rely on properties of the derivative we have yet to prove. Specifically, the proof properly belongs after the proof of Theorem 1.4.33, but we give it here since this is where it fits best in terms of the flow of ideas.

1.4.5 Theorem (The derivative is symmetric) *If $\mathrm{U} \subseteq \mathbb{R}^n$ is open and if $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^m$ is of class C^r, then $\mathbf{D}^r\mathbf{f} \in \mathrm{S}^r(\mathbb{R}^n; \mathbb{R}^m)$.*

Proof By Proposition 1.4.17 we can assume, without loss of generality, that $m = 1$. We thus take $m = 1$ and write our function as f. When $r = 1$ we have $S^1(\mathbb{R}^n; \mathbb{R}) = \mathrm{L}(\mathbb{R}^n; \mathbb{R})$ so the result is vacuous in this case. We next consider the case when $r = 2$. Let $x_0 \in U$ and let $u, v \in \mathbb{R}^n$. Let $a \in \mathbb{R}_{>0}$ be sufficiently small that $x_0 + su + tv \in U$ for all $(s,t) \in \mathsf{B}^2(a, (0,0))$, this being possible since U is open and since the map $(s,t) \mapsto x_0 + su + tv$ is linear, and so infinitely differentiable by Corollary 1.4.9. Then define $g\colon \mathsf{B}^2(a,(0,0)) \to \mathbb{R}$ by

$$g(s,t) = f(x_0 + su + tv).$$

The Chain Rule (Theorem 1.4.49) implies that g is of class C^2. We then compute the following iterated partial derivatives using the Chain Rule and Proposition 1.4.7:

$$\begin{aligned} D_1 g(s,t) \cdot 1 &= Df(x_0 + su + tv) \cdot u, \\ D_2 g(s,t) \cdot 1 &= Df(x_0 + su + tv) \cdot v, \\ D_2 D_1 g(s,t) \cdot (1,1) &= D^2 f(x_0 + su + tv) \cdot (v, u), \\ D_1 D_2 g(s,t) \cdot (1,1) &= D^2 f(x_0 + su + tv) \cdot (u, v). \end{aligned}$$

Thus the result for $r = 2$ will follow if $D_1 D_2 g(0,0) = D_2 D_1 g(0,0)$. This, however, is a special case of Theorem 1.4.33.

For $r > 2$ we proceed by induction, assuming the result true for $r = s - 1$ and then supposing that f is of class C^r. For $\boldsymbol{x} \in U$ and $\boldsymbol{v}_1, \dots, \boldsymbol{v}_s \in \mathbb{R}^n$ we compute

$$\begin{aligned}\boldsymbol{D}^s f(\boldsymbol{x}) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s) &= (\boldsymbol{D}^2(\boldsymbol{D}^{s-2} f)(\boldsymbol{x}) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_2)) \cdot (\boldsymbol{v}_3, \dots, \boldsymbol{v}_s) \\ &= (\boldsymbol{D}^2(\boldsymbol{D}^{s-2} f)(\boldsymbol{x}) \cdot (\boldsymbol{v}_2, \boldsymbol{v}_1)) \cdot (\boldsymbol{v}_3, \dots, \boldsymbol{v}_s) \\ &= \boldsymbol{D}^s f(\boldsymbol{x}) \cdot (\boldsymbol{v}_2, \boldsymbol{v}_1, \dots, \boldsymbol{v}_s),\end{aligned}$$

showing that

$$\boldsymbol{D}^s f(\boldsymbol{x}) \cdot (\boldsymbol{v}_{\sigma(1)}, \boldsymbol{v}_{\sigma(2)}, \dots, \boldsymbol{v}_{\sigma(s)}) = \boldsymbol{D}^s f(\boldsymbol{x}) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s)$$

for $\sigma = (1\ 2)$. Now let $\sigma \in \mathfrak{S}_{s-1}$ and by the induction hypothesis note that

$$\boldsymbol{D}^{s-1} f(\boldsymbol{x}) \cdot (\boldsymbol{v}_{\sigma(1)}, \dots, \boldsymbol{v}_{\sigma(s-1)}) = \boldsymbol{D}^{s-1} f(\boldsymbol{x}) \cdot (\boldsymbol{v}_1, \dots, \boldsymbol{v}_{s-1})$$

for all $\boldsymbol{x} \in U$ and $\boldsymbol{v}_1, \dots, \boldsymbol{v}_{s-1}$. Then, by Proposition 1.4.7, we have, for any $\boldsymbol{v}_0 \in \mathbb{R}^n$,

$$\begin{aligned}\boldsymbol{D}^s f(\boldsymbol{x}) \cdot (\boldsymbol{v}_0, \boldsymbol{v}_{\sigma(1)}, \dots, \boldsymbol{v}_{\sigma(s-1)}) &= (\boldsymbol{D}(\boldsymbol{D}^{s-1} f)(\boldsymbol{x}) \cdot \boldsymbol{v}_0) \cdot (\boldsymbol{v}_{\sigma(1)}, \dots, \boldsymbol{v}_{\sigma(s-1)}) \\ &= (\boldsymbol{D}(\boldsymbol{D}^{s-1} f)(\boldsymbol{x}) \cdot \boldsymbol{v}_0) \cdot (\boldsymbol{v}_1, \dots, \boldsymbol{v}_{s-1}) \\ &= \boldsymbol{D}^s f(\boldsymbol{x}) \cdot (\boldsymbol{v}_0, \boldsymbol{v}_1, \dots, \boldsymbol{v}_{s-1}),\end{aligned}$$

giving

$$\boldsymbol{D}^s f(\boldsymbol{x}) \cdot (\boldsymbol{v}_{\sigma(1)}, \boldsymbol{v}_{\sigma(2)}, \dots, \boldsymbol{v}_{\sigma(s)}) = \boldsymbol{D}^s f(\boldsymbol{x}) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_2, \dots, \boldsymbol{v}_s)$$

when σ leaves 1 fixed. Now, by Exercise I-4.1.12 any permutation $\sigma \in \mathfrak{S}_s$ can be written as a finite product of (1 2) and permutations leaving 1 fixed. From this the result follows. ■

We now deal with the problem of having potentially competing definitions of the derivative for a $\mathbb{R}$-valued function of a single real variable. Let us resolve this.

1.4.6 Theorem (Consistency of differentiability definitions for $\mathbb{R}$-valued functions of a single variable) *Let $\mathrm{I} \subseteq \mathbb{R}$ be an open interval, let $\mathrm{f}\colon \mathrm{I} \to \mathbb{R}$, let $\mathrm{x}_0 \in \mathrm{I}$, and let $\mathrm{r} \in \mathbb{Z}_{\geq 0}$. Then f is r times differentiable at x_0 in the sense of Definition I-3.2.5 if and only if f is r times differentiable at x_0 in the sense of Definition 1.4.4. Moreover, if f is r times continuously differentiable at x_0 then*

$$\mathbf{D}^{\mathrm{r}}\mathrm{f}(\mathrm{x}_0)(\mathrm{v}_1, \dots, \mathrm{v}_{\mathrm{r}}) = \mathrm{f}^{(\mathrm{r})}(\mathrm{x}_0)\mathrm{v}_1 \cdots \mathrm{v}_{\mathrm{r}}$$

for every $\mathrm{v}_1, \dots, \mathrm{v}_{\mathrm{r}} \in \mathbb{R}$.

Proof We first observe that there is a natural isomorphism from $\mathbb{R}$ to $S^r(\mathbb{R}; \mathbb{R})$ assigning to $a \in \mathbb{R}$ the symmetric multilinear map

$$(v_1, \dots, v_r) \mapsto a\, v_1 \cdots v_r.$$

This isomorphism is easily verified to preserve the standard norms on $\mathbb{R}$ and $S^r(\mathbb{R}; \mathbb{R})$. We shall implicitly use this isomorphism is the proof.

For $r = 0$ the result is clearly true since 0 times differentiable means continuous in the case of each definition. Assume the result is true for $r \in \{0, 1, \dots, k-1\}$. Thus assume that existence of $\boldsymbol{D}^{k-1}f(x_0)$ is equivalent to existence of $f^{(k-1)}(x_0)$ and that

$$\boldsymbol{D}^{k-1}f(x_0)(v_1, \dots, v_{k-1}) = f^{(k-1)}(x_0)v_1 \cdots v_{k-1}$$

for all $v_1, \dots, v_{k-1} \in \mathbb{R}$.

First let us suppose that f is k times differentiable at x_0 in the sense of Definition 1.4.4. Then $\boldsymbol{D}^{k-1}f$ is continuous at x_0. Let $g\colon I \to \mathbb{R}$ be defined by asking that $g(x)$ be the image of $\boldsymbol{D}^{k-1}f(x)$ under the isomorphism of $S^{k-1}(\mathbb{R};\mathbb{R})$ with $\mathbb{R}$. It then holds that g is differentiable at x_0 in the sense of Definition 1.4.4 since $\boldsymbol{D}^{k-1}f$ is differentiable at x_0 in the sense of Definition 1.4.4. By the induction hypothesis it then follows from Proposition I-3.2.4 that $f^{(k-1)}$ is differentiable in the sense of Definition I-3.2.5. This means that f is k times differentiable at x_0 in the sense of Definition I-3.2.5.

Next suppose that f is k times differentiable at x_0 in the sense of Definition 1.4.4. Let $\mathsf{L}\colon I \to S^{k-1}(\mathbb{R};\mathbb{R})$ be defined by asking that $\mathsf{L}(x)$ be the image of $f^{(k-1)}(x)$ under isomorphism of $\mathbb{R}$ with $S^{k-1}(\mathbb{R};\mathbb{R})$. Since $f^{(k-1)}$ is differentiable at x_0 in the sense of Definition I-3.2.5 it follows that L is differentiable at x_0 in the sense of Definition I-3.2.5. By the induction hypothesis and Proposition I-3.2.4 it follows that $\boldsymbol{D}^{k-1}f$ is differentiable at x_0 in the sense of Definition 1.4.4. This means that f is k times differentiable at x_0 in the sense of Definition 1.4.4.

For the final assertion of the proof, for fixed $v_1, \dots, v_{k-1} \in \mathbb{R}$ consider the function $h\colon I \to \mathbb{R}$ defined by

$$h(x) = f^{(k-1)}(x)v_1 \cdots v_{k-1}.$$

We claim that h is differentiable at x_0 if f is k times differentiable at x_0. We use the derivative of Definition I-3.2.5 to verify this assertion. We have

$$\begin{aligned}\lim_{x \to x_0} \frac{h(x) - h(x_0)}{x - x_0} &= \lim_{x \to x_0} \frac{f^{(k-1)}(x)v_1 \cdots v_{k-1} - f^{(k-1)}(x_0)v_1 \cdots v_{k-1}}{x - x_0} \\ &= \lim_{x \to x_0} \frac{f^{(k-1)}(x) - f^{(k-1)}(x_0)}{x - x_0} v_1 \cdots v_{k-1} \\ &= f^{(k)}(x_0)v_1 \cdots v_{k-1},\end{aligned}$$

where we have used Proposition I-2.3.23 and Proposition I-2.3.29. This gives the differentiability of h at x_0 as well as an explicit formula for the derivative. Using Proposition I-3.2.4 we have

$$\boldsymbol{D}h(x_0) \cdot v_0 = f^{(k)}(x_0)v_0 v_1 \cdots v_{k-1},$$

which gives the theorem. ■

The reader will have noticed that we give no examples to illustrate the multidimensional derivative. There is a reason for this. Based on the definition it is not that easy to actually compute the derivative in multiple-dimensions. However, it is actually easy to compute this derivative in practice only knowing how to differentiate $\mathbb{R}$-valued functions of a single variable. But the development of this connection is actually a little involved, and we postpone it until Theorem 1.4.22, at which time we will also provide some examples.

We close this section with a useful characterisation of differentiability that can simplify how one handles computations with derivatives.

1.4.7 Proposition (Swapping of differentiation and evaluation) *For* $U \subseteq \mathbb{R}^n$ *open, for* $\mathbf{f}\colon U \to \mathbb{R}^m$*, and for* $\mathbf{x}_0 \in U$*, the following statements are equivalent:*

(i) $\mathbf{f}$ *is* r *times differentiable at* $\mathbf{x}_0$*;*

(ii) $\mathbf{f}$ *is* $r-1$ *times continuously differentiable in a neighbourhood of* $\mathbf{x}_0$ *and, for each* $\mathbf{v}_1, \dots, \mathbf{v}_{r-1} \in \mathbb{R}^n$*, the map* $\boldsymbol{\delta}_{\mathbf{f};\mathbf{v}_1,\dots,\mathbf{v}_{r-1}}\colon U \to \mathbb{R}^m$ *defined by*

$$\boldsymbol{\delta}_{\mathbf{f};\mathbf{v}_1,\dots,\mathbf{v}_{r-1}}(\mathbf{x}) = \mathbf{D}^{r-1}\mathbf{f}(\mathbf{x}) \cdot (\mathbf{v}_1, \dots, \mathbf{v}_{r-1})$$

is differentiable at $\mathbf{x}_0$*.*

Moreover, if $\mathbf{f}$ *is* r *times differentiable at* $\mathbf{x}_0 \in U$ *then*

$$\mathbf{D}^r\mathbf{f}(\mathbf{x}_0) \cdot (\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{r-1}) = \mathbf{D}\boldsymbol{\delta}_{\mathbf{f};\mathbf{v}_1,\dots,\mathbf{v}_{r-1}}(\mathbf{x}_0) \cdot \mathbf{v}_0 \tag{1.17}$$

for every $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{r-1} \in \mathbb{R}^n$*.*

Proof First suppose that f is r times differentiable at x_0. From Proposition 1.4.35 it follows that f is $r-1$ times continuously differentiable in a neighbourhood of x_0. For $v_1, \dots, v_{r-1} \in \mathbb{R}^n$ let us define $\mathrm{Ev}_{v_1,\dots,v_{r-1}}\colon \mathrm{L}^{r-1}(\mathbb{R}^n;\mathbb{R}^m;\to)\mathbb{R}^m$ by

$$\mathrm{Ev}_{v_1,\dots,v_{r-1}}(\mathsf{L}) = \mathsf{L}(v_1, \dots, v_{r-1}).$$

Then we have $\boldsymbol{\delta}_{f;v_1,\dots,v_{r-1}} = \mathrm{Ev}_{v_1,\dots,v_{r-1}} \circ \boldsymbol{D}^{r-1}f$. Since $\mathrm{Ev}_{v_1,\dots,v_{r-1}}$ is linear (this is a simple verification), it follows from Corollary 1.4.9 that it is infinitely differentiable. Thus $\boldsymbol{\delta}_{f;v_1,\dots,v_{r-1}}$ is differentiable by the Chain Rule, Theorem 1.4.49. Moreover, also by the Chain Rule and Corollary 1.4.9, it follows that

$$\begin{aligned} \boldsymbol{D}\boldsymbol{\delta}_{f;v_1,\dots,v_{r-1}}(x_0) \cdot v_0 &= \mathrm{Ev}_{v_1,\dots,v_{r-1}}(\boldsymbol{D}(\boldsymbol{D}^{r-1}f)(x_0) \cdot v_0) \\ &= \boldsymbol{D}^r f(x_0) \cdot (v_0, v_1, \dots, v_{r-1}), \end{aligned}$$

using Proposition I-5.6.7. This gives (1.17).

Next suppose that f is $r-1$ times continuously differentiable in a neighbourhood of x_0 and that $\boldsymbol{\delta}_{f;v_1,\dots,v_{r-1}}$ is differentiable at x_0 for every $v_1, \dots, v_{r-1} \in \mathbb{R}^n$. To show that f is r times differentiable at x_0 we claim that it suffices to show that the components of $\boldsymbol{D}^{r-1}f$ are differentiable at x_0 (see Definition I-5.6.8 for definition of the components of a multilinear map). That this is so essentially follows from Proposition 1.4.17 below. However, the "essentially" warrants a little explanation.

In Proposition 1.4.17 we show that a map taking values in $\mathbb{R}^m$ is differentiable if and only if each of its components is differentiable. But here we are not talking about a map taking values in $\mathbb{R}^m$, but taking values in $\mathrm{L}^{r-1}(\mathbb{R}^n;\mathbb{R}^m)$. But, the assignment taking a multilinear map in $\mathrm{L}^{r-1}(\mathbb{R}^n;\mathbb{R}^m)$ to its components is a linear isomorphism taking values in $\mathbb{R}^{mn^{r-1}}$. Moreover, the Frobenius norm on $\mathrm{L}^{r-1}(\mathbb{R}^n;\mathbb{R}^m)$ is "the same
what as" the Euclidean norm on $\mathbb{R}^{mn^{r-1}}$ under this isomorphism; in the language of , the

isomorphism is norm-preserving. Therefore, Proposition 1.4.17 can essentially be applied to assert that $\boldsymbol{D}^{r-1}\boldsymbol{f}$ is differentiable at $\boldsymbol{x}_0$ if its components are differentiable at $\boldsymbol{x}_0$.

The matter of showing that the components of $\boldsymbol{D}^{r-1}\boldsymbol{f}$ are differentiable at $\boldsymbol{x}_0$ is straightforward. Indeed, the components of $\boldsymbol{D}^{r-1}\boldsymbol{f}$ are simply given by the $\mathbb{R}$-valued functions

$$\boldsymbol{x} \mapsto (\boldsymbol{D}^{r-1}\boldsymbol{f}(\boldsymbol{x}) \cdot (\boldsymbol{e}_{j_1}, \dots, \boldsymbol{e}_{j_{r-1}}))_a = (\delta_{\boldsymbol{f};\boldsymbol{e}_{j_1},\dots,\boldsymbol{e}_{j_{r-1}}}(\boldsymbol{x}))_a,$$
$$j_1, \dots, j_{r-1} \in \{1, \dots, n\},\ a \in \{1, \dots, m\},$$

defined in a neighbourhood of $\boldsymbol{x}_0$. By assumption and by Proposition 1.4.17 these functions are, indeed, differentiable at $\boldsymbol{x}_0$. ■

While in these volumes we do not adhere to presentation dictated solely by logical implications always flowing forwards, we do feel compelled to warn the reader that in this section we make an abuse of logical ordering so dire as to merit comment. We shall in the next several sections (and already in the proofs of Theorem 1.4.5 and Proposition 1.4.7 above) make repeated and crucial use of the multivariable Chain Rule which we do not prove until Theorem 1.4.49. A reader who might be bothered by this can go ahead and read the Chain Rule and its proof right now since the proof relies only on ideas that are presently at our disposal.

1.4.2 Derivatives of multilinear maps

In this section we consider a special class of maps, and show that they are infinitely differentiable and compute their derivatives of all orders. The maps we consider are multilinear maps $\mathsf{L}\colon \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k} \to \mathbb{R}^m$. It will turn out that these maps come up many times for various reasons, and for this reason it is useful to determine their derivatives. Moreover, it is a good exercise in using the definition of the derivatives to compute the derivatives of multilinear maps.

Since derivatives are themselves multilinear maps, it will be useful to discriminate notationally between points in the domain of the map and points in the domain of the derivative of the map. Thus we shall write a point in $\mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ as $(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$ when we mean it to be in the domain of the map L and we shall write a point in $\mathbb{R}^{n_1} \oplus \cdots \oplus \mathbb{R}^{n_k}$ as $(\boldsymbol{v}_1, \dots, \boldsymbol{v}_k)$ when we mean it to be an argument of the derivative. The argument of the rth derivative is an element of $(\mathbb{R}^{n_1} \oplus \cdots \oplus \mathbb{R}^{n_k})^r$ and will be written as

$$((\boldsymbol{v}_{11}, \dots, \boldsymbol{v}_{1k}), \dots, (\boldsymbol{v}_{r1}, \dots, \boldsymbol{v}_{rk})).$$

For $r \in \{1, \dots, k\}$ define

$$D_{r,k} = \{\{j_1, \dots, j_r\} \mid\ j_1, \dots, j_r \in \{1, \dots, k\} \text{ distinct}\}.$$

For $\{j_1, \dots, j_r\} \in D_{r,k}$ let us denote by $\{j'_1, \dots, j'_{k-r}\}$ the complement of $\{j_1, \dots, j_r\}$ in $\{1, \dots, k\}$. Now, for $\{j_1, \dots, j_r\} \in D_{r,k}$ define

$$\boldsymbol{\lambda}_{j_1,\dots,j_r} \in \mathsf{L}((\mathbb{R}^{n_{j'_1}} \oplus \cdots \oplus \mathbb{R}^{n_{j'_{k-r}}}) \oplus (\mathbb{R}^{n_{j_1}} \oplus \cdots \oplus \mathbb{R}^{n_{j_r}}); \mathbb{R}^{n_1} \oplus \cdots \oplus \mathbb{R}^{n_k})$$

by asking that

$$\boldsymbol{\lambda}_{j_1,\dots,j_k}((\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-r}),(\boldsymbol{v}_1,\dots,\boldsymbol{v}_r))$$

be obtained by placing $\boldsymbol{x}_l$ in slot j'_l for $l \in \{1,\dots,k-r\}$ and by placing $\boldsymbol{v}_l$ in slot j_l for $l \in \{1,\dots,r\}$.

With the above notation we have the following description of the derivative of a multilinear map.

1.4.8 Theorem (Derivatives of multilinear maps) *If* $\mathsf{L} \in \mathrm{L}(\mathbb{R}^{n_1} \oplus \cdots \oplus \mathbb{R}^{n_k}; \mathbb{R}^m)$ *then* L *is infinitely differentiable. Moreover, for* $r \in \{1,\dots,k\}$ *we have*

$$\begin{aligned}\mathbf{D}^r\mathsf{L}(\mathbf{x}_1,\dots,\mathbf{x}_k)\cdot((\mathbf{v}_{11},\dots,\mathbf{v}_{1k}),\dots,(\mathbf{v}_{r1},\dots,\mathbf{v}_{rk}))\\ = \sum_{\sigma\in\mathfrak{S}_r}\sum_{\{j_1,\dots,j_r\}\in D_{r,k}} \mathsf{L}\circ\boldsymbol{\lambda}_{j_1,\dots,j_r}((\mathbf{x}_{j'_1},\dots,\mathbf{x}_{j'_{k-r}}),(\mathbf{v}_{\sigma(1)j_1},\dots,\mathbf{v}_{\sigma(r)j_r}))\end{aligned}$$

and for $r > k$ *we have* $\mathbf{D}^r\mathsf{L}(\mathbf{x}_1,\dots,\mathbf{x}_k) = \mathbf{0}$.

Proof We prove the result by induction on r. For $r = 1$ the theorem asserts that

$$\begin{aligned}D\mathsf{L}(\boldsymbol{x}_{01},\dots,\boldsymbol{x}_{0k})\cdot(\boldsymbol{v}_1,\dots,\boldsymbol{v}_k) = \mathsf{L}(\boldsymbol{v}_1,\boldsymbol{x}_{02},\dots,\boldsymbol{x}_{0k})\\ + \mathsf{L}(\boldsymbol{x}_{01},\boldsymbol{v}_2,\dots,\boldsymbol{x}_{0k}) + \cdots + \mathsf{L}(\boldsymbol{x}_{01},\boldsymbol{x}_{02},\dots,\boldsymbol{v}_k).\end{aligned}$$

To verify this we must show that

$$\begin{aligned}\lim_{\substack{(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\\ \to(\boldsymbol{x}_{01},\dots,\boldsymbol{x}_{0k})}} \big\|\mathsf{L}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k) - \mathsf{L}(\boldsymbol{x}_{01},\dots,\boldsymbol{x}_{0k}) - \mathsf{L}(\boldsymbol{x}_1-\boldsymbol{x}_{01},\dots,\boldsymbol{x}_{0k})\\ - \mathsf{L}(\boldsymbol{x}_{01},\dots,\boldsymbol{x}_k-\boldsymbol{x}_{0k})\big\|_{\mathbb{R}^m}/\|(\boldsymbol{x}_1-\boldsymbol{x}_{01},\dots,\boldsymbol{x}_k-\boldsymbol{x}_{0k})\|_{\mathbb{R}^{n_1+\cdots+n_k}} = 0. \quad (1.18)\end{aligned}$$

We do this by induction on k. For $k = 1$ we have

$$\mathsf{L}(\boldsymbol{x}_1) - \mathsf{L}(\boldsymbol{x}_{01}) - \mathsf{L}(\boldsymbol{x}_1 - \boldsymbol{x}_{01}) = \mathbf{0},$$

and so (1.18) holds trivially. Now suppose that (1.18) holds for $k = s \geq 2$ and let $\mathsf{L} \in \mathrm{L}(\mathbb{R}^{n_1},\dots,\mathbb{R}^{n_{s+1}};\mathbb{R}^m)$. We first note that the numerator in the limit in (1.18) can be written as

$$\begin{aligned}\mathsf{L}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_s,\boldsymbol{x}_{0(s+1)}) - \mathsf{L}(\boldsymbol{x}_{01},\dots,\boldsymbol{x}_{0s},\boldsymbol{x}_{0(s+1)}) + \mathsf{L}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_s,\boldsymbol{x}_s-\boldsymbol{x}_{0(s+1)})\\ - \mathsf{L}(\boldsymbol{x}_1-\boldsymbol{x}_{01},\dots,\boldsymbol{x}_{0s},\boldsymbol{x}_{0(s+1)}) - \cdots - \mathsf{L}(\boldsymbol{x}_{01},\dots,\boldsymbol{x}_s-\boldsymbol{x}_{0s},\boldsymbol{x}_{0(s+1)})\\ - \mathsf{L}(\boldsymbol{x}_{01},\dots,\boldsymbol{x}_{0s},\boldsymbol{x}_{s+1}-\boldsymbol{x}_{0(s+1)}).\end{aligned}$$

By the induction hypothesis we have

$$\begin{aligned}\lim_{\substack{(\boldsymbol{x}_1,\dots,\boldsymbol{x}_s)\\ \to(\boldsymbol{x}_{01},\dots,\boldsymbol{x}_{0s})}} \big\|\mathsf{L}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_s,\boldsymbol{x}_{0(s+1)}) - \mathsf{L}(\boldsymbol{x}_{01},\dots,\boldsymbol{x}_{0s},\boldsymbol{x}_{0(s+1)})\\ - \mathsf{L}(\boldsymbol{x}_1-\boldsymbol{x}_{01},\dots,\boldsymbol{x}_{0s},\boldsymbol{x}_{0(s+1)}) - \mathsf{L}(\boldsymbol{x}_{01},\dots,\boldsymbol{x}_s-\boldsymbol{x}_{0s},\boldsymbol{x}_{0(s+1)})\big\|_{\mathbb{R}^m}\\ /\|(\boldsymbol{x}_1-\boldsymbol{x}_{01},\dots,\boldsymbol{x}_s-\boldsymbol{x}_{0s})\|_{\mathbb{R}^{n_1+\cdots+n_s}} = 0.\end{aligned}$$

Since

$$\|(x_1 - x_{01}, \dots, x_s - x_{0s})\|_{\mathbb{R}^{n_1+\dots+n_s}} \le \|(x_1 - x_{01}, \dots, x_s - x_{0s}, x_{s+1} - x_{0(s+1)})\|_{\mathbb{R}^{n_1+\dots+n_s+n_{s+1}}}$$

this implies that

$$\lim_{\substack{(x_1,\dots,x_s,x_{s+1})\\ \to(x_{01},\dots,x_{0s},x_{0(s+1)})}} \Big\|\mathsf{L}(x_1,\dots,x_s,x_{0(s+1)}) - \mathsf{L}(x_{01},\dots,x_{0s},x_{0(s+1)}) \\ - \mathsf{L}(x_1 - x_{01},\dots,x_{0s},x_{0(s+1)}) - \mathsf{L}(x_{01},\dots,x_s - x_{0s},x_{0(s+1)})\Big\|_{\mathbb{R}^m} \\ /\|(x_1 - x_{01},\dots,x_s - x_{0s},x_{s+1} - x_{0(s+1)})\|_{\mathbb{R}^{n_1+\dots+n_s+n_{s+1}}} = 0. \quad (1.19)$$

We also have

$$\lim_{\substack{(x_1,\dots,x_s,x_{s+1})\\ \to(x_{01},\dots,x_{0s},x_{0(s+1)})}} \Big\|\mathsf{L}\Big(x_1,\dots,x_s,\tfrac{x_{s+1}-x_{0(s+1)}}{\|x_{s+1}-x_{0(s+1)}\|_{\mathbb{R}^{n_{s+1}}}}\Big) - \mathsf{L}\Big(x_{01},\dots,x_{0s},\tfrac{x_{s+1}-x_{0(s+1)}}{\|x_{s+1}-x_{0(s+1)}\|_{\mathbb{R}^{n_{s+1}}}}\Big)\Big\|_{\mathbb{R}^m} = 0$$

by continuity of L. Since

$$\|x_{s+1} - x_{0(s+1)}\|_{\mathbb{R}^{n_{s+1}}} \le \|(x_1 - x_{01},\dots,x_s - x_{0s},x_{s+1} - x_{0(s+1)})\|_{\mathbb{R}^{n_1+\dots+n_s+n_{s+1}}}$$

this gives

$$\lim_{\substack{(x_1,\dots,x_s,x_{s+1})\\ \to(x_{01},\dots,x_{0s},x_{0(s+1)})}} \Big\|\mathsf{L}(x_1,\dots,x_s,x_{s+1} - x_{0(s+1)}) - \mathsf{L}(x_{01},\dots,x_{0s},x_{s+1} - x_{0(s+1)})\Big\|_{\mathbb{R}^m} \\ /\|(x_1 - x_{01},\dots,x_s - x_{0s},x_{s+1} - x_{0(s+1)})\|_{\mathbb{R}^{n_1+\dots+n_s+n_{s+1}}} = 0. \quad (1.20)$$

Combining (1.19) and (1.20) gives (1.18) for the case when $k = s+1$ and so gives the conclusion of the theorem in the case when $r = 1$.

Now suppose that the theorem holds for $r \in \{1,\dots,s\}$ with $s < k$ and let $\mathsf{L} \in \mathrm{L}(\mathbb{R}^{n_1},\dots,\mathbb{R}^{n_k};\mathbb{R}^m)$. Let us fix $\{j_1,\dots,j_s\} \in D_{s,k}$ and denote the complement of $\{j_1,\dots,j_s\}$ in $\{1,\dots,k\}$ by $\{j'_1,\dots,j'_{k-s}\}$, just as in our definitions before the theorem statement. Let us also fix $v_{j_l} \in \mathbb{R}^{n_{j_l}}$ for $l \in \{1,\dots,s\}$. Then define

$$\mathsf{P}_{v_{j_1},\dots,v_{j_s}} : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_k} \to (\mathbb{R}^{n_{j'_1}} \times \dots \times \mathbb{R}^{n'_{j_{k-s}}}) \times (\mathbb{R}^{n_{j_1}} \times \dots \times \mathbb{R}^{n_{j_s}}) \\ (x_1,\dots,x_k) \mapsto ((x_{j'_1},\dots,x_{j'_{k-s}}),(v_{j_1},\dots,v_{j_s})).$$

Now define $g_{v_{j_1},\dots,v_{j_s}} : \mathbb{R}^{n_1} \times \dots \times \mathbb{R}^{n_k} \to \mathbb{R}^m$ by $g_{v_{j_1},\dots,v_{j_s}} = \mathsf{L} \circ \lambda_{j_1,\dots,j_s} \circ \mathsf{P}_{v_{j_1},\dots,v_{j_s}}$ and note that

$$g_{v_{j_1},\dots,v_{j_s}}(x_1,\dots,x_k) = \mathsf{L} \circ \lambda_{j_1,\dots,j_s}((x_{j'_1},\dots,x_{j'_{k-s}}),(v_{j_1},\dots,v_{j_s})).$$

By the Chain Rule, Theorem 1.4.49 below, we have

$$\boldsymbol{D}g_{v_{j_1},\dots,v_{j_s}}(x_1,\dots,x_k) \cdot (u_1,\dots,u_k) \\ = \boldsymbol{D}(\mathsf{L} \circ \lambda_{j_1,\dots,j_r})(\boldsymbol{P}(x_1,\dots,x_k)) \circ \boldsymbol{D}\mathsf{P}_{v_{j_1},\dots,v_{j_s}}(x_1,\dots,x_k) \cdot (u_1,\dots,u_k).$$

Note that since $\mathsf{P}_{v_{j_1},\dots,v_{j_s}}$ is essentially a linear map (precisely, it is affine, meaning linear plus constant) we have

$$\boldsymbol{D}\mathsf{P}_{v_{j_1},\dots,v_{j_s}}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\cdot(\boldsymbol{u}_1,\dots,\boldsymbol{u}_k) = ((\boldsymbol{u}_{j'_1},\dots,\boldsymbol{u}_{j'_{k-s}}),(\boldsymbol{0},\dots,\boldsymbol{0})).$$

Note that since $\mathsf{L}\circ\boldsymbol{\lambda}_{j_1,\dots,j_s}\in\mathrm{L}(\mathbb{R}^{n_{j'_1}},\dots,\mathbb{R}^{n_{j'_{k-s}}},\mathbb{R}^{n_{j_1}},\dots,\mathbb{R}^{n_{j_s}};\mathbb{R}^m)$ (as is readily verified), by the induction hypothesis,

$$\begin{aligned}\boldsymbol{D}(\mathsf{L}\circ\boldsymbol{\lambda}_{j_1,\dots,j_s})(\boldsymbol{x}_{j'_1},\dots,\boldsymbol{x}_{j'_{k-s}},\boldsymbol{x}_{j_1},\dots,\boldsymbol{x}_{j_s})\cdot((\boldsymbol{u}_{j'_1},\dots,\boldsymbol{u}_{j'_{k-s}}),(\boldsymbol{u}_{j_1},\dots,\boldsymbol{u}_{j_s}))\\ = \mathsf{L}\circ\boldsymbol{\lambda}_{j_1,\dots,j_s}((\boldsymbol{u}_{j'_1},\dots,\boldsymbol{x}_{j'_{k-s}}),(\boldsymbol{x}_{j_1},\dots,\boldsymbol{x}_{j_s}))+\dots\\ +\mathsf{L}\circ\boldsymbol{\lambda}_{j_1,\dots,j_s}((\boldsymbol{x}_{j'_1},\dots,\boldsymbol{x}_{j'_{k-s}}),(\boldsymbol{x}_{j_1},\dots,\boldsymbol{u}_{j_s})).\end{aligned}$$

Therefore,

$$\begin{aligned}\boldsymbol{D}\boldsymbol{g}_{v_{j_1},\dots,v_{j_s}}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\cdot(\boldsymbol{u}_1,\dots,\boldsymbol{u}_k)\\ = \mathsf{L}\circ\boldsymbol{\lambda}_{j_1,\dots,j_s}((\boldsymbol{u}_{j'_1},\dots,\boldsymbol{x}_{j'_{k-s}}),(\boldsymbol{v}_{j_1},\dots,\boldsymbol{v}_{j_s}))+\dots\\ +\mathsf{L}\circ\boldsymbol{\lambda}_{j_1,\dots,j_s}((\boldsymbol{x}_{j'_1},\dots,\boldsymbol{u}_{j'_{k-s}}),(\boldsymbol{v}_{j_1},\dots,\boldsymbol{v}_{j_s})).\end{aligned}$$

Thus, for $\boldsymbol{v}_j\in\mathbb{R}^{n_j}$, $j\in\{1,\dots,k\}$, we have

$$\begin{aligned}\boldsymbol{D}\boldsymbol{g}_{v_{j_1},\dots,v_{j_s}}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\cdot(\boldsymbol{v}_1,\dots,\boldsymbol{v}_k)\\ = \sum_{j_{s+1}\notin\{j_1,\dots,j_s\}}\mathsf{L}\circ\boldsymbol{\lambda}_{j_1,\dots,j_s,j_{s+1}}((\boldsymbol{x}_{j'_1},\dots,\boldsymbol{x}_{j'_{k-(s+1)}}),(\boldsymbol{v}_{j_1},\dots,\boldsymbol{v}_{j_{s+1}})).\end{aligned}$$

Thus, using this relation along with Proposition 1.4.7, linearity of the derivative (see Proposition 1.4.47), the Chain Rule (see Theorem 1.4.49), and the induction hypothesis, we compute

$$\begin{aligned}\boldsymbol{D}^{s+1}\boldsymbol{f}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\cdot((\boldsymbol{v}_{11},\dots,\boldsymbol{v}_{1k}),(\boldsymbol{v}_{21},\dots,\boldsymbol{v}_{2k}),\dots,\dots,(\boldsymbol{v}_{(s+1)1},\dots,\boldsymbol{v}_{(s+1)k}))\\ = \sum_{\sigma\in\mathfrak{S}_s}\sum_{\{j_2,\dots,j_{s+1}\}\in D_{s,k}}\boldsymbol{D}\boldsymbol{g}_{v_{\sigma(2)j_2},\dots,v_{\sigma(s+1)j_{s+1}}}(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\cdot(\boldsymbol{v}_{11},\dots,\boldsymbol{v}_{1k})\\ = \sum_{\sigma\in\mathfrak{S}_s}\sum_{\{j_2,\dots,j_{s+1}\}\in D_{s,k}}\sum_{j_1\notin\{j_2,\dots,j_{s+1}\}}\mathsf{L}\circ\boldsymbol{\lambda}_{j_1,\dots,j_s,j_{s+1}}((\boldsymbol{x}_{j'_1},\dots,\boldsymbol{x}_{j'_{k-(s+1)}}),\\ (\boldsymbol{v}_{j_1},\boldsymbol{v}_{\sigma(2)j_2},\dots,\boldsymbol{v}_{\sigma(s+1)j_{s+1}}))\\ = \sum_{\sigma\in\mathfrak{S}_{s+1}}\sum_{\{j_1,\dots,j_{s+1}\}\in D_{s+1,k}}\mathsf{L}\circ\boldsymbol{\lambda}_{\{j_1,\dots,j_{s+1}\}}((\boldsymbol{x}_{j'_1},\dots,\boldsymbol{x}_{j'_{k-(s+1)}}),\\ (\boldsymbol{v}_{\sigma(1)j_1},\dots,\boldsymbol{v}_{\sigma(s+1)j_{s+1}})),\end{aligned}$$

where, in the second and third line, we define $\sigma\in\mathfrak{S}_s$ to be a bijection of $\{1,\dots,s+1\}$ by permutation of the last s elements.

The preceding argument gives the result when $r\in\{1,\dots,k\}$. For $r>k$ we argue as follows. We first note that

$$\boldsymbol{D}^k\mathsf{L}((\boldsymbol{v}_{11},\dots,\boldsymbol{v}_{1k}),\dots,(\boldsymbol{v}_{k1},\dots,\boldsymbol{v}_{kk})) = \sum_{\sigma\in\mathfrak{S}_k}\mathsf{L}(\boldsymbol{v}_{\sigma(1)1},\dots,\boldsymbol{v}_{\sigma(k)k}). \tag{1.21}$$

By Proposition 1.4.7 it follows that $\boldsymbol{D}^r\mathsf{L}=\boldsymbol{0}$ for $r>k$. ■

The proof of the preceding theorem, and indeed the statement, is marred by notational baggage needed to state the result in full generality. However, the result is actually simple to use, and to illustrate this we explicitly write the result when $k = 3$. In this case we have the following formulae:

$$\begin{aligned}
\boldsymbol{D}\mathsf{L}(x_1,x_2,x_3)\cdot(v_{11},v_{12},v_{13}) &= \mathsf{L}(v_1,x_2,x_3)+\mathsf{L}(x_1,v_2,x_3)+\mathsf{L}(x_1,x_2,v_3),\\
\boldsymbol{D}^2\mathsf{L}(x_1,x_2,x_3)\cdot((v_{11},v_{12},v_{13}),(v_{21},v_{22},v_{23}))&\\
&= \mathsf{L}(v_{21},v_{12},x_3)+\mathsf{L}(v_{11},v_{22},x_3)+\mathsf{L}(v_{21},x_2,v_{13})\\
&\quad+\mathsf{L}(v_{11},x_2,v_{23})+\mathsf{L}(x_1,v_{22},v_{13})+\mathsf{L}(x_1,v_{12},v_{23}),\\
\boldsymbol{D}^3\mathsf{L}(x_1,x_2,x_3)\cdot((v_{11},v_{12},v_{13}),(v_{21},v_{22},v_{23}),(v_{31},v_{32},v_{33}))&\\
&= \mathsf{L}(v_{11},v_{22},v_{33})+\mathsf{L}(v_{11},v_{32},v_{23})+\mathsf{L}(v_{21},v_{12},v_{33})\\
&\quad+\mathsf{L}(v_{21},v_{32},v_{13})+\mathsf{L}(v_{31},v_{12},v_{23})+\mathsf{L}(v_{31},v_{22},v_{13}).
\end{aligned}$$

For readers who understand the product rule of differentiation well, cf. Theorem 1.4.48, the preceding formulae are easy to derive. For readers for whom the formulae look mysterious, it is well to develop some facility in using them and like formulae since they come up often.

A case of particular importance occurs when $n_1 = \cdots = n_k = n$ and when all arguments of L are the same.

1.4.9 Corollary (Derivatives of multilinear maps II) *Let* $\mathsf{L} \in \mathrm{L}^k(\mathbb{R}^n;\mathbb{R}^m)$ *and define* $\mathbf{f}_{\mathsf{L}}\colon \mathbb{R}^n \to \mathbb{R}^m$ *by* $\mathbf{f}_{\mathsf{L}}(\mathbf{x}) = \mathsf{L}(\mathbf{x},\dots,\mathbf{x})$*. Then* $\mathbf{f}_{\mathsf{L}}$ *is infinitely differentiable and, moreover, for* $\mathrm{r} \in \{1,\dots,\mathrm{k}\}$ *we have*

$$\mathbf{D}^{\mathrm{r}}\mathbf{f}_{\mathsf{L}}(\mathbf{x})\cdot(\mathbf{v}_1,\dots,\mathbf{v}_{\mathrm{r}}) = \sum_{\sigma\in\mathfrak{S}_{\mathrm{r}}}\ \sum_{\{\mathrm{j}_1,\dots,\mathrm{j}_{\mathrm{r}}\}\in\mathrm{D}_{\mathrm{r},\mathrm{k}}} \mathsf{L}\circ\boldsymbol{\lambda}_{\mathrm{j}_1,\dots,\mathrm{j}_{\mathrm{r}}}((\mathbf{x},\dots,\mathbf{x}),(\mathbf{v}_{\sigma(1)},\dots,\mathbf{v}_{\sigma(\mathrm{r})})).$$

Proof Define $\mathsf{D} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^n\oplus\cdots\oplus\mathbb{R}^n)$ by $\mathsf{D}(x) = (x,\dots,x)$. Then $f_{\mathsf{L}} = \mathsf{L}\circ\mathsf{D}$. Let us also define, for any $r \in \mathbb{Z}_{>0}$, $\mathsf{D}_r^*\colon \mathrm{L}^r((\mathbb{R}^n)^k;\mathbb{R}^m) \to \mathrm{L}^r(\mathbb{R}^n;\mathbb{R}^m)$ by

$$\mathsf{D}_r^*(\mathsf{A})\cdot(v_1,\dots,v_r) = \mathsf{A}(\mathsf{D}(v_1),\dots,\mathsf{D}(v_r)).$$

Let us record the derivative of f_{L} in this case.

1 Lemma $\mathbf{D}^{\mathrm{r}}\mathbf{f}_{\mathsf{L}} = \mathsf{D}_{\mathrm{r}}^*\circ\mathbf{D}^{\mathrm{r}}\mathsf{L}\circ\mathsf{D}$.

Proof We prove the lemma by induction on r. For $r = 1$ we have

$$\boldsymbol{D}f_{\mathsf{L}}(x)\cdot v_1 = \boldsymbol{D}\mathsf{L}(\mathsf{D}(x))\circ\mathsf{D}(v_1),$$

using the Chain Rule below and the fact that the derivative of D is D since D is linear. This gives the result when $r = 1$, using the definition of D_1^*. So suppose the result holds for $r \in \{1,\dots,s\}$. Thus

$$\boldsymbol{D}^s f_{\mathsf{L}}(x)\cdot(v_1,\dots,v_s) = \boldsymbol{D}^s\mathsf{L}(\mathsf{D}(x))\cdot(\mathsf{D}(v_1),\dots,\mathsf{D}(v_s)).$$

Using Proposition 1.4.7 and the Chain Rule we then have

$$\begin{aligned}(\boldsymbol{D}^{s+1}f_{\mathsf{L}}(x)\cdot(v_0))\cdot(v_1,\dots,v_s) &= (\boldsymbol{D}(\boldsymbol{D}^s\mathsf{L})(\mathsf{D}(x))\circ\mathsf{D}(v_0))\cdot(\mathsf{D}(v_1),\dots,\mathsf{D}(v_s))\\ &= \boldsymbol{D}^{s+1}\mathsf{L}(\mathsf{D}(x))\cdot(\mathsf{D}(v_0),\mathsf{D}(v_1),\dots,\mathsf{D}(v_s)),\end{aligned}$$

where we use the isomorphism of Proposition I-5.6.7. This gives the lemma. ▼

The result now follows directly from Theorem 1.4.8. ■

The following trivial corollary is also worth recording separately.

1.4.10 Corollary (The derivative of a linear map) *If* $\mathsf{L} \in \mathsf{L}(\mathbb{R}^n;\mathbb{R}^m)$ *then* $\mathbf{D}\mathsf{L}(\mathbf{x}) = \mathsf{L}$ *for each* $\mathbf{x} \in \mathbb{R}^n$.

1.4.3 The directional derivative

In this section we describe another way of differentiating a function. As we shall see, this type of derivative is weaker than the derivative in the preceding section. However, it is perhaps a more intuitive notion of derivative, so we discuss it here to assist in understanding how one might interpret the derivative.

1.4.11 Definition (Directional derivative) Let $U \subseteq \mathbb{R}^n$ be open, let $f\colon U \to \mathbb{R}^m$, let $x_0 \in U$, and let $v \in \mathbb{R}^n$. The map f is ***differentiable in the direction*** $\mathbf{v}$ at x_0 if the map $s \mapsto f(x_0 + sv)$ is differentiable at $s = 0$. If f has a directional derivative at x_0 in the direction v then we denote by

$$\boldsymbol{D}f(x_0;v) = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} f(x_0 + sv)$$

the ***directional derivative***. If, for all $v \in \mathbb{R}^n$, f is differentiable in the direction v at x_0 then f is ***Gâteaux differentiable*** at x_0. •

We advise the reader to carefully note the distinction in the notation between the derivative at x_0 evaluated at v and the directional derivative at x_0 in the direction v. The former is denoted by $\boldsymbol{D}f(x_0)\cdot v$ while the latter is denoted by $\boldsymbol{D}f(x_0;v)$.

It is probably the case that the directional derivative is a more easily understood concept that the derivative. The idea of the directional derivative of f at x_0 in the direction of v is that one measures what is happening to the values of f as one steps away from x_0 in a specific direction. One might imagine that the existence of the derivative is equivalent to the existence of all partial derivatives. This, however, is false! Let us explore, therefore, the relationship between the derivative and the directional derivative.

1.4.12 Proposition (Differentiable maps are directionally differentiable) *Let* $\mathrm{U} \subseteq \mathbb{R}^n$ *be open and let* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^m$ *be differentiable at* $\mathbf{x}_0$. *Then, for any* $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{f}$ *has a directional derivative at* $\mathbf{x}_0$ *in the direction of* $\mathbf{v}$ *and, moreover,*

$$\mathbf{Df}(\mathbf{x}_0;\mathbf{v}) = \mathbf{Df}(\mathbf{x}_0)\cdot\mathbf{v}.$$

Proof Let $\epsilon \in \mathbb{R}_{>0}$ be such that $\boldsymbol{x}_0 + s\boldsymbol{v} \in U$ for each $s \in (-\epsilon, \epsilon)$; this is possible since U is open. Then let $\boldsymbol{g}\colon (-\epsilon, \epsilon) \to U$ be given by $\boldsymbol{g}(s) = \boldsymbol{x}_0 + s\boldsymbol{v}$. The existence of the directional derivative of $\boldsymbol{f}$ at $\boldsymbol{x}_0$ in the direction of $\boldsymbol{v}$ is then exactly the differentiability of $s \mapsto \boldsymbol{f} \circ \boldsymbol{g}(s)$ at $s = 0$. However, by the Chain Rule (Theorem 1.4.49), this function is indeed differentiable at $s = 0$ and, moreover,

$$\boldsymbol{Df}(\boldsymbol{x}_0; \boldsymbol{v}) = \boldsymbol{Df}(\boldsymbol{x}_0) \circ \boldsymbol{Dg}(0).$$

Note that $\boldsymbol{Dg}(0) \in \mathrm{L}(\mathbb{R}; \mathbb{R}^n)$ is simply the linear map $\alpha \mapsto \alpha\boldsymbol{v}$ and so

$$\boldsymbol{Df}(\boldsymbol{x}_0) \circ \boldsymbol{Dg}(0) \in \mathrm{L}(\mathbb{R}; \mathbb{R}^m)$$

is the linear map

$$\alpha \mapsto \alpha(\boldsymbol{Df}(\boldsymbol{x}_0) \cdot \boldsymbol{v}).$$

Upon making the natural identification of $\mathbb{R}^m$ with $\mathrm{L}(\mathbb{R}; \mathbb{R}^m)$ (i.e., the identification which assigns to $\boldsymbol{u} \in \mathbb{R}^m$ the linear map $\alpha \mapsto \alpha\boldsymbol{u}$) we see that we have the equality of derivatives asserted in the proposition. ■

In some sense the preceding result is reassuring since it tells us that the directional derivative interpretation can be made for the derivative when the latter exists. The following example shows, however, that the converse of the preceding result does not hold in general. Thus it is not the case that differentiability in all directions is equivalent to differentiability.

1.4.13 Example (Discontinuous function possessing all directional derivatives) We consider the function of Example 1.3.27:

$$f(x_1, x_2) = \begin{cases} \frac{x_1^2 x_2}{x_1^4 + x_2^2}, & (x_1, x_2) \neq (0, 0), \\ 0, & (x_1, x_2) = (0, 0). \end{cases}$$

In Example 1.3.27 we show that f is discontinuous at $(0, 0)$.

We further claim that f possesses all directional derivatives at $(0, 0)$. Indeed, let $(u_1, u_2) \in \mathbb{R}^2$ and consider the line

$$s \mapsto (0, 0) + s(u_1, u_2), \qquad s \in \mathbb{R},$$

through $(0, 0)$ in the direction of (u_1, u_2). Along this line we have

$$f((0, 0) + s(u_1, u_2)) = \frac{s u_1^2 u_2}{s^2 u_1^4 + u_2^2}.$$

A direct computation gives

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} f((0, 0) + s(u_1, u_2)) = \begin{cases} \frac{u_1^2}{u_2}, & u_2 \neq 0, \\ 0, & u_2 = 0, \end{cases}$$

which shows that f possesses all directional derivatives at $(0, 0)$. •

Having settled the relationship between the derivative and the directional derivative, let us give some of the properties of the directional derivative.

1.4.14 Proposition (Properties of the directional derivative) *Let* $U \subseteq \mathbb{R}^n$*, let* $\mathbf{f}, \mathbf{g}\colon U \to \mathbb{R}^m$*, let* $\mathbf{x}_0 \in U$*, let* $\mathbf{v} \in \mathbb{R}^n$*, and let* $a \in \mathbb{R}$*. If* $\mathbf{f}$ *and* $\mathbf{g}$ *are differentiable in the direction* $\mathbf{v}$ *at* $\mathbf{x}_0$ *then the following statements hold:*

(i) $\mathbf{f}$ *is differentiable in the direction* $\alpha\mathbf{v}$ *at* $\mathbf{x}_0$ *for each* $\alpha \in \mathbb{R}$ *and the map* $\alpha \mapsto \mathbf{Df}(\mathbf{x}_0; \alpha\mathbf{v})$ *is linear;*

(ii) $\mathbf{f} + \mathbf{g}$ *is differentiable in the direction* $\mathbf{v}$ *at* $\mathbf{x}_0$ *and*

$$\mathbf{D}(\mathbf{f}+\mathbf{g})(\mathbf{x}_0; \mathbf{v}) = \mathbf{Df}(\mathbf{x}_0; \mathbf{v}) + \mathbf{Dg}(\mathbf{x}_0; \mathbf{v});$$

(iii) $a\mathbf{f}$ *is differentiable in the direction* $\mathbf{v}$ *at* $\mathbf{x}_0$ *and*

$$\mathbf{D}(a\mathbf{f})(\mathbf{x}_0; \mathbf{v}) = a(\mathbf{Df}(\mathbf{x}_0; \mathbf{v}).$$

Moreover, if $m = 1$ *and we denote* $\mathbf{f}$ *and* $\mathbf{g}$ *by* f *and* g*, respectively, then under the same hypotheses as above we additionally have the following statements:*

(iv) fg *is differentiable in the direction* $\mathbf{v}$ *at* $\mathbf{x}_0$ *and*

$$\mathbf{D}(fg)(\mathbf{x}_0; \mathbf{v}) = g(\mathbf{x}_0)\mathbf{D}f(\mathbf{x}_0; \mathbf{v}) + f(\mathbf{x}_0)\mathbf{D}g(\mathbf{x}_0; \mathbf{v});$$

(v) *if* $g(\mathbf{x}_0) \neq 0$ *then* $\frac{f}{g}$ *is differentiable in the direction* $\mathbf{v}$ *at* $\mathbf{x}_0$ *and*

$$\mathbf{D}(\tfrac{f}{g})(\mathbf{x}_0; \mathbf{v}) = \frac{g(\mathbf{x}_0)\mathbf{D}f(\mathbf{x}_0; \mathbf{v}) - f(\mathbf{x}_0)\mathbf{D}g(\mathbf{x}_0; \mathbf{v})}{g(\mathbf{x}_0)^2}.$$

Proof (i) For $\alpha = 0$ we clearly have $\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}_0; \alpha\boldsymbol{v}) = \boldsymbol{0}$. So suppose that $\alpha \neq 0$. Then, letting $\sigma = \alpha s$ and using the Chain Rule, Theorem 1.4.49,

$$\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \boldsymbol{f}(\boldsymbol{x}_0 + s\alpha\boldsymbol{v}) = \frac{\mathrm{d}\sigma}{\mathrm{d}s}\,\frac{\mathrm{d}}{\mathrm{d}\sigma}\Big|_{\sigma=0} \boldsymbol{f}(\boldsymbol{x}_0 + \sigma\boldsymbol{v}) = \alpha\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}_0; \boldsymbol{v}),$$

giving this part of the result.

(ii) We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} (\boldsymbol{f}+\boldsymbol{g})(\boldsymbol{x}_0 + s\boldsymbol{v}) &= \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \boldsymbol{f}(\boldsymbol{x}_0 + s\boldsymbol{v}) + \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} \boldsymbol{g}(\boldsymbol{x}_0 + s\boldsymbol{v}) \\ &= \boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}_0; \boldsymbol{v}) + \boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}_0; \boldsymbol{v}),\end{aligned}$$

as desired, where we have used Proposition I-3.2.10.

(iii) This part of the result also follows from Proposition I-3.2.10.

(iv) We have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} (fg)(\boldsymbol{x}_0 + s\boldsymbol{v}) &= \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} f(\boldsymbol{x}_0 + s\boldsymbol{v})g(\boldsymbol{x}_0 + s\boldsymbol{v}) \\ &= \boldsymbol{D}f(\boldsymbol{x}_0; \boldsymbol{v}) + \boldsymbol{D}g(\boldsymbol{x}_0; \boldsymbol{v}),\end{aligned}$$

where we have used Proposition I-3.2.10.

(v) This also follows from Proposition I-3.2.10. ■

It is also possible to define higher-order directional derivatives. We let $U \subseteq \mathbb{R}^n$ be open, let $f\colon U \to \mathbb{R}^m$, let $x \in U$, and let $v_1, v_2 \in \mathbb{R}^n$. We suppose that the directional derivative $\boldsymbol{D}f(x_0 + sv_2; v_1)$ exists for each s sufficiently close to zero for some $x_0 \in U$. This allows the possibility of defining the directional derivative of the directional derivative:

$$\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \boldsymbol{D}f(x_0 + sv_2; v_1).$$

This procedure can be continued inductively.

1.4.15 Definition (Higher-order directional derivatives) Let $U \subseteq \mathbb{R}^n$ be open, let $f\colon U \to \mathbb{R}^m$, let $x_0 \in U$ and $v_0, v_1, \dots, v_{r-1} \in \mathbb{R}^n$, and suppose that f is differentiable in the directions $v_1, \dots, v_{r-1}$ at $x_0 + sv_0$ for $s \in (-\epsilon, \epsilon)$ with $\epsilon \in \mathbb{R}_{>0}$, with $\boldsymbol{D}^{r-1}f(x_0 + sx_0; v_1, \dots, v_{r-1})$ be the directional derivative. The vector

$$\boldsymbol{D}^r f(x_0; v_0, v_1, \dots, v_{r-1}) \triangleq \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \boldsymbol{D}^{r-1} f(x_0 + sx_0; v_1, \dots, v_{r-1})$$

in $\mathbb{R}^m$ is the ***directional derivative of*** $\mathbf{f}$ at x_0 in the directions $v_0, v_1, \dots, v_{r-1}$, when the derivative exists. •

We now have the following generalisation of Proposition 1.4.12.

1.4.16 Proposition (Higher-order derivative and directional derivatives) *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *and let* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$ *be* r *times differentiable at* $\mathbf{x}_0 \in \mathrm{U}$. *Then, for any* $\mathbf{v}_1, \dots, \mathbf{v}_{\mathrm{r}} \in \mathbb{R}^{\mathrm{n}}$, *the directional derivative of* $\mathbf{f}$ *at* $\mathbf{x}_0$ *and in the directions* $\mathbf{v}_1, \dots, \mathbf{v}_{\mathrm{r}}$ *exists and, moreover,*

$$\mathbf{D}^{\mathrm{r}}\mathbf{f}(\mathbf{x}_0; \mathbf{v}_1, \dots, \mathbf{v}_{\mathrm{r}}) = \mathbf{D}^{\mathrm{r}}\mathbf{f}(\mathbf{x}_0) \cdot (\mathbf{v}_1, \dots, \mathbf{v}_{\mathrm{r}}).$$

Proof We prove the result by induction on r, the case of $r = 1$ being Proposition 1.4.12. Suppose the result holds for $r = s$ and let f be $s+1$ times differentiable at x_0. By Proposition 1.4.35 and by the induction hypothesis the directional derivatives $\boldsymbol{D}^s f(x; v_1, \dots, v_s)$ exist for x in a neighbourhood of x_0 and for all $v_1, \dots, v_s$. Since

$$\boldsymbol{D}^s f(x; v_1, \dots, v_s) = \boldsymbol{D}^s f(x) \cdot (v_1, \dots, v_s)$$

by the induction hypothesis, it follows from Proposition 1.4.7 that

$$x \mapsto \boldsymbol{D}^s f(x; v_1, \dots, v_s)$$

is differentiable at x_0. By Proposition 1.4.12 it then holds that this map has a directional derivative at x_0 in the direction $v_0 \in \mathbb{R}^n$. Also by Proposition 1.4.12 it follows that

$$\begin{aligned}\boldsymbol{D}^{s+1} f(x_0; v_0, v_1, \dots, v_s) &= (\boldsymbol{D}(\boldsymbol{D}^s f)(x) \cdot v_0) \cdot (v_1, \dots, v_s) \\ &= \boldsymbol{D}^{s+1} f(x_0) \cdot (v_0, v_1, \dots, v_s),\end{aligned}$$

giving the result. ■

1.4.4 Derivatives and products, partial derivatives

The notion of a partial derivatives is one that is easy to understand in practice. That is to say, if one can compute derivatives, the matter of computing partial derivatives poses no problems in principle. However, this simplicity of computation can serve to obscure the rather important contribution of the *concept* of partial derivative to the theory of the derivative, and particularly higher-order derivatives. Therefore, in this section we present the partial derivative in a slightly general setting in order to give the partial derivative a little context. The appropriate general setting is that of functions defined on and taking values in products.

We first consider the case when we have a map $\boldsymbol{f}\colon A \to \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_k}$ from a subset $A \subseteq \mathbb{R}^n$ into a product of Euclidean spaces. In this case, following Example I-1.3.3–9, we write $\boldsymbol{f} = \boldsymbol{f}_1 \times \cdots \times \boldsymbol{f}_k$ for maps $\boldsymbol{f}_j\colon A \to \mathbb{R}^{m_j}$, $j \in \{1, \dots, k\}$; that is,

$$\boldsymbol{f}(\boldsymbol{x}) = (\boldsymbol{f}_1(\boldsymbol{x}), \dots, \boldsymbol{f}_k(\boldsymbol{x})), \qquad \boldsymbol{x} \in A.$$

We note that if $\boldsymbol{f}$ is differentiable at $\boldsymbol{x}_0 \in A$ then $\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}_0) \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^{m_1} \oplus \cdots \oplus \mathbb{R}^{m_k})$. As in Exercise I-5.4.5 we note that a linear map L from $\mathbb{R}^n$ into $\mathbb{R}^{m_1} \oplus \cdots \oplus \mathbb{R}^{m_k}$ can be written as

$$\mathsf{L}(v) = \mathsf{L}_1(v) + \cdots + \mathsf{L}_k(v)$$

for linear maps $\mathsf{L}_j\colon \mathbb{R}^n \to \mathbb{R}^{m_j}$, $j \in \{1, \dots, k\}$. Let us use the notation $\mathsf{L} = \mathsf{L}_1 \oplus \cdots \oplus \mathsf{L}_k$ to represent this fact. This notation can be extended to multilinear maps as well. Thus if $\mathsf{L} \in \mathrm{L}^k(\mathbb{R}^n; \mathbb{R}^{m_1} \oplus \cdots \oplus \mathbb{R}^{m_k})$ then we can write

$$\mathsf{L}(v_1, \dots, v_k) = \mathsf{L}_1(v_1, \dots, v_k) + \cdots + \mathsf{L}_k(v_1, \dots, v_k)$$

for $\mathsf{L}_j \in \mathrm{L}^k(\mathbb{R}^n; \mathbb{R}^{m_j})$, $j \in \{1, \dots, k\}$. We also write $\mathsf{L} = \mathsf{L}_1 \oplus \cdots \oplus \mathsf{L}_k$ in this case.

With all this notation we have the following result.

1.4.17 Proposition (Derivatives of maps taking values in products) *Let $U \subseteq \mathbb{R}^n$ be open and let $\mathbf{f}\colon A \to \mathbb{R}^{m_1} \times \cdots \times \mathbb{R}^{m_k}$ be a map which we write as $\mathbf{f} = \mathbf{f}_1 \times \cdots \times \mathbf{f}_k$. Then $\mathbf{f}$ is r times differentiable at $\mathbf{x}_0 \in U$ if and only if $\mathbf{f}_j$ is r times differentiable at $\mathbf{x}_0$ for each $j \in \{1, \dots, k\}$. Moreover, if $\mathbf{f}$ is r times differentiable at $\mathbf{x}_0$ then*

$$\mathbf{D}^r\mathbf{f}(\mathbf{x}_0) = \mathbf{D}^r\mathbf{f}_1(\mathbf{x}_0) \oplus \cdots \oplus \mathbf{D}^r\mathbf{f}_k(\mathbf{x}_0).$$

Proof Via an elementary inductive argument it suffices to prove the result in the case of $r = 1$, and so we restrict ourselves to this case.

Suppose that $\boldsymbol{f}$ is differentiable at $\boldsymbol{x}_0$ with derivative written as $\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}_0) = \mathsf{L}_1 \oplus \cdots \oplus \mathsf{L}_k$. Then, using the triangle inequality,

$$\frac{\|\boldsymbol{f}_j(\boldsymbol{x}) - \boldsymbol{f}_j(\boldsymbol{x}_0) - \mathsf{L}_j(\boldsymbol{x} - \boldsymbol{x}_0)\|_{\mathbb{R}^{m_j}}}{\|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n}} \le \frac{\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_0) - \boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{x} - \boldsymbol{x}_0)\|_{\mathbb{R}^{m_1 + \cdots + m_k}}}{\|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n}}, \qquad j \in \{1, \dots, k\}.$$

Therefore,

$$\lim_{x\to x_0} \frac{\|f_j(x) - f_j(x_0) - \mathsf{L}_j(x - x_0)\|_{\mathbb{R}^{m_j}}}{\|x - x_0\|_{\mathbb{R}^n}} = 0, \qquad j \in \{1, \dots, k\},$$

giving differentiability of f_j at x_0 with derivative L_j for each $j \in \{1, \dots, k\}$.

For the converse, suppose that $f_1, \dots, f_k$ are differentiable at x_0 and let

$$\mathsf{L} = \boldsymbol{D}f_1(x_0) \oplus \cdots \oplus \boldsymbol{D}f_k(x_0).$$

Then, using the triangle inequality,

$$\frac{\|f(x) - f(x_0) - \mathsf{L}(x - x_0)\|_{\mathbb{R}^{m_1+\cdots+m_k}}}{\|x - x_0\|_{\mathbb{R}^n}} \le \sum_{j=1}^{k} \frac{\|f_j(x) - f_j(x_0) - \boldsymbol{D}f_j(x_0)\cdot(x - x_0)\|_{\mathbb{R}^{m_j}}}{\|x - x_0\|_{\mathbb{R}^n}}.$$

Thus

$$\lim_{x\to x_0} \frac{\|f(x) - f(x_0) - \mathsf{L}(x - x_0)\|_{\mathbb{R}^{m_1+\cdots+m_k}}}{\|x - x_0\|_{\mathbb{R}^n}} = 0,$$

giving differentiability of f at x_0. Uniqueness of the derivative now also ensures that the final assertion of the result holds. ■

Now we turn to the case of primary interest, that when the domain of the function is a product.

1.4.18 Definition (Partial derivative) Let $U \subseteq \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ be open, let $f\colon U \to \mathbb{R}^m$, let $x_0 = (x_{01}, \dots, x_{0k}) \in U$, and let $j \in \{1, \dots, k\}$.

(i) The map f is ***differentiable at*** $\mathbf{x}_0$ ***with respect to the*** $\mathbf{j}$***th component*** if the map

$$U \cap (\{x_{01}\} \times \cdots \times \mathbb{R}^{n_j} \times \cdots \times \{x_{0k}\} \ni x_j) \mapsto f(x_{01}, \dots, x_j, \dots, x_{0k}) \in \mathbb{R}^m \tag{1.22}$$

is differentiable at x_{0j}.

(ii) If f is differentiable at x_0 with respect to the jth component, then the derivative at x_{j0} of the map (1.22) is denoted by $\boldsymbol{D}_j f(x_0)$ and is called the $\mathbf{j}$***th partial derivative*** of f at x_0. •

For the reader who cannot quite imagine what is the connection with the usual notion of partial derivative, we ask that they hang on for just a moment as this will be made clear soon enough. First let us record the relationship between the derivative and the partial derivatives.

1.4.19 Theorem (Partial derivatives and derivatives) *If* $\mathrm{U} \subseteq \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ *is an open set and if* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^m$ *is a map differentiable at* $\mathbf{x}_0 \in \mathrm{U}$*, then* $\mathbf{f}$ *is differentiable at* $\mathbf{x}_0$ *with respect to the* j*th component for each* $\mathrm{j} \in \{1, \dots, \mathrm{k}\}$*. Moreover, if* $\mathbf{f}$ *is differentiable at* $\mathbf{x}_0$ *then we have the following relationships between the derivative and the partial derivatives:*

$$\mathbf{D}_j\mathbf{f}(\mathbf{x}_0)\cdot\mathbf{v}_j = \mathbf{D}\mathbf{f}(\mathbf{x}_0)\cdot(\mathbf{0}, \dots, \mathbf{v}_j, \dots, \mathbf{0})$$

$$\mathbf{D}\mathbf{f}(\mathbf{x}_0)\cdot(\mathbf{v}_1, \dots, \mathbf{v}_k) = \sum_{j=1}^{k} \mathbf{D}_j\mathbf{f}(\mathbf{x}_0)\cdot\mathbf{v}_j.$$

Proof Let us denote $x_0 = (x_{01}, \dots, x_{0k})$. Differentiability of f at x_0 implies, in particular, that

$$\lim_{x_j \to x_{0j}} \Big(\| f(x_{01}, \dots, x_j, \dots, x_{k0}) - f(x_{01}, \dots, x_{0j}, \dots, x_{0k}) - \boldsymbol{D} f(x_0) \cdot (\boldsymbol{0}, \dots, x_j - x_{0j}, \dots, \boldsymbol{0}) \|_{\mathbb{R}^m} \Big) / \Big(\| x_j - x_{0j} \|_{\mathbb{R}^{n_j}} \Big) = 0.$$

This precisely means that f is differentiable at x_0 with respect to the jth component.

Now let $v_j \in \mathbb{R}^{n_j}$ and denote $v = (\boldsymbol{0}, \dots, v_j, \dots, \boldsymbol{0})$. By twice applying Proposition 1.4.12 we have

$$\boldsymbol{D} f(x_0) \cdot v = \frac{\mathrm{d}}{\mathrm{d}s}\Big|_{s=0} f(x_0 + sv) = f(x_{01}, \dots, x_{0j} + sv_j, \dots, x_{0k}) = \boldsymbol{D}_j f(x_0) \cdot v_j.$$

By linearity of the derivative we then have

$$\boldsymbol{D} f(x_0) \cdot (v_1, \dots, v_k) = \sum_{j=1}^{k} \boldsymbol{D} f(x_0) \cdot (\boldsymbol{0}, \dots, v_j, \dots, \boldsymbol{0}) = \sum_{j=1}^{k} \boldsymbol{D}_j f(x_0) \cdot v_j,$$

which completes the proof. ■

If we combine Proposition 1.4.17 and Theorem 1.4.19 then we get the following general result concerning derivatives and products.

1.4.20 Corollary (Derivatives and products) *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}_1} \times \cdots \times \mathbb{R}^{\mathrm{n}_\mathrm{r}}$ *be an open set and let* $\mathbf{f} \colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}_1} \times \cdots \times \mathbb{R}^{\mathrm{m}_\mathrm{s}}$ *be a map that we write as* $\mathbf{f} = \mathbf{f}_1 \times \cdots \times \mathbf{f}_\mathrm{s}$. *If* $\mathbf{f}$ *is differentiable at* $\mathbf{x}_0 \in \mathrm{U}$ *then, for each* $\mathrm{j} \in \{1, \dots, \mathrm{r}\}$ *and* $\mathrm{k} \in \{1, \dots, \mathrm{s}\}$, $\mathbf{f}_\mathrm{k}$ *is differentiable at* $\mathbf{x}_0$ *with respect to the* j*th component. Moreover, if* $\mathbf{f}$ *is differentiable at* $\mathbf{x}_0 \in \mathrm{U}$ *then*

$$\mathbf{Df}(\mathbf{x}_0) \cdot (\mathbf{v}_1, \dots, \mathbf{v}_\mathrm{r}) = \Big(\sum_{\mathrm{j}_1=1}^{\mathrm{r}} \mathbf{D}_{\mathrm{j}_1} \mathbf{f}_1(\mathbf{x}_0) \cdot \mathbf{v}_{\mathrm{j}_1}, \dots, \sum_{\mathrm{j}_\mathrm{s}=1}^{\mathrm{r}} \mathbf{D}_{\mathrm{j}_\mathrm{s}} \mathbf{f}_\mathrm{s}(\mathbf{x}_0) \cdot \mathbf{v}_{\mathrm{j}_\mathrm{s}} \Big),$$

While the above presentation makes it look like the product structure is special, of course this is not the case. Every Euclidean space is a product of copies of $\mathbb{R}^1$, by definition. Therefore, the above presentation can always be applied to this natural product structure of every Euclidean space. Moreover, using this product structure sheds some light on the derivative and how to compute it. We see this as follows.

1.4.21 Definition (Jacobian matrix) Let $U \subseteq \mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$ be differentiable, let $f \colon U \to \mathbb{R}^m = \mathbb{R} \times \cdots \times \mathbb{R}$ be differentiable at $x_0 \in U$, and write $f = f_1 \times \cdots \times f_m$ for $f_1, \dots, f_m \colon U \to \mathbb{R}$.

(i) The ***j*th partial derivative** of f at x_0 is $\boldsymbol{D}_j f(x_0) \in \mathbb{R}^m$ (noting that $\mathrm{L}(\mathbb{R}; \mathbb{R}^m)$ is isomorphic to $\mathbb{R}^m$ by Exercise I-5.6.5).

(ii) The ***j*th partial derivative** of the kth component of f at x_0 is $\boldsymbol{D}_j f_k(x_0) \in \mathbb{R}$ (noting that $\mathrm{L}(\mathbb{R}; \mathbb{R})$ is isomorphic to $\mathbb{R}$ by Exercise I-5.6.5).

(iii) The ***Jacobian matrix*** of f at x_0 is the $m \times n$ matrix

$$\begin{bmatrix} D_1 f_1(x_0) & \cdots & D_n f_1(x_0) \\ \vdots & \ddots & \vdots \\ D_1 f_m(x_0) & \cdots & D_n f_m(x_0) \end{bmatrix}.$$ •

Note that we use the same terminology "jth partial derivative" for the specific case of the preceding definition as we used in the more general case of Definition 1.4.18. This is a legitimate source of possible confusion, but is also standard practice.

The next result follows immediately from Corollary 1.4.20, and is quite important since it tells us how one computes the derivative in practice.

1.4.22 Theorem (Explicit formula for the derivative) *If* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *is an open set and if* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$ *is a map differentiable at* $\mathbf{x}_0 \in \mathrm{U}$ *written as* $\mathbf{f} = \mathrm{f}_1 \times \cdots \times \mathrm{f}_{\mathrm{m}}$*, then the components* $\mathrm{f}_1, \dots, \mathrm{f}_{\mathrm{m}}\colon \mathrm{U} \to \mathbb{R}$ *of* $\mathbf{f}$ *are differentiable at* $\mathbf{x}_0$ *with respect to the* j*th coordinate for each* $\mathrm{j} \in \{1, \dots, \mathrm{n}\}$*. Furthermore, the matrix representative of* $\mathbf{Df}(\mathbf{x}_0)$ *with respect to the standard bases* $\mathscr{B}_{\mathrm{n}}$ *and* $\mathscr{B}_{\mathrm{m}}$ *for* $\mathbb{R}^{\mathrm{n}}$ *and* $\mathbb{R}^{\mathrm{m}}$ *is the Jacobian matrix of* $\mathbf{f}$ *at* $\mathbf{x}_0$*.*

We shall frequently think of the derivative as being *equal* to its Jacobian matrix with the understanding that we are using the standard basis to represent the components of the derivative as a linear map. This is convenient to do, and is only a mild abuse at worst.

1.4.23 Notation (Alternative notation for the partial derivative) As with the notation for the derivative as discussed in Notation I-3.2.2, there is notation for the partial derivative that sees more common use that the notation we give. Specifically, it is frequent to see the symbol $\frac{\partial f}{\partial x_j}$ used for what we denote by $D_j f$. This more common notation suffers from the same drawbacks as the notation $\frac{\mathrm{d}f}{\mathrm{d}x}$ for the ordinary derivative. Namely, it introduces the independent variable x_j in a potentially confusing way. Much of the time, this does not cause problems, and indeed we will use this notation when it is not imprudent to do so. •

In Exercise 1.4.3 the reader can provide a rule that is often helpful in computing partial derivatives with respect to coordinates. Let us give a couple of examples to illustrate the notion of partial derivative and its connection with the derivative.

1.4.24 Examples (Partial derivative)

1. Let $U = \mathbb{R}^2 \setminus \{(0,0)\}$ and define $f\colon U \to \mathbb{R}^2$ by $f(x_1, x_2) = \left(\frac{x_1}{\sqrt{x_1^2+x_2^2}}, \frac{x_2}{\sqrt{x_1^2+x_2^2}}\right)$. We claim that f possesses both partial derivatives at all points in U. Indeed, we

compute

$$\lim_{h\to 0}\frac{\Big(\frac{x_1+h}{\sqrt{(x_1+h)^2+x_2^2}},\frac{x_2}{\sqrt{(x_1+h)^2+x_2^2}}\Big)-\Big(\frac{x_1}{\sqrt{x_1^2+x_2^2}},\frac{x_2}{\sqrt{x_1^2+x_2^2}}\Big)}{h}$$
$$=\left(\lim_{h\to 0}\frac{\frac{x_1+h}{\sqrt{(x_1+h)^2+x_2^2}}-\frac{x_1}{\sqrt{x_1^2+x_2^2}}}{h},\lim_{h\to 0}\frac{\frac{x_2}{\sqrt{(x_1+h)^2+x_2^2}}-\frac{x_2}{\sqrt{x_1^2+x_2^2}}}{h}\right)$$
$$=\Big(\frac{x_2^2}{(x_1^2+x_2^2)^{3/2}},-\frac{x_1x_2}{(x_1^2+x_2^2)^{3/2}}\Big),$$

where, in the last step, we have simply computed the usual derivative, using the rules given in Section I-3.2. In like manner we have

$$\lim_{h\to 0}\frac{\Big(\frac{x_1}{\sqrt{x_1^2+x_2^2}},\frac{x_2+h}{\sqrt{x_1^2+(x_2+h)^2}}\Big)-\Big(\frac{x_1}{\sqrt{x_1^2+(x_2+h)^2}},\frac{x_2}{\sqrt{x_1^2+x_2^2}}\Big)}{h}=\Big(-\frac{x_1x_2}{(x_1^2+x_2^2)^{3/2}},\frac{x_1^2}{(x_1^2+x_2^2)^{3/2}}\Big)$$

Thus both partial derivatives indeed exist, and we moreover have

$$\boldsymbol{D}_1\boldsymbol{f}(x_1,x_2)=\Big(\frac{x_2^2}{(x_1^2+x_2^2)^{3/2}},-\frac{x_1x_2}{(x_1^2+x_2^2)^{3/2}}\Big),$$
$$\boldsymbol{D}_2\boldsymbol{f}(x_1,x_2)=\Big(-\frac{x_1x_2}{(x_1^2+x_2^2)^{3/2}},\frac{x_1^2}{(x_1^2+x_2^2)^{3/2}}\Big),$$

and so the partial derivatives are also continuous functions on U.

Therefore, *if* $\boldsymbol{f}$ is differentiable at some point $(x_{01},x_{02})\in\mathbb{R}^2\setminus\{(0,0)\}$ then it must hold that

$$\boldsymbol{D}\boldsymbol{f}(x_{01},x_{02})=\begin{bmatrix}\frac{x_{02}^2}{(x_{01}^2+x_{02}^2)^{3/2}} & -\frac{x_{01}x_{02}}{(x_{01}^2+x_{02}^2)^{3/2}}\\ -\frac{x_{01}x_{02}}{(x_{01}^2+x_{02}^2)^{3/2}} & \frac{x_{01}^2}{(x_{01}^2+x_{02}^2)^{3/2}}\end{bmatrix},$$

where we identify the derivative with its matrix representative in the standard basis. We should, at this point since we know no better, actually verify that $\boldsymbol{f}$ is differentiable with this derivative. This can be done directly using the definition of derivative. Thus one can check directly, using rules for limits as in Proposition I-2.3.23, that

$$\lim_{(x_1,x_2)\to(x_{01},x_{02})}\Big(\|\boldsymbol{f}(x_1,x_2)-\boldsymbol{f}(x_{01},x_{02})$$
$$-\boldsymbol{D}\boldsymbol{f}(x_{01},x_{02})\cdot(x_1-x_{01},x_2-x_{02})\|_{\mathbb{R}^2}\Big)/\|(x_1,x_2)-(x_{01},x_{02})\|_{\mathbb{R}^2}=0.$$

We leave the tedious verification of this to the reader, particularly as we shall see in Theorem 1.4.25 below that in this example there is an easy way to verify that this function is, in fact, of class C^1.

2. Define $f\colon \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x_1, x_2) = \begin{cases} \frac{2x_1^2 x_2}{x_1^4 + x_2^2}, & (x_1, x_2) \neq (0,0), \\ 0, & (x_1, x_2) = (0,0). \end{cases}$$

We claim that f possesses both partial derivatives at $(0,0)$, but is not differentiable at $(0,0)$. Let us first show that f possesses both partial derivative at $(0,0)$. By definition, this amounts to checking the differentiability (in the sense of Definition I-3.2.1) of the function $x_1 \mapsto f(x_1, 0) = 0$. This function, being constant, is obviously differentiable at $(0,0)$ with derivative zero. In like manner one can show that f possesses the second partial derivative at $(0,0)$ and that this second partial derivative is also zero. Now let us show that f is discontinuous, and therefore not differentiable, at $(0,0)$. Consider the sequence $((\frac{1}{j}, \frac{1}{j^2}))_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}^2$. This sequence converges to $(0,0)$. We directly compute that $f(\frac{1}{j}, \frac{1}{j^2}) = 1$ for all $j \in \mathbb{Z}_{>0}$. Therefore

$$\lim_{j\to\infty} f(\tfrac{1}{j}, \tfrac{1}{j^2}) = 1 \neq f(0,0).$$

Therefore, f is indeed discontinuous, and so not differentiable, at $(0,0)$ by Proposition 1.4.35 below.

Note that the function of Example 1.4.13 also has the property that its partial derivatives exist, but the function is not differentiable. •

The preceding examples illustrate one of the problems that one has with the derivative: it is often not so easy to verify its existence since the mere existence of all partial derivatives is not sufficient. There is an important case, however, where one can infer differentiability from the properties of the partial derivatives. Here we return to the general setup for the partial derivative in terms of products.

1.4.25 Theorem (Equivalence of continuous differentiability and continuity of partial derivatives) *For an open set* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}_1} \times \cdots \times \mathbb{R}^{\mathrm{n}_k}$*, a map* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$*, and for* $\mathrm{r} \in \mathbb{Z}_{>0}$*, the following statements are equivalent:*

(i) $\mathbf{f}$ *is of class* C^{r}*;*

(ii) the partial derivatives $\mathbf{D}_{\mathrm{j}}\mathbf{f}(\mathbf{x})$ *exist for each* $\mathrm{j} \in \{1, \dots, \mathrm{k}\}$ *and* $\mathbf{x} \in \mathrm{U}$*, and, moreover, the maps* $\mathbf{x} \mapsto \mathbf{D}_{\mathrm{j}}\mathbf{f}(\mathbf{x})$ *are of class* $\mathrm{C}^{\mathrm{r}-1}$*.*

Proof By induction we can assume without loss of generality that $k = 2$. Moreover, by Propositions 1.3.26 and 1.4.17 we can take $m = 1$ without loss of generality. Thus we prove the theorem for $k = 2$ and $m = 1$. Consistent with our standing conventions we write "$\boldsymbol{f}$" as "f."

(i) $\Longrightarrow$ (ii) From Theorem 1.4.19 we know that the partial derivatives $\boldsymbol{D}_1 f(\boldsymbol{x})$ and $\boldsymbol{D}_2 f(\boldsymbol{x})$ exist at all points $\boldsymbol{x} \in U$. To prove continuity of the partial derivatives, define maps

$$\phi_1\colon \mathrm{L}(\mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}; \mathbb{R}) \to \mathrm{L}(\mathbb{R}^{n_1}; \mathbb{R}), \quad \phi_2\colon \mathrm{L}(\mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}; \mathbb{R}) \to \mathrm{L}(\mathbb{R}^{n_2}; \mathbb{R})$$

by

$$\phi_1(\mathsf{L}_1)(\boldsymbol{v}_1) = \mathsf{L}_1(\boldsymbol{v}_1, \boldsymbol{0}), \quad \phi_2(\mathsf{L}_2)(\boldsymbol{v}_2) = \mathsf{L}_1(\boldsymbol{0}, \boldsymbol{v}_2)$$

for $\boldsymbol{v}_1 \in \mathbb{R}^{n_1}$ and $\boldsymbol{v}_2 \in \mathbb{R}^{n_2}$. These maps are easily verified to be linear and so in particular are infinitely differentiable, cf. Corollary 1.4.10. Moreover, we easily see that

$$\boldsymbol{D}_1 f = \phi_1 \circ \boldsymbol{D} f, \quad \boldsymbol{D}_2 f = \phi_2 \circ \boldsymbol{D} f.$$

Therefore, if $\boldsymbol{D} f$ is of class C^{r-1} (as it is by Proposition 1.4.35) then the partial derivatives are also of class C^{r-1}.

(ii) $\implies$ (i) First we show that $\boldsymbol{D} f(\boldsymbol{x})$ exists for all $\boldsymbol{x} \in U$ if all partial derivatives exist at each point, and are continuous. Let us fix $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2) \in U$ and let $(\boldsymbol{h}_1, \boldsymbol{h}_2) \in \mathbb{R}^{n_1} \times \mathbb{R}^{n_2}$ be such that $(\boldsymbol{x}_1 + s_1 \boldsymbol{h}_1, \boldsymbol{x}_2 + s_2 \boldsymbol{h}_2) \in U$ for $s_1, s_2 \in [0,1]$, this being possible since U is open. Consider the map

$$s \mapsto f(\boldsymbol{x}_1, \boldsymbol{x}_2 + s\boldsymbol{h}_2).$$

By the Chain Rule (Theorem 1.4.49), it being applicable since the partial derivative of f with respect to the second component exists, we have

$$\frac{\mathrm{d}}{\mathrm{d}s} f(\boldsymbol{x}_1, \boldsymbol{x}_2 + s\boldsymbol{h}_2) = \boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2 + s\boldsymbol{h}_2) \cdot \boldsymbol{h}_2.$$

ref? By the multivariable Fundamental Theorem of Calculus (this is obtained in this case by applying the single-variable Fundamental Theorem componentwise, but the reader can also refer ahead to) we have

$$f(\boldsymbol{x}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2) - f(\boldsymbol{x}_1, \boldsymbol{x}_2) = \int_0^1 \boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2 + s\boldsymbol{h}_2) \cdot \boldsymbol{h}_2 \, \mathrm{d}s. \tag{1.23}$$

The same argument can be applied to the map

$$s \mapsto f(\boldsymbol{x}_1 + s\boldsymbol{h}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2)$$

to give

$$f(\boldsymbol{x}_1 + \boldsymbol{h}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2) - f(\boldsymbol{x}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2) = \int_0^1 \boldsymbol{D}_1 f(\boldsymbol{x}_1 + s\boldsymbol{h}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2) \cdot \boldsymbol{h}_1 \, \mathrm{d}s. \tag{1.24}$$

Combining (1.23) and (1.24) we get

$$\begin{aligned} f(\boldsymbol{x}_1{+}\boldsymbol{h}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2) - f(\boldsymbol{x}_1, \boldsymbol{h}_2) - \boldsymbol{D}_1 f(\boldsymbol{x}_1, \boldsymbol{x}_2) \cdot \boldsymbol{h}_1 - \boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2) \cdot \boldsymbol{h}_2 \\ = \int_0^1 \boldsymbol{D}_1 f(\boldsymbol{x}_1 + s\boldsymbol{h}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2) \cdot \boldsymbol{h}_1 \, \mathrm{d}s + \int_0^1 \boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2 + s\boldsymbol{h}_2) \cdot \boldsymbol{h}_2 \, \mathrm{d}s \\ - \boldsymbol{D}_1 f(\boldsymbol{x}_1, \boldsymbol{x}_2) \cdot \boldsymbol{h}_1 - \boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2) \cdot \boldsymbol{h}_2 \\ = \Big(\int_0^1 \Big(\boldsymbol{D}_2 f(\boldsymbol{x}_1 + s\boldsymbol{h}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2) - \boldsymbol{D}_1 f(\boldsymbol{x}_1, \boldsymbol{x}_2) \Big) \mathrm{d}s \Big) \cdot \boldsymbol{h}_1 + \\ \Big(\int_0^1 \Big(\boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2 + s\boldsymbol{h}_2) - \boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2) \Big), \mathrm{d}s \Big) \cdot \boldsymbol{h}_2 \end{aligned}$$

Now let $\epsilon \in \mathbb{R}_{>0}$ and by continuity of the partial derivatives choose $(\boldsymbol{h}_1, \boldsymbol{h}_2)$ such that

$$\sup\{\|\boldsymbol{D}_2 f(\boldsymbol{x}_1 + s\boldsymbol{h}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2) - \boldsymbol{D}_1 f(\boldsymbol{x}_1, \boldsymbol{x}_2)\|_{\mathbb{R}^n,\mathbb{R}^m} \mid s \in [0,1]\} < \frac{\epsilon}{2\sqrt{2}}$$
$$\sup\{\|\boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2 + s\boldsymbol{h}_2) - \boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2)\|_{\mathbb{R}^n,\mathbb{R}^m} \mid s \in [0,1]\} < \frac{\epsilon}{2\sqrt{2}}.$$

With $(\boldsymbol{h}_1, \boldsymbol{h}_2)$ so chosen we have

$$\begin{aligned}\Big|\Big(\boldsymbol{D}_2 f(\boldsymbol{x}_1 + s\boldsymbol{h}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2) - \boldsymbol{D}_1 f(\boldsymbol{x}_1, \boldsymbol{x}_2)\Big) \cdot \boldsymbol{h}_1 \\ + \Big(\boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2 + s\boldsymbol{h}_2) - \boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2)\Big) \cdot \boldsymbol{h}_2\Big| &\le \tfrac{\epsilon}{2\sqrt{2}}\|\boldsymbol{h}_1\|_{\mathbb{R}^{n_1}} + \tfrac{\epsilon}{2\sqrt{2}}\|\boldsymbol{h}_2\|_{\mathbb{R}^{n_2}} \\ &\le \tfrac{\epsilon}{\sqrt{2}}(\|\boldsymbol{h}_1\|_{\mathbb{R}^{n_1}} + \|\boldsymbol{h}_2\|_{\mathbb{R}^{n_2}}) \le \epsilon\|(\boldsymbol{h}_1, \boldsymbol{h}_2)\|_{\mathbb{R}^{n_1+n_2}},\end{aligned}$$

using Lemma 1.2.67. Therefore,

$$\Big|f(\boldsymbol{x}_1 + \boldsymbol{h}_1, \boldsymbol{x}_2 + \boldsymbol{h}_2) - f(\boldsymbol{x}_1, \boldsymbol{h}_2) - \boldsymbol{D}_1 f(\boldsymbol{x}_1, \boldsymbol{x}_2) \cdot \boldsymbol{h}_1 - \boldsymbol{D}_2 f(\boldsymbol{x}_1, \boldsymbol{x}_2) \cdot \boldsymbol{h}_2\Big| / \|(\boldsymbol{h}_1, \boldsymbol{h}_2)\|_{\mathbb{R}^{n_1+n_2}} < \epsilon,$$

and so we conclude that f is differentiable at $(\boldsymbol{x}_1, \boldsymbol{x}_2)$.

Finally, we show that $\boldsymbol{D}f$ is of class C^{r-1} if both $\boldsymbol{D}_1 f$ and $\boldsymbol{D}_2 f$ are of class C^{r-1}. Define maps

$$\psi_1 \colon \mathsf{L}(\mathbb{R}^{n_1}; \mathbb{R}) \to \mathsf{L}(\mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}; \mathbb{R}), \quad \psi_2 \colon \mathsf{L}(\mathbb{R}^{n_2}; \mathbb{R}) \to \mathsf{L}(\mathbb{R}^{n_1} \oplus \mathbb{R}^{n_2}; \mathbb{R})$$

by

$$\psi_1(\mathsf{L}_1)(\boldsymbol{v}_1, \boldsymbol{v}_2) = \mathsf{L}_1(\boldsymbol{v}_1), \quad \psi_2(\mathsf{L}_2)(\boldsymbol{v}_1, \boldsymbol{v}_2) = \mathsf{L}_2(\boldsymbol{v}_2).$$

These maps are linear and so infinitely differentiable. Moreover, since

$$\boldsymbol{D}f(\boldsymbol{x}) = \psi_1 \circ \boldsymbol{D}_1 f + \psi_2 \circ \boldsymbol{D}_2 f$$

it follows that $\boldsymbol{D}f$ is of class C^{r-1} if $\boldsymbol{D}_1 f$ and $\boldsymbol{D}_2 f$ are of class C^{r-1} by virtue of Proposition 1.4.47. ■

Let us consider the theorem in view of the examples we introduced above.

1.4.26 Examples (Partial derivatives (cont'd))

1. We take $U = \mathbb{R}^2 \setminus \{(0,0)\}$ and take $\boldsymbol{f} \colon U \to \mathbb{R}^2$ given by $\boldsymbol{f}(x_1, x_2) = \Big(\frac{x_1}{\sqrt{x_1^2+x_2^2}}, \frac{x_2}{\sqrt{x_1^2+x_2^2}}\Big)$. In Example 1.4.24–1 we computed

$$\boldsymbol{D}\boldsymbol{f}(x_1, x_2) = \begin{bmatrix} \frac{x_2^2}{(x_1^2+x_2^2)^{3/2}} & -\frac{x_1 x_2}{(x_1^2+x_2^2)^{3/2}} \\ -\frac{x_1 x_2}{(x_1^2+x_2^2)^{3/2}} & \frac{x_1^2}{(x_1^2+x_2^2)^{3/2}} \end{bmatrix}.$$

Since the components of this matrix are continuous functions on U, it follows from Theorem 1.4.25 that $\boldsymbol{f}$ is of class C^1 on U.

2. Here we take $f\colon \mathbb{R}^2 \to \mathbb{R}$ to be defined by

$$f(x_1,x_2) = \begin{cases} \frac{2x_1^2x_2}{x_1^4+x_2^2}, & (x_1,x_2) \neq (0,0), \\ 0, & (x_1,x_2) = (0,0). \end{cases}$$

In Example 1.4.24–2 we showed that both partial derivatives of f exist at $(0,0)$ and are zero. For $(x_1,x_2) \neq (0,0)$ we can compute, using Theorem 1.4.22,

$$\boldsymbol{D}_1 f(x_1,x_2) = 2x_1x_2\frac{x_2^2 - x_1^4}{(x_1^4+x_2^2)^2}, \quad \boldsymbol{D}_2 f(x_1,x_2) = x_1^2\frac{x_1^4 - x_2^2}{(x_1^4+x_2^2)^2}.$$

These partial derivatives are continuous on $\mathbb{R}^2 \setminus \{(0,0)\}$, and so it follows from Theorem 1.4.25 that f is of class C^1 on $\mathbb{R}^2 \setminus \{(0,0)\}$. However, the partial derivatives are readily verified to be discontinuous at $(0,0)$, cf. Example 1.4.24–2, and so it follows from Theorem 1.4.25 that f is not of class C^1 in any neighbourhood of $(0,0)$. Of course, we knew this already since f is actually discontinuous at $(0,0)$. •

1.4.5 Iterated partial derivatives

Now that we have used the notion of partial derivative to get better handle on how to compute the derivative of a multivariable map, let us see if we can similarly compute higher-order derivatives of multivariable maps using partial derivatives. In addressing this matter we will also shed some light on an important property of higher-order derivatives in the usual sense. In particular, we shall illuminate clearly the significance of the classical statement that "partial derivatives commute" by showing that this statement is not true in general.

Suppose we have an open set $U \subseteq \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ and a map $\boldsymbol{f}\colon U \to \mathbb{R}^m$. Suppose that for $j_1 \in \{1,\dots,k\}$, $\boldsymbol{f}$ is continuously differentiable with respect to the j_1st component. That is, the map

$$U \ni \boldsymbol{x} \mapsto \boldsymbol{D}_{j_1}\boldsymbol{f}(\boldsymbol{x}) \in \mathrm{L}(\mathbb{R}^{n_1};\mathbb{R}^m)$$

is defined and continuous. While there are weaker conditions that will guarantee this, to keep things simple let us suppose that $\boldsymbol{f}$ is of class C^1 so the existence and continuity of the partial derivative is ensured by Theorem 1.4.19. Now let $j_2 \in \{1,\dots,k\}$. We can then talk about the differentiability of the map $U \ni \boldsymbol{x} \mapsto \boldsymbol{D}_{j_1}\boldsymbol{f}(\boldsymbol{x})$ with respect to the j_2nd component. Indeed, while again weaker hypotheses are possible, if we assume that $\boldsymbol{f}$ is of class C^2 then the map

$$U \ni \boldsymbol{x} \mapsto \boldsymbol{D}_{j_2}\boldsymbol{D}_{j_1}\boldsymbol{f}(\boldsymbol{x}) \in \mathrm{L}(\mathbb{R}^{n_{j_2}},\mathbb{R}^{n_{j_1}};\mathbb{R}^m)$$

is defined and continuous by virtue of Theorem 1.4.19. (We use Proposition I-5.6.7 to describe the codomain of this map.) Clearly, if $\boldsymbol{f}$ is of class C^r and if $j_1,\dots,j_r \in \{1,\dots,k\}$ then we can inductively define

$$U \ni \boldsymbol{x} \mapsto \boldsymbol{D}_{j_r}\cdots\boldsymbol{D}_{j_1}\boldsymbol{f}(\boldsymbol{x}) \in \mathrm{L}(\mathbb{R}^{n_{j_r}},\dots,\mathbb{R}^{n_{j_1}};\mathbb{R}^m),$$

again using Proposition I-5.6.7.

Let us organise the preceding discussion by naming the objects.

1.4.27 Definition (Iterated partial derivative) Let $U \subseteq \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ be open, let $f\colon U \to \mathbb{R}^m$, let $x_0 \in U$, and let $j_1, \dots, j_r \in \{1, \dots, k\}$. The multilinear map

$$D_{j_r} \cdots D_{j_1} f(x_0) \in \mathrm{L}(\mathbb{R}^{n_{j_r}}, \dots, \mathbb{R}^{n_{j_1}}; \mathbb{R}^m),$$

when it is defined, is an ***iterated partial derivative*** of f at x_0. The number $r \in \mathbb{Z}_{>0}$ is the ***degree*** of the iterated partial derivative. •

Let us relate the rth derivative of f to the iterated partial derivatives of degree r. To do so we generalise the relationship in the case of $r = 1$ given in Theorem 1.4.19. This requires that we represent elements of $(\mathbb{R}^{n_1} \oplus \cdots \oplus \mathbb{R}^{n_k})^r$ is an appropriate way. A vector in $\mathbb{R}^{n_1} \oplus \cdots \oplus \mathbb{R}^{n_k}$ we write as $(v_1, \dots, v_k)$ for $v_j \in \mathbb{R}^{n_j}$, $j \in \{1, \dots, k\}$. Thus we write an element of $(\mathbb{R}^{n_1} \oplus \cdots \oplus \mathbb{R}^{n_k})^r$ as

$$((v_{r1}, \dots, v_{rk}), \dots, (v_{11}, \dots, v_{1k}))$$

for $v_{aj} \in \mathbb{R}^{n_j}$, $a \in \{1, \dots, r\}$, $j \in \{1, \dots, k\}$. Note the ordering with respect to the first index: we list the vectors from r to 1, not from 1 to r. This is to be consistent with our ordering of indices for iterated partial derivatives from r to 1 as we go from left to right.

We now have the following generalisation of Theorem 1.4.19.

1.4.28 Theorem (Iterated partial derivatives and higher-order derivatives) *If $\mathrm{U} \subseteq \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ is an open set and if $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^m$ is a map that is r times differentiable at $\mathbf{x}_0 \in \mathrm{U}$, then all iterated partial derivatives of $\mathbf{f}$ degree r are defined $\mathbf{x}_0$. Moreover, if $\mathbf{f}$ is r times differentiable at $\mathbf{x}_0$ then we have the following relationships between the derivative and the partial derivatives:*

(i) for $((\mathbf{v}_{r1}, \dots, \mathbf{v}_{rk}), \dots, (\mathbf{v}_{11}, \dots, \mathbf{v}_{1k})) \in (\mathbb{R}^{n_1} \oplus \cdots \oplus \mathbb{R}^{n_k})^r$ we have

$$\mathbf{D}^r\mathbf{f}(\mathbf{x}) \cdot ((\mathbf{v}_{r1}, \dots, \mathbf{v}_{1k}), \dots, (\mathbf{v}_{11}, \dots, \mathbf{v}_{1k})) = \sum_{j_1, \dots, j_r = 1}^{k} \mathbf{D}_{j_r} \cdots \mathbf{D}_{j_1}\mathbf{f}(\mathbf{x}) \cdot (\mathbf{v}_{rj_r}, \dots, \mathbf{v}_{1j_1}); \quad (1.25)$$

(ii) for $j_1, \dots, j_r \in \{1, \dots, k\}$ and $(\mathbf{v}_r, \dots, \mathbf{v}_1) \in \mathbb{R}^{n_{j_r}} \oplus \cdots \oplus \mathbb{R}^{n_{j_1}}$ we have

$$\mathbf{D}_{j_r} \cdots \mathbf{D}_{j_1}\mathbf{f}(\mathbf{x}) \cdot (\mathbf{v}_r, \dots, \mathbf{v}_1) = \mathbf{D}^r\mathbf{f}(\mathbf{x}) \cdot (\underbrace{(\mathbf{0}, \dots, \mathbf{v}_r, \dots, \mathbf{0})}_{\mathbf{v}_r \textit{ in } j_r\textit{th slot}}, \dots, \underbrace{(\mathbf{0}, \dots, \mathbf{v}_1, \dots, \mathbf{0})}_{\mathbf{v}_1 \textit{ in } j_1\textit{st slot}}). \quad (1.26)$$

.

Proof We prove the first implication of the theorem by induction on r. We do this by simultaneously proving (1.26) by in the induction argument. For $r = 1$ the assertion and (1.26) is simply Theorem 1.4.19. So suppose the result true for $r \in \{1, \dots, s\}$ and suppose that $\boldsymbol{f}$ is $s+1$ times differentiable at $\boldsymbol{x}_0$. By the induction hypothesis we have that all iterated partial derivatives of degree s exist and satisfy

$$\boldsymbol{D}_{j_s} \cdots \boldsymbol{D}_{j_1} \boldsymbol{f}(\boldsymbol{x}) \cdot (v_s, \dots, v_1) = \boldsymbol{D}^s \boldsymbol{f}(\boldsymbol{x}) \cdot (\underbrace{(\boldsymbol{0}, \dots, v_s, \dots, \boldsymbol{0})}_{v_s \text{ in } j_s\text{th slot}}, \dots, \underbrace{(\boldsymbol{0}, \dots, v_1, \dots, \boldsymbol{0})}_{v_1 \text{ in } j_1\text{st slot}}).$$

By Proposition 1.4.7 and Theorem 1.4.19, differentiability of $\boldsymbol{D}^s\boldsymbol{f}$ at $\boldsymbol{x}_0$ implies that all iterated partial derivatives of degree $s+1$ exist at $\boldsymbol{x}_0$. To prove that (1.26) holds for $r = s+1$ we compute

$$\begin{aligned}
\boldsymbol{D}_{j_{s+1}} & \boldsymbol{D}_{j_s} \cdots \boldsymbol{D}_{j_1} \boldsymbol{f}(\boldsymbol{x}) \cdot (v_{s+1}, v_s, \dots, v_1) \\
&= \Big(\boldsymbol{D}_{j_{s+1}}(\boldsymbol{D}^s \boldsymbol{f}(\boldsymbol{x}) \cdot (\underbrace{(\boldsymbol{0}, \dots, v_s, \dots, \boldsymbol{0})}_{v_s \text{ in } j_s\text{th slot}}, \dots, \underbrace{(\boldsymbol{0}, \dots, v_1, \dots, \boldsymbol{0})}_{v_1 \text{ in } j_1\text{st slot}}))\Big) \cdot v_{s+1} \\
&= \boldsymbol{D}^{s+1} \boldsymbol{f}(\boldsymbol{x}) \cdot (\underbrace{(\boldsymbol{0}, \dots, v_{s+1}, \dots, \boldsymbol{0})}_{v_{s+1} \text{ in } j_{s+1}\text{st slot}}, \underbrace{(\boldsymbol{0}, \dots, v_s, \dots, \boldsymbol{0})}_{v_s \text{ in } j_s\text{th slot}}, \dots, \\
&\qquad \underbrace{(\boldsymbol{0}, \dots, v_1, \dots, \boldsymbol{0})}_{v_1 \text{ in } j_1\text{st slot}})
\end{aligned}$$

using the induction hypotheses and Theorem 1.4.19. This gives (1.26) for $r = s+1$.

Finally we need to show that (1.25) holds. We prove this also by induction on r. For $r = 1$ the formula holds by Theorem 1.4.19. Suppose, then, that (1.25) holds for $r = s$ and that $\boldsymbol{f}$ is $s+1$ times differentiable at $\boldsymbol{x}_0$. Using the fact that the formula holds for $r = s$, we compute

$$\begin{aligned}
\boldsymbol{D}^{s+1} & \boldsymbol{f}(\boldsymbol{x}) \cdot ((v_{(s+1)1}, \dots, v_{(s+1)k}), (v_{s1}, \dots, v_{sk}), \dots, (v_{11}, \dots, v_{1k})) \\
&= \sum_{j_{s+1}=1}^{k} \Big(\boldsymbol{D}_{j_{s+1}}(\boldsymbol{D}^s \boldsymbol{f} \cdot ((v_{s1}, \dots, v_{sk}), \dots, (v_{11}, \dots, v_{1k})))\Big) \cdot v_{(s+1)j_{s+1}} \\
&= \sum_{j_{s+1}=1}^{k} \Big(\boldsymbol{D}_{j_{s+1}}\Big(\sum_{j_1,\dots,j_s=1}^{k} \boldsymbol{D}_{j_s} \cdots \boldsymbol{D}_{j_1} \boldsymbol{f}(\boldsymbol{x}) \cdot (v_{sj_s}, \dots, v_{1j_1})\Big)\Big) \cdot v_{(s+1)j_{s+1}} \\
&= \sum_{j_1,\dots,j_s,j_{s+1}=1}^{k} \boldsymbol{D}_{j_{s+1}} \boldsymbol{D}_{j_s} \cdots \boldsymbol{D}_{j_1} \boldsymbol{f}(\boldsymbol{x}) \cdot (v_{(s+1)j_{s+1}}, v_{sj_s}, \dots, v_{1j_1}),
\end{aligned}$$

giving (1.25) for $r = s+1$. ■

Since the preceding theorem contains Theorem 1.4.19 as a special case, it follows that the converse does not hold. That is to say, the existence of iterated partial derivatives of degree r does not imply that $\boldsymbol{f}$ is r times differentiable. We refer to the discussion surrounding Theorem 1.4.19 for more details.

Just as Theorem 1.4.19 allowed us to give an explicit formula for the derivative in Theorem 1.4.22, we can use apply Theorem 1.4.28 to give an explicit formula for higher-order derivatives.

1.4.29 Definition (Iterated partial derivative) Let $U \subseteq \mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$ be open, let $f\colon U \to \mathbb{R}^m$, let $x_0 \in U$, and let $j_1, \dots, j_r \in \{1, \dots, n\}$. The multilinear map

$$D_{j_r} \cdots D_{j_1} f(x_0) \in \mathbb{R}^m$$

(noting that $L^r(\mathbb{R}; \mathbb{R}^m)$ is isomorphic to $\mathbb{R}^m$ by Exercise I-5.6.5) when it is defined, is an ***iterated partial derivative*** of f at x_0. The number $r \in \mathbb{Z}_{>0}$ is the ***degree*** of the iterated partial derivative. •

Now, an application of Proposition 1.4.17 and Theorem 1.4.28 gives the following result.

1.4.30 Theorem (Explicit formula for higher-order derivatives) *If* $U \subseteq \mathbb{R}^n$ *is open and if* $\mathbf{f}\colon U \to \mathbb{R}^m$ *is a map that is* r *times differentiable at* $\mathbf{x}_0$ *and is written as* $\mathbf{f} = f_1 \times \cdots \times f_m$, *then all iterated partial derivatives of degree* r *of components* $f_1, \dots, f_m\colon U \to \mathbb{R}$ *exist at* $\mathbf{x}_0$. *Furthermore, the components of* $\mathbf{D}^r\mathbf{f}(\mathbf{x}_0) \in L^r(\mathbb{R}^n; \mathbb{R}^m)$ *are defined by*

$$(\mathbf{D}^r\mathbf{f}(\mathbf{x}_0)(\mathbf{e}_{j_r}, \dots, \mathbf{e}_{j_1}))_a = \mathbf{D}_{j_r} \cdots \mathbf{D}_{j_1} f_a(\mathbf{x}_0),$$

for $j_1, \dots, j_r \in \{1, \dots, n\}$ *and* $a \in \{1, \dots, m\}$.

In terms of more commonly used notation, the components of $D^r f(x_0)$ are written as

$$\frac{\partial^r f_a}{\partial x_{j_r} \cdots \partial x_{j_1}}(x_0), \qquad j_1, \dots, j_r \in \{1, \dots, n\},\ a \in \{1, \dots, m\}.$$

The following theorem generalises Theorem 1.4.25 and shows that, as long as the iterated partial derivatives are continuous, one can assert higher-order continuous differentiability.

1.4.31 Theorem (Higher-order continuous differentiability and continuity of iterated partial derivatives) *Let* $U \subseteq \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ *be open, let* $\mathbf{f}\colon U \to \mathbb{R}^m$, *and let* $r \in \mathbb{Z}_{>0}$. *Then the following statements are equivalent:*

(i) $\mathbf{f}$ *is of class* C^r;

(ii) all iterated partial derivatives of $\mathbf{f}$ *of degree* r *exist and are continuous.*

Proof (i) $\Longrightarrow$ (ii) By Theorem 1.4.25, if f is of class C^r then $D_{j_1} f$ is of class C^{r-1} for every $j_1 \in \{1, \dots, k\}$. Inductively using Theorem 1.4.25, it then follows that $D_{j_r} \cdots D_{j_1} f$ is defined and continuous for every $j_1, \dots, j_r \in \{1, \dots, k\}$.

(ii) $\Longrightarrow$ (i) We prove this implication by induction on r. As part of the proof we shall prove, included in the induction, that (1.25) holds under the assumption that iterated partial derivatives of degree r exist. By Theorem 1.4.25 it holds that if all iterated partial derivatives of degree 1 (i.e., all partial derivatives) exist and are continuous then f is of class C^1. Moreover, we showed in the proof of Theorem 1.4.25 that (1.25) holds for $r = 1$. Suppose the implication and (1.25) are true for $r \in \{1, \dots, s\}$ and suppose that all iterated partial derivatives of degree $s+1$ exist and are continuous. By the induction hypothesis the map $x \mapsto D^s f(x)$ is defined and continuous. Moreover, the assumption that all iterated derivatives of degree $s+1$ exist and are continuous implies, by Proposition 1.4.7 and (1.25) with $r = s$, that all partial derivatives of $D^s f$ exist and are continuous. Thus, by Theorem 1.4.25, $D^s f$ is continuously differentiable and so f is of class C^{s+1}. The proof that (1.25) holds for $r = s+1$ is then carried out just as in the proof of Theorem 1.4.28. ■

Next we discuss an important idea, that of commutativity of iterated partial derivatives. That is, we consider an open subset $U \subseteq \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ and a map $f: U \to \mathbb{R}^m$ for which the iterated partial derivatives $D_{j_1} D_{j_2} f$ and $D_{j_2} D_{j_1} f$ exist at $x_0 \in U$ for some $j_1, j_2 \in \{1, \dots, k\}$. The question is, "When are these iterated partial derivatives equal?" Clearly they cannot be equal when $n_{j_1} \neq n_{j_2}$ since $D_{j_1} D_{j_2} f \in L(\mathbb{R}^{n_{j_1}}, \mathbb{R}^{n_{j_2}}; \mathbb{R}^m)$ and $D_{j_2} D_{j_1} f \in L(\mathbb{R}^{n_{j_2}}, \mathbb{R}^{n_{j_1}}; \mathbb{R}^m)$. Even when $n_{j_1} = n_{j_2}$ they are not generally equal.

1.4.32 Example (Partial derivatives do not generally commute) Let $\mathsf{B} \in \bigwedge^2(\mathbb{R}^n; \mathbb{R}^m)$; that is, B is a skew-symmetric bilinear map from $\mathbb{R}^n \times \mathbb{R}^n$ to $\mathbb{R}^m$. Let us define $f_B: \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^m$ by $f_{\mathsf{B}}(x_1, x_2) = \mathsf{B}(x_1, x_2)$. By Theorem 1.4.8 we have

$$\begin{aligned} D_1 f_{\mathsf{B}}(x_1, x_2) \cdot v &= \mathsf{B}(v, x_2) \\ D_2 f_{\mathsf{B}}(x_1, x_2) \cdot v &= \mathsf{B}(x_1, v) \\ D_1 D_1 f_{\mathsf{B}}(x_1, x_2) \cdot (v_1, v_2) &= \mathbf{0} \\ D_1 D_2 f_{\mathsf{B}}(x_1, x_2) \cdot (v_1, v_2) &= \mathsf{B}(v_1, v_2) \\ D_2 D_1 f_{\mathsf{B}}(x_1, x_2) \cdot (v_1, v_2) &= \mathsf{B}(v_2, v_1) \\ D_2 D_2 f_{\mathsf{B}}(x_1, x_2) \cdot (v_1, v_2) &= \mathbf{0} \end{aligned}$$

for all $x_1, x_2, v, v_1, v_2 \in \mathbb{R}^n$. Since B is skew-symmetric, we have

$$D_1 D_2 f_{\mathsf{B}}(x_1, x_2) = D_2 D_1 f_{\mathsf{B}}(x_1, x_2)$$

if and only if $\mathsf{B} = 0$ (why?). Since the only case when B *must* be zero is when $n = 1$ (why?), we conclude that there are lots of possible choices for B when $n \geq 2$ for which the partial derivatives do not commute. •

The preceding example showing that partial derivative do not generally commute is not deep. However, it does help to provide a context as to why, when $n_{j_1} = n_{j_2} = 1$ it follows that partial derivatives do, indeed, commute. In particular, we hope that this suggests that the commuting of partial derivatives in this case is somewhat deep.

1.4.33 Theorem (One-dimensional partial derivatives commute) *Let* $U \subseteq \mathbb{R}^{n_1} \times \cdots \times \mathbb{R}^{n_k}$ *be open, let* $\mathbf{f}: U \to \mathbb{R}^m$ *be of class* C^2*, and let* $j_1, j_2 \in \{1, \dots, k\}$ *have the property that* $n_{j_1} = n_{j_2} = 1$*. Then*

$$\mathbf{D}_{j_1} \mathbf{D}_{j_2} \mathbf{f}(\mathbf{x}) = \mathbf{D}_{j_2} \mathbf{D}_{j_1} \mathbf{f}(\mathbf{x})$$

for all $\mathbf{x} \in U$.

Proof By Proposition 1.4.17 we can assume that $m = 1$ without loss of generality. We thus denote "$\boldsymbol{f}$" by "f." Let us write a point in U as

$$(\boldsymbol{x}_1, \dots, x_{j_1}, \dots, x_{j_2}, \dots, \boldsymbol{x}_k) \in \mathbb{R}^{n_1} \times \dots \times \mathbb{R} \times \dots \times \mathbb{R} \times \dots \times \mathbb{R}^{n_k}.$$

For each $j \in \{1, \dots, k\}$ choose $\boldsymbol{x}_{0j} \in \mathbb{R}^{n_j}$ so that

$$\boldsymbol{x}_0 \triangleq (\boldsymbol{x}_{01}, \dots, \boldsymbol{x}_{0k}) \in U.$$

If f is of class C^2 then the map

$$g \colon (s_1, s_2) \mapsto f(\boldsymbol{x}_{01}, \dots, x_{j_1 0} + s_1, \dots, x_{j_2 0} + s_2, \dots, \boldsymbol{x}_{0k})$$

is of class C^2 in a neighbourhood of $(0, 0) \in \mathbb{R}^2$. Moreover, by definition of the partial derivatives,

$$\boldsymbol{D}_1\boldsymbol{D}_2 g(0,0) = \boldsymbol{D}_{j_1}\boldsymbol{D}_{j_2} f(\boldsymbol{x}_0), \quad \boldsymbol{D}_2\boldsymbol{D}_1 g(0,0) = \boldsymbol{D}_{j_2}\boldsymbol{D}_{j_1} f(\boldsymbol{x}_0).$$

Thus is suffices to show that $\boldsymbol{D}_1\boldsymbol{D}_2 g(0,0) = \boldsymbol{D}_2\boldsymbol{D}_1 g(0,0)$.

For (s_1, s_2) in a neighbourhood of $(0, 0)$ define

$$D(s_1, s_2) = g(s_1, s_2) - g(s_1, 0) - g(0, s_2) + g(0, 0).$$

For fixed s_2 define $g_{s_2}(s_1) = g(s_1, s_2) - g(s_1, 0)$ so that $D(s_1, s_2) = g_{s_2}(s_1) - g_{s_2}(0)$. By the Mean Value Theorem, Theorem I-3.2.19, we have

$$D(s_1, s_2) = g_{s_2}(s_1) - g_{s_2}(0) = s_1 g'_{s_2}(\tilde{s}_1) = s_1(\boldsymbol{D}_1 g(\tilde{s}_1, s_2) - \boldsymbol{D}_1 g(\tilde{s}_1, 0))$$

for some $\tilde{s}_1 \in [0, s_1]$. Now we apply the Mean Value Theorem again to the function $s_2 \mapsto \boldsymbol{D}_1 g(\tilde{s}_1, s_2)$ to get

$$\boldsymbol{D}_1 g(\tilde{s}_1, s_2) - \boldsymbol{D}_1 g(\tilde{s}_1, 0) = s_2 \boldsymbol{D}_2\boldsymbol{D}_1 g(\tilde{s}_1, \tilde{s}_2).$$

Putting the preceding two formulae together we get

$$\boldsymbol{D}_2\boldsymbol{D}_1 g(\tilde{s}_1, \tilde{s}_2) = \frac{D(s_1, s_2)}{s_1 s_2}.$$

Continuity of the iterated partial derivatives of length two gives

$$\boldsymbol{D}_2\boldsymbol{D}_1 g(0,0) = \lim_{(s_1,s_2)\to(0,0)} \frac{D(s_1, s_2)}{s_1 s_2}$$

The above construction can be repeated, swapping the rôles of s_1 and s_2, to give

$$\boldsymbol{D}_1\boldsymbol{D}_2 g(0,0) = \lim_{(s_1,s_2)\to(0,0)} \frac{D(s_1, s_2)}{s_1 s_2},$$

giving the result. ■

Let us give a few examples to illuminate this important theorem.

1.4.34 Examples (Commutativity of one-dimensional partial derivatives)

1.

1.4.6 The derivative and function behaviour

Why is the derivative and differentiability important? Of course, this is an important question, and in this section we give some simple results that indicate why one might study the derivative of a map. Somewhat more profound illustrations of this are given in Section 1.5.

As in the single-variable case, differentiability implies continuity.

1.4.35 Proposition (Differentiable maps are continuous) *If* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *is an open set and if* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$ *is differentiable at* $\mathbf{x}_0 \in \mathrm{U}$*, then there exists* $\mathrm{M} \in \mathbb{R}_{>0}$ *and a neighbourhood* $\mathrm{V} \subseteq \mathrm{U}$ *of* $\mathbf{x}_0$ *such that*

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)\|_{\mathbb{R}^{\mathrm{m}}} \le \mathrm{M}\|\mathbf{x} - \mathbf{x}_0\|_{\mathbb{R}^{\mathrm{n}}}, \qquad \mathbf{x} \in \mathrm{V}.$$

In particular, $\mathbf{f}$ *is continuous at* $\mathbf{x}_0$*.*

Proof By definition of "differentiable at x_0" there exists a neighbourhood V of x_0 such that

$$\frac{\|f(x) - f(x_0) - Df(x_0)\cdot(x - x_0)\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}} < 1$$
$$\implies \quad \|f(x) - f(x_0) - Df(x_0)\cdot(x - x_0)\|_{\mathbb{R}^m} < \|x - x_0\|_{\mathbb{R}^n}$$

for $x \in V$. By Proposition 1.1.13 we have

$$\|Df(x_0)\cdot v\|_{\mathbb{R}^m} \le \|Df(x_0)\|_{\mathbb{R}^n,\mathbb{R}^m}\|v\|_{\mathbb{R}^n}$$

for all $v \in \mathbb{R}^n$. Thus the triangle inequality gives

$$\begin{aligned}\|f(x) - f(x_0)\|_{\mathbb{R}^m} &\le \|f(x) - f(x_0) - Df(x_0)\cdot(x - x_0)\|_{\mathbb{R}^m} \\ &\quad + \|Df(x_0)\|_{\mathbb{R}^n,\mathbb{R}^m}\|x - x_0\|_{\mathbb{R}^n} \\ &\le \|x - x_0\|_{\mathbb{R}^n} + \|Df(x_0)\|_{\mathbb{R}^n,\mathbb{R}^m}\|x - x_0\|_{\mathbb{R}^n}\end{aligned}$$

for all $x \in V$, giving the first assertion of the result if we take $M = 1 + \|Df(x_0)\|_{,\mathbb{R}^n}\mathbb{R}^m$.

For the final assertion, let $\epsilon \in \mathbb{R}_{>0}$ and let $\delta' \in \mathbb{R}_{>0}$ be such that $\mathsf{B}(\delta', x_0) \subseteq V$. Taking $\delta = \min\{\delta', \frac{\epsilon}{M}\}$ and letting $x \in \mathsf{B}(\delta, x_0)$ gives

$$\|f(x) - f(x_0)\|_{\mathbb{R}^m} \le M\|x - x_0\|_{\mathbb{R}^n} < \epsilon,$$

giving continuity of f at x_0. ■

If the derivative of the function is bounded, then one can infer uniform continuity.

1.4.36 Proposition (Functions with bounded derivatives are sometimes uniformly continuous) *If* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *is open and if* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$ *is continuously differentiable, then the following two statements hold:*

(i) if U *is convex (see the comments before the statement of Theorem 1.4.38 below) and if* $\mathbf{Df}$ *is bounded, then* $\mathbf{f}$ *is uniformly continuous;*

(ii) if $\mathrm{K} \subseteq \mathrm{U}$ *is compact, then* $\mathbf{f}|\mathrm{K}$ *is uniformly continuous.*

Proof (i) From the Mean Value Theorem, Theorem I-3.2.19 below. there exists $M \in \mathbb{R}_{>0}$ such that

$$\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{y})\|_{\mathbb{R}^m} \leq M\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n}$$

for every $\boldsymbol{x}, \boldsymbol{y} \in U$. Now let $\epsilon \in \mathbb{R}_{>0}$ and let $\boldsymbol{x} \in U$. Define $\delta = \frac{\epsilon}{M}$ and note that if $\boldsymbol{y} \in U$ satisfies $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} < \delta$ then we have

$$\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{y})\|_{\mathbb{R}^m} < \epsilon,$$

giving the desired uniform continuity.

(ii) Let

$$A = \sup\{\|\boldsymbol{f}\|_{\mathbb{R}^m}(\boldsymbol{x}) \mid \boldsymbol{x} \in K\},$$
$$B = \sup\{\|\boldsymbol{Df}(\boldsymbol{x})\|_{,|}\ \mathbb{R}^n\}\mathbb{R}^m \boldsymbol{x} \in K,$$

noting that $A, B < \infty$ by Theorem 1.3.31. Let $\boldsymbol{x} \in K$ and let $r_{\boldsymbol{x}} \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^n(2r_{\boldsymbol{x}}, \boldsymbol{x}) \subseteq U$. For $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathsf{B}^n(r_{\boldsymbol{x}}, \boldsymbol{x})$, the Mean Value Theorem gives

$$\|\boldsymbol{f}(\boldsymbol{y}_1) - \boldsymbol{f}(\boldsymbol{y}_2)\|_{\mathbb{R}^m} \leq B\|\boldsymbol{y}_1 - \boldsymbol{y}_2\|_{\mathbb{R}^n}.$$

Since $(\mathsf{B}^n(r_{\boldsymbol{x}}, \boldsymbol{x}))_{\boldsymbol{x} \in K}$ covers K, there exists $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in K$ such that $K \subseteq \cup_{j=1}^k \mathsf{B}^n(r_{\boldsymbol{x}_j}, \boldsymbol{x}_j)$. Let us abbreviate $N_j = \mathsf{B}^n(r_{\boldsymbol{x}_j}, \boldsymbol{x}_j)$ for $j \in \{1, \dots, k\}$. By Theorem 1.2.38 there exists $r \in \mathbb{R}_{>0}$ such that if $\boldsymbol{x}, \boldsymbol{y} \in K$ satisfy $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} < r$ then $\boldsymbol{x}, \boldsymbol{y} \in N_j$ for some $j \in \{1, \dots, k\}$.

We let $\boldsymbol{x}, \boldsymbol{y} \in K$. If $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} < r$ then $\boldsymbol{x}, \boldsymbol{y} \in N_j$ for some $j \in \{1, \dots, k\}$ and so

$$\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{y})\|_{\mathbb{R}^m} \leq B\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n}.$$

If $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} \geq r$ then

$$\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{y})\|_{\mathbb{R}^m} \leq \|\boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^m} + \|\boldsymbol{f}(\boldsymbol{y})\|_{\mathbb{R}^m} \leq 2A = \frac{2Ar}{r} \leq 2r^{-1}A\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n}.$$

Taking $M = \max\{B, 2r^{-1}A\}$, we then have

$$\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{y})\|_{\mathbb{R}^m} \leq M\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in K$. Uniform continuity of $\boldsymbol{f}$ follows as in the proof of the first part of the result. ■

The two conditions in the preceding result are generally necessary, as the following example shows.

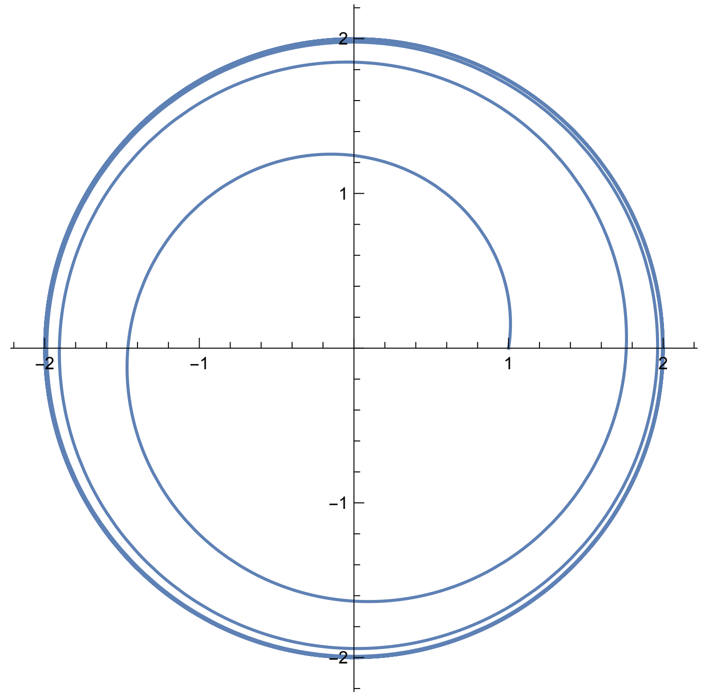

Figure 1.10 A spiral curve

1.4.37 Example (A function with a bounded derivative that is not uniformly continuous) Consider the curve $\gamma\colon (1,\infty) \to \mathbb{R}^2$ defined by

$$\gamma(t) = (1 + \tanh(t-1))(\cos(2\pi t), \sin(2\pi t)).$$

In Figure 1.10 we depict the traces of this curve, which is a spiral whose radius grows from a radius of 1 to a limiting radius of 2. Define

$$\begin{aligned}\phi\colon (-\tfrac{1}{8}, \tfrac{1}{8}) \times (1,\infty) &\to \mathbb{R}^2\\ (s,t) &\mapsto (1 + \tanh(t+s-1)(\cos(2\pi t), \sin(2\pi t)).\end{aligned}$$

Let us verify some of the elementary properties of this map.

1 Lemma *The map ϕ has the following properties:*

(i) it is injective;

(ii) it is continuously differentiable and there exists $\mathrm{c} \in \mathbb{R}_{>0}$ such that $\|\mathbf{D}\phi(\mathrm{s},\mathrm{t})\|_{\mathbb{R}^2,\mathbb{R}^2} \geq \mathrm{c}$ for all $(\mathrm{s},\mathrm{t}) \in (-\frac{1}{8}, \frac{1}{8}) \times (1,\infty)$;

(iii) ϕ^{-1} is continuously differentiable with bounded derivative.

Proof (i) Suppose that $\phi(s_1,t_1) = \phi(s_2,t_2)$, and without loss of generality take $t_2 \geq t_1$. Since the two image points must lie on the same ray through the origin we must have

$$(\cos(2\pi t_1), \sin(2\pi t_1)) = (\cos(2\pi t_2), \sin(2\pi t_2)),$$

implying that $t_2 - t_1 \in \mathbb{Z}_{\geq 0}$. If $t_1 = t_2$ then we must immediately have $1 + \tanh(t_1 + s_1 - 1) = 1 + \tanh(t_1 + s_2 - 1)$ giving $s_1 = s_2$ since tanh is injective (see Exercise I-3.8.5). So suppose that $t_2 - t_1 = k \in \mathbb{Z}_{>0}$. Then we must have

$$\begin{aligned} & 1 + \tanh(t_1 + s_1 - 1) = 1 + \tanh(t_1 + k + s_2 - 1) \\ \Longrightarrow \quad & t_1 + s_1 - 1 = s_1 + k + s_2 - 1 \\ \Longrightarrow \quad & s_1 - s_2 = k. \end{aligned}$$

again using injectivity of tanh. However, since $s_1, s_2 \in (\frac{1}{8}, \frac{1}{8})$ we have

$$|s_1 - s_2| < \tfrac{1}{4} \neq k,$$

and so we conclude that we must have $t_2 = t_1$ and so $s_2 = s_1$. This gives the desired injectivity of $\boldsymbol{\phi}$.

(ii) We directly compute

$$\|\boldsymbol{D\phi}(s, t)\|_{\mathbb{R}^2,\mathbb{R}^2} = 2(\cosh(t + s - 1)^(- 4) + 2\pi^2(1 - \tanh(t + s - 1))^2).$$

Note that, for (s, t) in the domain of $\boldsymbol{\phi}$, we have

$$\tanh(t + s - 1) \geq \tanh(-\tfrac{1}{8}) = -\tanh(\tfrac{1}{8}) \in (-1, 0).$$

Thus

$$\|\boldsymbol{D\phi}(s, t)\|_{\mathbb{R}^2,\mathbb{R}^2} \geq 4\pi^2(1 + \tanh(\tfrac{1}{8})^2 > 0,$$

giving this part of the lemma.

(iii) This follows from the Inverse Function Theorem, Theorem 1.5.2 below. ▼

Now we let $U = \operatorname{image}(\boldsymbol{\phi})$, noting that U is a "thickening" of the trace from Figure 1.10. By U is open. Next define what

$$\begin{aligned} g\colon (-\tfrac{1}{8}, \tfrac{1}{2}) \times (1, \infty) &\to \mathbb{R} \\ (s, t) &\mapsto t \end{aligned}$$

and we note that clearly g is continuously differentiable with a bounded derivative. If we define $f\colon U \to \mathbb{R}$ by $f = g \circ \boldsymbol{\phi}^{-1}$, then, by the Chain Rule and Proposition 1.1.16(vi), f is continuously differentiable with a bounded derivative. It remains to show that f is not uniformly continuous.

For $k \in \mathbb{Z}$ with $k \geq 2$ note that $\boldsymbol{x}_k \triangleq (1 + \tanh(k - 1)(1, 0) \in U$ and that $f(\boldsymbol{x}_k) = k$. Let $\delta \in \mathbb{R}_{>0}$. Since $\lim_{k\to\infty}(1 + \tanh(k - 1)) = 2$ let $N \in \mathbb{Z}_{>0}$ be sufficiently large that

$$|(1 + \tanh(j - 1)) - (1 + \tanh(k - 1))| < \delta, \qquad j, k \geq N.$$

Then let $k \in \mathbb{Z}_{>0}$ and note that

$$|f(\boldsymbol{x}_{N+k}) - f(\boldsymbol{x}_N)| = k.$$

Note that

$$\|\boldsymbol{x}_{N+k} - \boldsymbol{x}_N\|_{\mathbb{R}^2} = |(1 + \tanh(N + k - 1)) - (1 + \tanh(N - 1))| < \delta.$$

Therefore, for any $\delta \in \mathbb{R}_{>0}$, there are points $\boldsymbol{x}, \boldsymbol{y} \in U$ such that $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^2} < \delta$ but $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \geq 1$. This prohibits uniform continuity of f. •

As we showed to dramatic effect in Example I-3.2.9, it is very much not the case that a continuous function is differentiable.

Next we consider the multivariable version of the Mean Value Theorem that we stated in the single-variable case as Theorem I-3.2.19. The fact that the natural domain for functions in the single-variable case is an interval needs to be appropriately generalised to the multivariable case. A natural way to do this is with the notion of a convex set. We shall investigate convexity in some detail in Section 1.9, so let us just recall the basic definition here. For $x_1, x_2 \in \mathbb{R}^n$ denote by $\{(1-s)x_1 + sx_2 \mid s \in [0,1]\}$ the line segment between x_1 and x_2. A subset $C \subseteq \mathbb{R}^n$ is ***convex*** if the line segment between any two points in C is a subset of C.

1.4.38 Theorem (Mean Value Theorem) *Let $C \subseteq \mathbb{R}^n$ be an open convex set and let $\mathbf{f}\colon C \to \mathbb{R}^m$ be of class C^1. If $\mathbf{x}_1, \mathbf{x}_2 \in C$ then*

$$\|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\|_{\mathbb{R}^m} \leq \sup\{\|\mathbf{Df}((1-s)\mathbf{x}_1 + s\mathbf{x}_2)\|_{\mathbb{R}^n,\mathbb{R}^m} \mid s \in [0,1]\}\|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbb{R}^n}.$$

Moreover, if $\mathbf{Df}$ is uniformly bounded, i.e., if there exists $M \in \mathbb{R}_{>0}$ such that $\|\mathbf{Df}(\mathbf{x})\|_{\mathbb{R}^n,\mathbb{R}^m} \leq M$ for every $\mathbf{x} \in C$, then

$$\|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\|_{\mathbb{R}^m} \leq M\|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbb{R}^n}.$$

Proof Let $\gamma\colon [0,1] \to \mathbb{R}^n$ be defined by $\gamma(s) = (1-s)x_1 + sx_2$. Then $\text{image}(\gamma) \subseteq C$ since C is convex. By the Chain Rule, Theorem 1.4.49, we have

$$\boldsymbol{D}(f \circ \gamma)(s) = \boldsymbol{D}f(\gamma(s)) \circ \boldsymbol{D}\gamma(s).$$

Using the Fundamental Theorem of Calculus applied to the components of the map $g = f \circ \gamma\colon C \to \mathbb{R}^m$ we have

$$g(1) - g(0) = \int_0^1 \boldsymbol{D}g(s)\,\mathrm{d}s,$$

which gives

$$f(x_1) - f(x_2) = \int_0^1 \boldsymbol{D}f((1-s)x_1 + sx_2)\cdot(x_2 - x_1)\,\mathrm{d}s.$$

integral triangle ineq Thus, using Proposition 1.1.16(v) and ,

$$\begin{aligned}\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} &= \Big\|\int_0^1 \boldsymbol{D}f((1-s)x_1 + sx_2)\cdot(x_2 - x_1)\,\mathrm{d}s\Big\|_{\mathbb{R}^m} \\ &\leq \int_0^1 \|\boldsymbol{D}f((1-s)x_1 + sx_2)\cdot(x_2 - x_1)\|_{\mathbb{R}^m}\,\mathrm{d}s \\ &\leq \Big(\int_0^1 \|\boldsymbol{D}f((1-s)x_1 + sx_2)\|_{\mathbb{R}^n,\mathbb{R}^m}\,\mathrm{d}s\Big)\cdot\|x_1 - x_2\|_{\mathbb{R}^m} \\ &\leq \sup\{\|\boldsymbol{D}f((1-s)x_1 + sx_2)\|_{\mathbb{R}^n,\mathbb{R}^m} \mid s \in [0,1]\}\|x_1 - x_2\|_{\mathbb{R}^m},\end{aligned}$$

as desired.

The final assertion of the theorem follows immediately from the first. ■

1.4.7 Derivatives and maxima and minima

Next we generalise to multiple-dimensions the relationships between derivatives and maxima and minima of functions. First let us define the relevant function properties.

1.4.39 Definition (Local maximum and local minimum) Let $A \subseteq \mathbb{R}^n$ and let $A\colon I \to \mathbb{R}$ be a function. A point $x_0 \in A$ is a:

(i) ***local maximum*** if there exists a neighbourhood U of x_0 such that $f(\boldsymbol{x}) \leq f(x_0)$ for every $\boldsymbol{x} \in U \cap A$;

(ii) ***strict local maximum*** if there exists a neighbourhood U of x_0 such that $f(\boldsymbol{x}) < f(x_0)$ for every $\boldsymbol{x} \in U \cap (A \setminus \{x_0\})$;

(iii) ***local minimum*** if there exists a neighbourhood U of x_0 such that $f(\boldsymbol{x}) \geq f(x_0)$ for every $\boldsymbol{x} \in U \cap A$;

(iv) ***strict local minimum*** if there exists a neighbourhood U of x_0 such that $f(\boldsymbol{x}) > f(x_0)$ for every $\boldsymbol{x} \in U \cap (A \setminus \{x_0\})$. •

To generalise the single-variable characterisation of maxima and minima given in Theorem I-3.2.16 the reader will want to recall properties of symmetric bilinear maps from Section I-5.6.4.

1.4.40 Theorem (Derivatives, and maxima and minima) *If $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ is open, if $\mathrm{f}\colon \mathrm{U} \to \mathbb{R}$ is a function, and if $\mathbf{x}_0 \in \mathrm{U}$ then the following statements hold:*

(i) if f is differentiable at $\mathbf{x}_0$ and if $\mathbf{x}_0$ is a local maximum or a local minimum for f, then $\mathbf{D}\mathrm{f}(\mathbf{x}_0) = \mathbf{0}$;

(ii) if f is twice differentiable at $\mathbf{x}_0$, and if $\mathbf{x}_0$ is a local maximum (resp. local minimum) for f, then $\mathbf{D}^2\mathrm{f}(\mathbf{x}_0)$ is negative-semidefinite (resp. positive-semidefinite);

(iii) if f is twice differentiable at $\mathbf{x}_0$, and if $\mathbf{D}\mathrm{f}(\mathbf{x}_0) = \mathbf{0}$ and $\mathbf{D}^2\mathrm{f}(\mathbf{x}_0)$ is negative definite (resp. positive-definite), then $\mathbf{x}_0$ is a strict local maximum (resp. strict local minimum) for f;

(iv) if f is twice differentiable at $\mathbf{x}_0$, if $\mathbf{D}\mathrm{f}(\mathbf{x}_0) = \mathbf{0}$ and if $\mathbf{D}^2\mathrm{f}(\mathbf{x}_0)$ is neither positive- nor negative-semidefinite, then $\mathbf{x}_0$ is neither a local minimum nor a local maximum for f.

Proof (i) We shall give the proof for the case when x_0 is a local minimum; the case of a local maximum is similar. Let $v \in \mathbb{R}^n$. Since x_0 a local minimum we have

$$f(x_0 + sv) - f(x_0) \geq 0$$

for all s sufficiently near 0. Thus

$$\tfrac{1}{s}(f(\boldsymbol{x_0} + sv) - f(x_0)) \geq 0$$

for $s \in \mathbb{R}_{\geq 0}$ and so, by Proposition 1.4.12,

$$\boldsymbol{D}f(x_0) \cdot v = \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} f(x_0 + sv) = \lim_{s \downarrow 0} \tfrac{1}{s}(f(\boldsymbol{x_0} + sv) - f(x_0)) \geq 0.$$

Similarly, since

$$\frac{1}{s}(f(\boldsymbol{x_0}+s\boldsymbol{v})-f(\boldsymbol{x}_0))\leq 0$$

for $s\in\mathbb{R}_{\leq 0}$ we have $\boldsymbol{D}f(\boldsymbol{x}_0)\cdot\boldsymbol{v}\leq 0$ and so we conclude that $\boldsymbol{D}f(\boldsymbol{x}_0)\cdot\boldsymbol{v}=0$. Since this holds for any $\boldsymbol{v}\in\mathbb{R}^n$ we must have $\boldsymbol{D}f(\boldsymbol{x}_0)=\boldsymbol{0}$.

what?

where? and get the notation right

(ii) We prove the result for the case when $\boldsymbol{x}_0$ is a local minimum; the case of a local maximum is proved similarly. By the multivariable Taylor Theorem, , and noting the definition of the Landau symbol from , we have

$$0\leq f(\boldsymbol{x}_0+s\boldsymbol{v})-f(\boldsymbol{x}_0)=\tfrac{1}{2}s^2\boldsymbol{D}^2f(\boldsymbol{x}_0)\cdot(\boldsymbol{v},\boldsymbol{v})+o((s\boldsymbol{v})^2)$$

for every $\boldsymbol{v}\in\mathbb{R}^n$ and for s sufficiently near 0. Therefore,

$$\begin{aligned}&\boldsymbol{D}^2f(\boldsymbol{x}_0)\cdot(\boldsymbol{v},\boldsymbol{v})+\tfrac{2}{s^2}o((s\boldsymbol{v})^2)\geq 0\\ \Longrightarrow\quad&\boldsymbol{D}^2f(\boldsymbol{x}_0)\cdot(\boldsymbol{v},\boldsymbol{v})+\lim_{s\to 0}\tfrac{2}{s^2}o((s\boldsymbol{v})^2)\geq 0\\ \Longrightarrow\quad&\boldsymbol{D}^2f(\boldsymbol{x}_0)\cdot(\boldsymbol{v},\boldsymbol{v})\geq 0,\end{aligned}$$

giving $\boldsymbol{D}^2f(\boldsymbol{x}_0)$ as positive-semidefinite, as desired.

(iii) We first prove a lemma.

1 Lemma *If* $\mathsf{B}\in\mathsf{S}^2(\mathbb{R}^n;\mathbb{R})$ *is positive-definite then there exists* $m,M\in\mathbb{R}_{>0}$ *such that*

$$m\|\mathbf{v}\|^2_{\mathbb{R}^n}\leq\mathsf{B}(\mathbf{v},\mathbf{v})\leq M\|\mathbf{v}\|^2_{\mathbb{R}^n}$$

for every $\mathbf{v}\in\mathbb{R}^n$.

Proof Define $\boldsymbol{B}\in\mathrm{Mat}_{n\times n}(\mathbb{R})$ by $\boldsymbol{B}(i,j)=\mathsf{B}(\boldsymbol{e}_i,\boldsymbol{e}_j)$ so that

$$\mathsf{B}(\boldsymbol{v},\boldsymbol{v})=\sum_{i,j=1}^n\boldsymbol{B}(i,j)\boldsymbol{v}(i)\boldsymbol{v}(j).$$

Then $\boldsymbol{B}^T=\boldsymbol{B}$ and

$$\sum_{i,j=1}^n\boldsymbol{B}(i,j)\boldsymbol{v}(i)\boldsymbol{v}(j)>0$$

what

for every $\boldsymbol{v}\in\mathbb{R}^n$, cf. the proof of Theorem I-5.6.19. By there exists an orthogonal matrix $\boldsymbol{R}\in\mathsf{O}(n)$ such that

$$\boldsymbol{B}=\boldsymbol{R}^T\begin{bmatrix}d_1&0&\cdots&0\\0&d_2&\cdots&0\\\vdots&\vdots&\ddots&\vdots\\0&0&\cdots&d_n\end{bmatrix}\boldsymbol{R}$$

for $d_1,\dots,d_n\in\mathbb{R}_{>0}$. Therefore, for any $\boldsymbol{v}\in\mathbb{R}^n$, we have

$$\sum_{i,j=1}^n\boldsymbol{B}(i,j)\boldsymbol{v}(i)\boldsymbol{v}(j)=\sum_{j=1}^n d_j(\boldsymbol{R}\boldsymbol{v})(j)^2=\sum_{j=1}^n d_j\boldsymbol{v}(j)^2.$$

Therefore, we directly have

$$\min\{d_1,\dots,d_n\}\|\boldsymbol{v}\|^2_{\mathbb{R}^n}\leq\mathsf{B}(\boldsymbol{v},\boldsymbol{v})\leq\max\{d_1,\dots,d_n\}\|\boldsymbol{v}\|^2_{\mathbb{R}^n}$$

for every $\boldsymbol{v}\in\mathbb{R}^n$, giving the result. ▼

We now prove this part of the theorem for the case when $D^2f(x_0)$ is positive-definite; the case when it is negative-definite follows in the same manner with a suitable trivial modification to the signs of m and M in the lemma above.

From the lemma there exists $m \in \mathbb{R}_{>0}$ such that $D^2f(x_0) \cdot (v, v) \geq m\|v\|^2_{\mathbb{R}^n}$ for every $v \in \mathbb{R}^n$. Therefore, by the multivariable Taylor Theorem, , we have what?

$$f(x_0 + v) - f(x_0) = \tfrac{1}{2}D^2f(x_0) \cdot (v, v) + o(v^2) \geq \tfrac{1}{2}m\|v\|^2_{\mathbb{R}^n} + o(v^2),$$

for v sufficiently small in norm that $x_0 + v \in U$. Now choose $\epsilon \in \mathbb{R}_{>0}$ sufficiently small that $|o(v^2)| \leq \frac{1}{4}m\|v\|^2_{\mathbb{R}^n}$ for $v \in \overline{\mathsf{B}}^n(\epsilon, \mathbf{0})$. Then

$$f(x_0 + v) - f(x_0) \geq \tfrac{1}{4}m\|v\|^2_{\mathbb{R}^n},$$

for all $v \in \overline{\mathsf{B}}^n(\epsilon, \mathbf{0})$, giving x_0 as a strict local minimum for f.

(iv) Since $D^2f(x_0)$ is neither positive- nor negative-semidefinite, there exists $v_-, v_+ \in \mathbb{R}^n$ such that

$$D^2f(x_0) \cdot (v, v_-) \in \mathbb{R}_{<0}, \quad D^2f(x_0) \cdot (v+, v_+) \in \mathbb{R}_{>0}.$$

As above, write

$$f(x_0 + sv_-) - f(x_0) = \tfrac{1}{2}s^2D^2f(x_0) \cdot (v_-, v_-) + o((sv_-)^2).$$

for $s \in \mathbb{R}_{>0}$ be sufficiently small that $x_0 + sv_- \in U$. Further choosing s_0 sufficiently small that

$$\left|\frac{o((sv_-)^2)}{s^2}\right| < \frac{1}{4}Df(x_0) \cdot (v_-, v_-)$$

for $s \in (0, s_0]$, we have

$$\begin{aligned}\tfrac{1}{2}s^2D^2f(x_0) \cdot (v_-, v_-) + o((sv_-)^2) &= s^2\Big(\frac{1}{2}D^2f(x_0) \cdot (v_-, v_-) + \frac{o((sv_-)^2)}{s^2}\Big)\\ &< \frac{s^2}{4}Df(x_0) \cdot (v_-, v_-) < 0,\end{aligned}$$

giving $f(x_0+sv_-) < f(x_0)$ for $s \in (0, s_0]$. In a similar manner, one shows that $f(x_0+sv_+) > f(x_0)$ for s sufficiently small. Thus x_0 is neither a local minimum nor a local minimum. ■

We refer to Example I-3.2.17 for illustrations of the above theorem in the single-variable case. The same conclusions concerning the lack of converses to the theorem hold as were drawn from Example I-3.2.17. It is, however, slightly insightful to give a few additional examples in multiple-variables.

1.4.41 Examples (Derivatives, and maxima and minima)

1. We define $f_\alpha : \mathbb{R}^2 \to \mathbb{R}$ by $f_\alpha(x_1, x_2) = x_1^2 + \alpha x_2^2$ for $\alpha \in \mathbb{R}$. We see that $(0, 0)$ is a local minimum (resp. strict local minimum) when $\alpha \in \mathbb{R}_{\geq 0}$ (resp. $\alpha \in \mathbb{R}_{>0}$). When $\alpha \in \mathbb{R}_{<0}$ we have that $(0, 0)$ is neither a local minimum nor a local maximum. We compute

$$Df_\alpha(0, 0) = 0, \quad D^2f_\alpha(0, 0) \cdot ((v_1, v_2), (v_1, v_2)) = 2v_1^2 + 2\alpha v_2^2.$$

Thus $\boldsymbol{D}^2 f_\alpha$ is positive-semidefinite when $\alpha = 0$, positive-definite when $\alpha \in \mathbb{R}_{>0}$, and indefinite when $\alpha \in \mathbb{R}_{<0}$. From Theorem 1.4.40 we see that $(0,0)$ is a strict local minimum for f_α when $\alpha \in \mathbb{R}_{>0}$. When $\alpha \in \mathbb{R}_{\le 0}$ we can only conclude that $(0,0)$ is not a local minimum for f_α.

2. We take $f_\alpha\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $f(x_1,x_2) = x_1^2 + \alpha x_2^2$ for $\alpha \in \mathbb{R}$. When $\alpha \in \mathbb{R}_{>0}$ we see that $(0,0)$ is a strict local minimum for f_α and that when $\alpha \in \mathbb{R}_{\ge 0}$ we have $(0,0)$ as a (not strict) local minimum. When $\alpha \in \mathbb{R}_{<0}$, $(0,0)$ is neither a local minimum nor a local maximum. We compute

$$\boldsymbol{D} f_\alpha(0,0) = 0, \quad \boldsymbol{D}^2 f_\alpha(0,0) \cdot ((v_1,v_2),(v_1,v_2)) = 2v_1^2.$$

Thus $\boldsymbol{D}^2 f_\alpha(0,0)$ is positive-semidefinite for every α. By Theorem 1.4.40 we can conclude that $(0,0)$ cannot be a local minimum for f_α for every α, and this is indeed the case. However, the conclusion that $(0,0)$ is a strict local minimum of f_α for $\alpha \in \mathbb{R}_{>0}$ cannot be deduced from Theorem 1.4.40.

3. Finally, we take $f_\alpha\colon \mathbb{R}^2 \to \mathbb{R}$ defined by $f_\alpha(x_1,x_2) = x_1^4 + \alpha x_2^4$. We see that when $\alpha \in \mathbb{R}_{>0}$ (resp. $\alpha \in \mathbb{R}_{\ge 0}$), $(0,0)$ is a strict local minimum (resp. local minimum) for f_α. For $\alpha \in \mathbb{R}_{<0}$ we have that $(0,0)$ is neither a local minimum nor a local maximum. Moreover, we compute $\boldsymbol{D}^2 f(0,0) \cdot ((v_1,v_2),(v_1,v_2)) = 0$ and so $\boldsymbol{D}^2 f(0,0)$ is both positive- and negative-semidefinite. No conclusions can be drawn using Theorem 1.4.40 to determine whether $(0,0)$ is a local maximum or minimum. •

1.4.8 Derivatives and constrained extrema

Let us next consider an important modification of the problem of finding minima and maxima, that where constraints are added to the mix. We wish to allow equality and inequality constraints, so let us set this up properly. Given $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, let us write $\boldsymbol{x} \le \boldsymbol{y}$ when $x_j \le y_j$ for each $j \in \{1,\dots,n\}$. With this convention, we make the following definition.

1.4.42 Definition (Equality and inequality constraints) Let $A \subseteq \mathbb{R}^n$ and let $\boldsymbol{g}\colon A \to \mathbb{R}^m$. A point $\boldsymbol{x} \in A$ satisfies the ***equality constraint*** defined by $\boldsymbol{g}$ if $\boldsymbol{g}(\boldsymbol{x}) = 0$ and satisfies the ***inequality constraint*** defined by $\boldsymbol{g}$ if $\boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0}$. •

Thus, with the notation of the definition, the set of points in A satisfying the equality constraint is $\boldsymbol{g}^{-1}(\boldsymbol{0})$ and the set of points satisfying the inequality constraint defined by $\boldsymbol{g}$ is $\boldsymbol{g}^{-1}(\mathbb{R}^m_{\le 0})$. We can now define the sorts of minima and maxima in which we are interested.

1.4.43 Definition (Constrained local maximum and minimum) Let $A \subseteq \mathbb{R}^n$ and consider maps $f\colon A \to \mathbb{R}$ and $\boldsymbol{g}\colon A \to \mathbb{R}^m$ and $\boldsymbol{h}\colon A \to \mathbb{R}^k$. A point $\boldsymbol{x}_0 \in \boldsymbol{g}^{-1}(\boldsymbol{0})$ is

(i) ***local maximum*** of the triple $(f, \boldsymbol{g}, \boldsymbol{h})$ if there exists a relative neighbourhood U of $\boldsymbol{x}_0$ in A such that $f(\boldsymbol{x}) \le f(\boldsymbol{x}_0)$ for every $\boldsymbol{x} \in \boldsymbol{g}^{-1}(\boldsymbol{0}) \cap \boldsymbol{h}^{-1}(\mathbb{R}^k_{\le 0}) \cap U$;

(ii) ***strict local maximum*** of the triple $(f, \boldsymbol{g}, \boldsymbol{h})$ if there exists a relative neighbourhood U of $\boldsymbol{x}_0$ in A such that $f(\boldsymbol{x}) < f(\boldsymbol{x}_0)$ for every $\boldsymbol{x} \in \boldsymbol{g}^{-1}(\boldsymbol{0}) \cap \boldsymbol{h}^{-1}(\mathbb{R}^k_{\leq 0}) \cap (U \setminus \{\boldsymbol{x}_0\})$;

(iii) ***local minimum*** of the triple $(f, \boldsymbol{g}, \boldsymbol{h})$ if there exists a relative neighbourhood U of $\boldsymbol{x}_0$ in A such that $f(\boldsymbol{x}) \geq f(\boldsymbol{x}_0)$ for every $\boldsymbol{x} \in \boldsymbol{g}^{-1}(\boldsymbol{0}) \cap \boldsymbol{h}^{-1}(\mathbb{R}^k_{\leq 0}) \cap U$;

(iv) ***strict local minimum*** of the triple $(f, \boldsymbol{g}, \boldsymbol{h})$ if there exists a relative neighbourhood U of $\boldsymbol{x}_0$ in A such that $f(\boldsymbol{x}) > f(\boldsymbol{x}_0)$ for every $\boldsymbol{x} \in \boldsymbol{g}^{-1}(\boldsymbol{0}) \cap \boldsymbol{h}^{-1}(\mathbb{R}^k_{\leq 0}) \cap (U \setminus \{\boldsymbol{x}_0\})$.

If there are no inequality constraints, we shall say that $\boldsymbol{x}_0$ is a local maximum (etc.) of $(f, \boldsymbol{g})$ with equality constraints. If there are no inequality constraints, we shall say that $\boldsymbol{x}_0$ is a local maximum (etc.) of $(f, \boldsymbol{h})$ with inequality constraints. •

The following theorem gives conditions for minimising $(f, \boldsymbol{g}, \boldsymbol{h})$ under hypotheses of differentiability.

1.4.44 Theorem (Lagrange Multiplier Rule) *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *be open, and let* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}$, $\mathbf{g}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$, *and* $\mathbf{h}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{k}}$ *be continuously differentiable. For* $\lambda_0 \in \mathbb{R}$, $\boldsymbol{\lambda} \in \mathbb{R}^{\mathrm{m}}$, *and* $\boldsymbol{\mu} \in \mathbb{R}^{\mathrm{k}}$, *define*

$$\begin{aligned} \mathrm{f}_{\lambda_0,\boldsymbol{\lambda},\boldsymbol{\mu}}\colon \mathrm{U} &\to \mathbb{R} \\ \mathbf{x} &\mapsto \lambda_0 \mathrm{f}(\mathbf{x}) + \langle \boldsymbol{\lambda}, \mathbf{g}(\mathbf{x}) \rangle_{\mathbb{R}^{\mathrm{m}}} + \langle \boldsymbol{\mu}, \mathbf{h}(\mathbf{x}) \rangle_{\mathbb{R}^{\mathrm{k}}} \end{aligned}.$$

If $\mathbf{x}_0$ *is a local minimum of* $(\mathrm{f}, \mathbf{g}, \mathbf{h})$, *then there exist* $\lambda_0 \in \mathbb{R}$, $\boldsymbol{\lambda} \in (\mathbb{R}^{\mathrm{m}})^*$, *and* $\boldsymbol{\mu} \in \mathbb{R}^{\mathrm{k}}$, *not simultaneously zero, such that* $\mathbf{D}\mathrm{f}_{\lambda_0,\boldsymbol{\lambda},\boldsymbol{\mu}}(\mathbf{x}_0) = \mathbf{0}$. *Furthermore, the following statements hold:*

(i) $\lambda_0 \in \mathbb{R}_{\geq 0}$ *and* $\boldsymbol{\mu} \geq \mathbf{0}$;

(ii) if, for $\mathrm{r} \in \{1, \dots, \mathrm{k}\}$, $\mathrm{h_r}(\mathbf{x}_0) < 0$, *then* $\mu_{\mathrm{r}} = 0$;

(iii) if the vectors satisfy the ***Kuhn–Tucker condition****, namely that*

$$\{\mathbf{D}\mathrm{g}_1(\mathbf{x}_0), \dots, \mathbf{D}\mathrm{g}_{\mathrm{m}}(\mathbf{x}_0)\} \cup \{\mathbf{D}\mathrm{h}_{\mathrm{r}}(\mathbf{x}_0) \mid \mathrm{h}_{\mathrm{r}}(\mathbf{x}_0)\}$$

are linearly independent, then λ_0 *can be taken to be* 1.

Proof We assume, without loss of generality, that $\boldsymbol{x}_0 = \boldsymbol{0}$, that $f(\boldsymbol{0}) = 0$, and that

$$h_1(\boldsymbol{0}) = \dots = h_s(\boldsymbol{0}) = 0, \; h_{s+1}(\boldsymbol{0}), \dots, h_k(\boldsymbol{0}) \in \mathbb{R}_{<0}.$$

It will be convenient to denote $a_+ = \max\{0, a\}$ for $a \in \mathbb{R}$.

Suppose that $\bar{\epsilon} \in \mathbb{R}_{>0}$ is such that $\overline{\mathsf{B}}^n(\bar{\epsilon}, \boldsymbol{0}) \subseteq U$ and such that $h_r(\boldsymbol{x}) < 0$ for every $\boldsymbol{x} \in \overline{\mathsf{B}}(\bar{\epsilon}, \boldsymbol{0})$, $r \in \{s+1, \dots, k\}$, the latter being possible since $\boldsymbol{h}$ is continuous. We prove a lemma.

1 Lemma *If* $\epsilon \in (0, \bar{\epsilon}]$, *then there exists* $\mathrm{M} \in \mathbb{R}_{>0}$ *such that*

$$\mathrm{f}(\mathbf{x}) + \|\mathbf{x}\|^2_{\mathbb{R}^{\mathrm{n}}} + \mathrm{M}\Big(\sum_{\mathrm{a}=1}^{\mathrm{m}} \mathrm{g_a}(\mathbf{x})^2 + \sum_{\mathrm{r}=1}^{\mathrm{s}} (\mathrm{h_r}(\mathbf{x})_+)^2\Big) \in \mathbb{R}_{>0}$$

for all $\mathbf{x}$ *such that* $\|\mathbf{x}\|_{\mathbb{R}^{\mathrm{n}}} = \epsilon$.

Proof Suppose the conclusions of the lemma do not hold. Then there exists a sequence $(M_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}_{>0}$ and a sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ such that (1) $\lim_{j\to\infty} M_j = \infty$, (2) $\|\boldsymbol{x}_j\|_{\mathbb{R}^n} = \epsilon$ for each $j \in \mathbb{Z}_{>0}$, and (3)

$$f(\boldsymbol{x}_j) + \|\boldsymbol{x}_j\|^2_{\mathbb{R}^n} \le -M_j\Big(\sum_{a=1}^{m} g_a(\boldsymbol{x}_j)^2 + \sum_{r=1}^{s}(h_r(\boldsymbol{x}_j)_+)^2\Big) \tag{1.27}$$

for each $j \in \mathbb{Z}_{>0}$. Note that the set of points

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_{\mathbb{R}^n} = \epsilon\}$$

is closed and bounded, and so compact. By the Bolzano–Weierstrass Theorem, we can assume that the sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $\bar{\boldsymbol{x}}$ such that $\|\bar{\boldsymbol{x}}\|_{\mathbb{R}^n} = \epsilon$. Since $\boldsymbol{g}$ is continuous and since the function $\boldsymbol{x} \mapsto h_r(\boldsymbol{x})_+$ is continuous, we have

$$\sum_{a=1}^{m} g_a(\bar{\boldsymbol{x}})^2 + \sum_{r=1}^{s}(h_r(\bar{\boldsymbol{x}})_+)^2 = \lim_{j\to\infty}\Big(\sum_{a=1}^{m} g_a(\boldsymbol{x}_j)^2 + \sum_{r=1}^{s}(h_r(\boldsymbol{x}_j)_+)^2\Big) = 0.$$

Thus $\boldsymbol{g}(\bar{\boldsymbol{x}}) = \boldsymbol{0}$ and $h_r(\bar{\boldsymbol{x}}) = 0$, $r \in \{1,\dots,s\}$. Then $\bar{\boldsymbol{x}}$ satisfies the equality constraints defined by $\boldsymbol{g}$ and the inequality constraints defined by $\boldsymbol{h}$. As such, since $\boldsymbol{0}$ is a local minimum of $(f,\boldsymbol{g},\boldsymbol{h})$, $f(\bar{\boldsymbol{x}}) \ge f(\boldsymbol{0}) = 0$. However, by (1.27), $f(\boldsymbol{x}_j) \le -\epsilon^0$ for each $j \in \mathbb{Z}_{>0}$, and so, by continuity of f,

$$f(\bar{\boldsymbol{x}}) = \lim_{j\to\infty} f(\boldsymbol{x}_j) \le -\epsilon^2,$$

giving a contradiction. ▼

Now another lemma.

2 Lemma *If $\epsilon \in (0,\bar{\epsilon}]$, then there exists $\bar{\mathbf{x}} \in \mathsf{B}(\epsilon,\mathbf{0})$, $\lambda_0 \in \mathbb{R}$, $\boldsymbol{\lambda} \in \mathbb{R}^{\mathrm{m}}$, and $\boldsymbol{\mu} \in \mathbb{R}^{\mathrm{k}}$ such that*

- *(i)* $\lambda_0, \mu_1, \dots, \mu_{\mathrm{s}} \in \mathbb{R}_{\ge 0}$,
- *(ii)* $\mu_{\mathrm{s}+1} = \cdots = \mu_{\mathrm{k}} = 0$,
- *(iii)* $\|(\lambda_0, \lambda_1, \dots, \lambda_{\mathrm{m}}, \mu_1, \dots, \mu_{\mathrm{k}})\|_{\mathbb{R}^{\mathrm{m+k+1}}} = 1$, *and*
- *(iv)* *for each* $\mathrm{j} \in \{1,\dots,\mathrm{n}\}$,

$$\lambda_0(\mathbf{D}_2\mathrm{f}(\bar{\mathbf{x}}) + 2\bar{\mathbf{x}}) + \sum_{\mathrm{a}=1}^{\mathrm{m}} \lambda_{\mathrm{a}}\mathbf{D}_{\mathrm{j}}\mathrm{g}_{\mathrm{a}}(\bar{\mathbf{x}}) + \sum_{\mathrm{r}=1}^{\mathrm{s}} \mu_{\mathrm{r}}\mathbf{D}_{\mathrm{j}}\mathrm{h}_{\mathrm{r}}(\bar{\mathbf{x}}) = 0.$$

Proof Let M be as in Lemma 1. Define

$$F(\boldsymbol{x}) = f(\boldsymbol{x}) + \|\boldsymbol{x}\|^2_{\mathbb{R}^n} + M\Big(\sum_{a=1}^{m} g_a(\boldsymbol{x})^2 + \sum_{r=1}^{s}(h_r(\boldsymbol{x})_+)^2\Big)$$

for $\boldsymbol{x} \in U$. Since $\overline{\mathsf{B}}^n(\epsilon,\boldsymbol{0})$ is compact and F is continuous, by Theorem 1.3.32 there exists $\bar{\boldsymbol{x}} \in \overline{\mathsf{B}}^n(\epsilon,\boldsymbol{0})$ such that

$$F(\bar{\boldsymbol{x}}) = \inf\{F(\boldsymbol{x}) \mid \boldsymbol{x} \in \overline{\mathsf{B}}^n(\epsilon,\boldsymbol{0})\}.$$

In particular, $F(\bar{\boldsymbol{x}}) \le F(\boldsymbol{0}) = 0$. Thus, by the definition of M from Lemma 1, $\|\bar{\boldsymbol{x}}\|_{\mathbb{R}^n} \ne \epsilon$. By Theorem 1.4.40 it follows that $\boldsymbol{D}F(\bar{\boldsymbol{x}}) = \boldsymbol{0}$ since $\bar{\boldsymbol{x}}$ is a local minimum for $F|\mathsf{B}^n(\epsilon,\boldsymbol{0})$.

Note that the function $x \mapsto (x_+)^2$ is continuously differentiable. Therefore, by the Chain Rule, the function $\boldsymbol{x} \mapsto (h_r(\boldsymbol{x})_+)^2$ is continuously differentiable for each $r \in \{1, \dots, s\}$. Moreover, also by the Chain Rule, its j partial derivative is given by

$$\boldsymbol{x} \mapsto 2h_s(\boldsymbol{x})_+ \boldsymbol{D}_j h_r(\boldsymbol{x}).$$

Thus an elementary computation gives

$$0 = \boldsymbol{D}_j F(\bar{\boldsymbol{x}}) = \boldsymbol{D}_j f(\bar{\boldsymbol{x}}) + 2x_j + \sum_{a=1}^{m} 2Mg_a(\bar{\boldsymbol{x}})\boldsymbol{D}_j g_a(\bar{\boldsymbol{x}}) + \sum_{r=1}^{s} 2Mh_s(\boldsymbol{x})_+ \boldsymbol{D}_j h_r(\boldsymbol{x}).$$

Now define

$$\lambda'_0 = 1, \quad \lambda'_a = 2Mg_a(\bar{\boldsymbol{x}}),\ a \in \{1, \dots, m\},$$
$$\mu'_r = 2Mh_r(\bar{\boldsymbol{x}})_+,\ r \in \{1, \dots, s\}, \quad \mu'_{s+1} = \dots = \mu'_k = 0.$$

Then let $\ell = \|(\lambda'_0, \lambda'_1, \dots, \lambda'_m, \mu'_1, \dots, \mu'_k)\|_{\mathbb{R}^{m+k+1}}$ and define $\lambda_a = \ell^{-1}\lambda'_a$, $a \in \{0, 1, \dots, m\}$, and $\mu_r = \ell^{-1}\mu'_r$, $r \in \{1, \dots, k\}$. One easily sees that these definitions satisfy the conclusions of the lemma. ▼

Now let $(\epsilon_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in $(0, \bar{\epsilon}]$ converging to 0. For each $j \in \mathbb{Z}_{>0}$, let $\bar{\boldsymbol{x}}_j \in \mathsf{B}^n(\epsilon_j, \boldsymbol{0})$, $\lambda_{0,j} \in \mathbb{R}_{\geq 0}$, $\boldsymbol{\lambda}_j \in \mathbb{R}^m$, and $\boldsymbol{\mu}_j \in \mathbb{R}^k$ satisfy the conclusions of Lemma 2 for ϵ_j. Then, since $\lim_{j \to \infty} \bar{\boldsymbol{x}}_j = \boldsymbol{0}$,

$$0 = \lim_{j \to \infty} \Big(\lambda_0 (\boldsymbol{D}_2 f(\bar{\boldsymbol{x}}_j) + 2\bar{\boldsymbol{x}}_j) + \sum_{a=1}^{m} \lambda_a \boldsymbol{D}_j g_a(\bar{\boldsymbol{x}}_j) + \sum_{r=1}^{s} \mu_r \boldsymbol{D}_j h_r(\bar{\boldsymbol{x}}_j) \Big)$$
$$= \lambda_0 \boldsymbol{D}_2 f(\boldsymbol{0}) + \sum_{a=1}^{m} \lambda_a \boldsymbol{D}_j g_a(\boldsymbol{0}) + \sum_{r=1}^{s} \mu_r \boldsymbol{D}_j h_r(\boldsymbol{0}).$$

This gives the conclusions of the theorem, with the exception of the final assertion.

For the final assertion, if $\lambda_0 = 0$ then the condition $\boldsymbol{D} f_{\lambda_0, \boldsymbol{\lambda}, \boldsymbol{\mu}}(\boldsymbol{0}) = 0$ with $\lambda = 0$ ensures that the set

$$\{\boldsymbol{D}g_1(\boldsymbol{0}), \dots, \boldsymbol{D}g_m(\boldsymbol{0}), \boldsymbol{D}h_1(\boldsymbol{0}), \dots, \boldsymbol{D}h_s(\boldsymbol{0})\}$$

is linearly dependent. As $\lambda_0 \in \mathbb{R}_{>0}$ we can define $\lambda'_0 = 1$, $\lambda'_a = \lambda_0^{-1}\lambda_a$, $a \in \{1, \dots, m\}$, and $\mu'_r = \lambda_0^{-1}\mu_r$, $r \in \{1, \dots, k\}$, and the resulting λ'_0, $\boldsymbol{\lambda}'$, and $\boldsymbol{\mu}'$ will satisfy the conclusions of the theorem with $\lambda_0 = 1$. ■

Many presentations of the Lagrange Multiplier Rule will omit the rôle of the constant λ_0, assuming it to be equal to 1. However, this is only valid when the condition (iii) of the theorem is satisfied, as the following example shows.

1.4.45 Example (Constrained extrema when the constraints are not linearly independent) We take $n = 2$, and define $f\colon \mathbb{R}^2 \to \mathbb{R}$ by $f(x_1, x_2) = x_1$ $g\colon \mathbb{R}^2 \to \mathbb{R}$ by $g(x_1, x_2) = x_1^2 + x_2^2$. We do not consider inequality constraints. Note that the only point satisfying the equality constraint defined by g is $(0, 0)$. Thus there is only one choice for a local minimum of (f, g) and so the solution of the problem is trivial. However, it is not possible to satisfy the conclusions of Theorem 1.4.44 for this solution unless $\lambda_0 = 0$. Indeed, if $\lambda_0 = 0$ then Theorem 1.4.44 tells us that there exists $\lambda_1, \lambda_2 \in \mathbb{R}$, not both zero, such that

$$\boldsymbol{D}_j f(0,0) + \lambda_1 \boldsymbol{D}_j g_1(0,0) + \lambda_2 \boldsymbol{D}_j g_2(0,0) = 0, \qquad j \in \{1, 2\}.$$

Thus gives $1 = 0$, which is rather absurd. However, the conclusions of Theorem 1.4.44 *are* satisfied for arbitrary $\lambda_1, \lambda_2 \in \mathbb{R}$ if we take $\lambda_0 = 0$. •

The preceding result gives necessary conditions for a point $\boldsymbol{x}_0$ to be a local minimum for $(f, \boldsymbol{g}, \boldsymbol{h})$. Let us now consider sufficient conditions involving the second derivative.

To conveniently state the theorem, we introduce some notation. If $\boldsymbol{\lambda} \in \mathbb{R}^m$ then we denote $\boldsymbol{f}_{\boldsymbol{\lambda}}\colon U \to \mathbb{R}$ the function given by

$$f_{\boldsymbol{\lambda}}(\boldsymbol{x}) = f(\boldsymbol{x}) + \langle \boldsymbol{\lambda}, \boldsymbol{g}(\boldsymbol{x}) \rangle_{\mathbb{R}^m}.$$

Let $Q_{\boldsymbol{\lambda}}(\boldsymbol{x})$ denote the restriction of the symmetric bilinear map $\boldsymbol{D}^2 f_{\boldsymbol{\lambda}}(\boldsymbol{x})$ to the subspace $\ker(\boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}))$. With this notation, we have the following theorem.

1.4.46 Theorem (Second-derivative tests for constrained minima) *Let $U \subseteq \mathbb{R}^n$ be open, and let $f\colon U \to \mathbb{R}$ and $\mathbf{g}\colon U \to \mathbb{R}^m$ be twice continuously differentiable. For $\mathbf{x}_0 \in U$, assume that $\mathbf{Dg}(\mathbf{x}_0)$ has rank m and that there exists $\boldsymbol{\lambda} \in \mathbb{R}^m$ such that $\mathbf{D}f_{\boldsymbol{\lambda}}(\mathbf{x}_0) = \mathbf{0}$. Then the following statements hold:*

(i) if $\mathbf{x}_0$ is a local maximum (resp. local minimum) for $(f, \mathbf{g})$, then $Q_{\boldsymbol{\lambda}}(\mathbf{x}_0)$ is negative-semidefinite (resp. positive-semidefinite);

(ii) if $Q_{\boldsymbol{\lambda}}(\mathbf{x}_0)$ is negative definite (resp. positive-definite), then $\mathbf{x}_0$ is a strict local maximum (resp. strict local minimum) for f;

(iii) if $Q_{\boldsymbol{\lambda}}(\mathbf{x}_0)$ is neither positive- nor negative-semidefinite, then $\mathbf{x}_0$ is neither a local minimum nor a local maximum.

Proof Let

$$S = \{\boldsymbol{v} \in \ker(\boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}_0)) \mid \|\boldsymbol{v}\|_{\mathbb{R}^n} = 1\}.$$

The following lemma, relying on the Implicit Function Theorem stated below as Theorem 1.5.3, is key to our proof.

1 Lemma *If $\mathbf{v} \in S$ there exists $\delta \in \mathbb{R}_{>0}$ and a continuously differentiable curve $\boldsymbol{\gamma}\colon [-\delta, \delta] \to \mathbf{g}^{-1}(\mathbf{0})$ such that $\boldsymbol{\gamma}(0) = \mathbf{x}_0$ and $\mathbf{D}\boldsymbol{\gamma}(0) = \mathbf{v}$.*

Proof For $\sigma \in \mathfrak{S}_n$ let $\mathsf{L}_\sigma\colon \mathbb{R}^n \to \mathbb{R}^n$ be defined by

$$\mathsf{L}_\sigma(x_1, \dots, x_n) = (x_{\sigma(1)}, \dots, x_{\sigma(n)}).$$

Note that

$$\boldsymbol{D}(\boldsymbol{g} \circ \mathsf{L}_\sigma)(\mathsf{L}_\sigma^{-1}(\boldsymbol{x}_0)) = \boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}_0) \circ \mathsf{L}_\sigma.$$

Let $\sigma \in \mathfrak{S}_n$ be such that the matrix

$$\begin{bmatrix} \boldsymbol{D}_{\sigma(1)} g_1(\boldsymbol{x}_0) & \cdots & \boldsymbol{D}_{\sigma(m)} g_1(\boldsymbol{x}_0) \\ \vdots & \ddots & \vdots \\ \boldsymbol{D}_{\sigma(1)} g_m(\boldsymbol{x}_0) & \cdots & \boldsymbol{D}_{\sigma(m)} g_m(\boldsymbol{x}_0) \end{bmatrix}$$

is invertible. Such a σ exists since $\boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}_0)$ has rank m, and so has m linearly independent columns. The permutation σ is chosen to shift these columns to be leftmost. Let $U' \subseteq \mathbb{R}^m$ and $V' \subseteq \mathbb{R}^{n-m}$ be open sets such that $\boldsymbol{x}_0 \in \mathsf{L}_\sigma(U' \times V') \subseteq U$, making the obvious identification of $\mathbb{R}^n$ with $\mathbb{R}^m \times \mathbb{R}^{n-m}$. Now note that the map $\boldsymbol{g} \circ \mathsf{L}_\sigma\colon U' \times V' \to \mathbb{R}^m$ satisfies the hypotheses of the Implicit Function Theorem at $\mathsf{L}_\sigma^{-1}(\boldsymbol{x}_0)$, and so, after shrinking V' if necessary, there exists a continuously differentiable map $\boldsymbol{h}\colon V' \to U'$ such that

$$(\boldsymbol{h}(\boldsymbol{y}), \boldsymbol{y}) = (\boldsymbol{g} \circ \mathsf{L}_\sigma)^{-1}(\boldsymbol{0}) \cap U' \times V'.$$

Moreover, also by the Implicit Function Theorem,

make sure this goes in there

$$\ker(\boldsymbol{D}(\boldsymbol{g} \circ \mathsf{L}_\sigma)(\mathsf{L}_\sigma^{-1}(\boldsymbol{x}_0))) = \{(\boldsymbol{D}\boldsymbol{h}(\boldsymbol{0}) \cdot \boldsymbol{u}, \boldsymbol{u}) \mid \boldsymbol{u} \in \mathbb{R}^{n-m}\}.$$

Let $\boldsymbol{y}_0 \in V'$ be such that $\boldsymbol{x}_0 = \mathsf{L}_\sigma(\boldsymbol{h}(\boldsymbol{y}_0), \boldsymbol{y}_0)$. Note that

$$\mathsf{L}_\sigma^{-1}(\boldsymbol{v}) \in \ker(\boldsymbol{D}(\boldsymbol{g} \circ \mathsf{L}_\sigma)(\mathsf{L}_\sigma^{-1}(\boldsymbol{x}_0)))$$

and thus there exists $\boldsymbol{u} \in \mathbb{R}^{n-m}$ such that $(\boldsymbol{D}\boldsymbol{h}(\boldsymbol{y}_0) \cdot \boldsymbol{u}, \boldsymbol{u}) = \mathsf{L}_\sigma^{-1}(\boldsymbol{v})$. The curve

$$\boldsymbol{\gamma}'(s) = (\boldsymbol{h}(\boldsymbol{y}_0 + s\boldsymbol{u}), \boldsymbol{y}_0 + s\boldsymbol{u}),$$

defined for s sufficiently small, satisfies $\boldsymbol{D}\boldsymbol{\gamma}'(0) = \mathsf{L}_\sigma^{-1}(\boldsymbol{v})$. Therefore, the curve

$$\boldsymbol{\gamma}(s) = \mathsf{L}_\sigma \circ \boldsymbol{\gamma}'(s)$$

satisfies $\boldsymbol{D}\boldsymbol{\gamma}(0) = \boldsymbol{v}$. Moreover,

$$\boldsymbol{g}(\boldsymbol{\gamma}(s)) = \boldsymbol{g} \circ \mathsf{L}_\sigma(\boldsymbol{\gamma}'(s)) = \boldsymbol{0}$$

by definition of $\boldsymbol{\gamma}'$, and so we get the lemma. ▼

With the lemma at hand, the remainder of the proof is more or less straightforward, following the proofs of parts (ii), (ii), and (iv) of Theorem 1.4.40. Moreover, we shall only prove the statements corresponding to local maxima, as the statements for local minima follow using the same ideas.

(i) Suppose that $Q_\lambda(\boldsymbol{x}_0)$ is not positive-semidefinite. Then there exists $\boldsymbol{v} \in S$ such that $Q_\lambda(\boldsymbol{x}_0) \cdot (\boldsymbol{v}, \boldsymbol{v}) < 0$. By the lemma, let $\boldsymbol{\gamma}$ be a curve in $\boldsymbol{g}^{-1}(\boldsymbol{0})$ such that $\boldsymbol{\gamma}(0) = \boldsymbol{x}_0$ and $\boldsymbol{D}\boldsymbol{\gamma}(0) = \boldsymbol{v}$. Following the ideas in Theorem 1.4.40, write

$$f_\lambda(\boldsymbol{\gamma}(s)) - f_\lambda(\boldsymbol{x}_0) = \tfrac{1}{2}\boldsymbol{D}^2 f_\lambda(\boldsymbol{x}_0) + o(s^2).$$

Let $s_0 \in \mathbb{R}_{>0}$ be sufficiently small that

$$|o(s^2)| < \frac{1}{4}\boldsymbol{D}^2 f_\lambda(\boldsymbol{x}_0) \cdot (\boldsymbol{v}, \boldsymbol{v})$$

for every $s \in (0, s_0]$. Then

$$\tfrac{1}{2}\boldsymbol{D}^2 f_\lambda(\boldsymbol{x}_0) + o(s^2) < 0$$

and so $f_\lambda(\boldsymbol{\gamma}(s)) < f_\lambda(\boldsymbol{x}_0)$ for every $s \in (0, s_0]$, showing that $\boldsymbol{x}_0$ is not a local minimum for f_λ. Since $f_\lambda|\boldsymbol{g}^{-1}(\boldsymbol{0}) = f_\lambda|\boldsymbol{g}^{-1}(\boldsymbol{0})$, this part of the result follows.

(ii) Suppose that $\boldsymbol{D}^2 f_\lambda(\boldsymbol{x}_0)$ is positive-definite. Let

$$m = \inf\{\tfrac{1}{2}\boldsymbol{D}^2 f_\lambda(\boldsymbol{x}_0) \cdot (\boldsymbol{v}, \boldsymbol{v}) \mid \boldsymbol{v} \in S\},$$

noting that $m \in \mathbb{R}_{>0}$. Let

$$M = \{\boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{x}_0 + \boldsymbol{v} \in \boldsymbol{g}^{-1}(\boldsymbol{0})\}.$$

Note that

$$\begin{aligned}\lim_{\substack{\boldsymbol{v}\to\boldsymbol{0}\\ \boldsymbol{v}\in M}} \boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}_0) \cdot (\tfrac{\boldsymbol{v}}{\|\boldsymbol{v}\|_{\mathbb{R}^n}}) &= \lim_{\substack{\boldsymbol{v}\to\boldsymbol{0}\\ \boldsymbol{v}\in M}} \Big(\boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}_0) \cdot (\tfrac{\boldsymbol{v}}{\|\boldsymbol{v}\|_{\mathbb{R}^n}}) - \frac{\boldsymbol{g}(\boldsymbol{x}_0 + \boldsymbol{v}) - \boldsymbol{g}(\boldsymbol{x}_0)}{\|\boldsymbol{v}\|_{\mathbb{R}^n}}\Big) \\ &= \lim_{\substack{\boldsymbol{v}\to\boldsymbol{0}\\ \boldsymbol{v}\in M}} \Big(\frac{\boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}_0) \cdot \boldsymbol{v} - \boldsymbol{g}(\boldsymbol{x}_0 + \boldsymbol{v}) - \boldsymbol{g}(\boldsymbol{x}_0)}{\|\boldsymbol{v}\|_{\mathbb{R}^n}}\Big) = 0.\end{aligned}$$

Thus

$$\lim_{\substack{\boldsymbol{v}\to\boldsymbol{0}\\ \boldsymbol{v}\in M}} \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_{\mathbb{R}^n}} \in S.$$

Thus, given $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $\|\boldsymbol{v}\|_{\mathbb{R}^n} < \delta$ then there exists $\boldsymbol{u} \in S$ such that $\|\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_{\mathbb{R}^n}} - \boldsymbol{u}\|_{\mathbb{R}^n} < \epsilon$. Because the function

$$\boldsymbol{v} \mapsto \tfrac{1}{2}\boldsymbol{D}^2 f_\lambda(\boldsymbol{x}_0) \cdot (\boldsymbol{v}, \boldsymbol{v})$$

is continuous, it follows from Theorem 1.3.33 that it is uniformly continuous on the compact set

$$\{\boldsymbol{u} + \boldsymbol{v} \in \mathbb{R}^n \mid \boldsymbol{u} \in S,\ \|\boldsymbol{v}\|_{\mathbb{R}^n} \le \epsilon\}.$$

Therefore, by choosing δ (and thus ϵ) sufficiently small, we can ensure that $\|\frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_{\mathbb{R}^n}}\|_{\mathbb{R}^n} > \frac{1}{2}m$ for $\boldsymbol{v} \in M$ such that $\|\boldsymbol{v}\|_{\mathbb{R}^n} < \delta$. As in the proof of Theorem 1.4.40, write

$$f_\lambda(\boldsymbol{x}_0 + \boldsymbol{v}) - f_\lambda(\boldsymbol{x}_0) = \tfrac{1}{2}\boldsymbol{D}^2 f_\lambda(\boldsymbol{x}_0) \cdot (\boldsymbol{v}, \boldsymbol{v}) + o(\boldsymbol{v}^2).$$

By making δ smaller if necessary, we can ensure that

$$\frac{o(\boldsymbol{v}^2)}{\|\boldsymbol{v}\|^2_{\mathbb{R}^n}} < \frac{1}{4}m.$$

In this case, for $\boldsymbol{v} \in M$,

$$\tfrac{1}{2}\boldsymbol{D}^2 f_\lambda(\boldsymbol{x}_0) \cdot (\boldsymbol{v}, \boldsymbol{v}) + o(\boldsymbol{v}^2) = \|\boldsymbol{v}\|^2_{\mathbb{R}^n}\Big(\tfrac{1}{2}\boldsymbol{D}^2 f_\lambda(\boldsymbol{x}_0) \cdot (\tfrac{\boldsymbol{v}}{\|\boldsymbol{v}\|_{\mathbb{R}^n}}, \tfrac{\boldsymbol{v}}{\|\boldsymbol{v}\|^2_{\mathbb{R}^n}}) + \frac{o(\boldsymbol{v}^2)}{\|\boldsymbol{v}\|^2_{\mathbb{R}^n}}\Big) \ge \frac{1}{4}m\|\boldsymbol{v}\|^2_{\mathbb{R}^n} > 0.$$

This shows that $f_\lambda(\boldsymbol{x}) > f_\lambda(\boldsymbol{x}_0)$ for $\boldsymbol{x} \in \boldsymbol{g}^{-1}(\boldsymbol{0})$ is a neighbourhood of $\boldsymbol{x}_0$. Since $f_\lambda|\boldsymbol{g}^{-1}(\boldsymbol{0}) = f|\boldsymbol{g}^{-1}(\boldsymbol{0})$, this part of the theorem follows.

(iii) The proof here follows the proof of part (iii). ■

1.4.9 The derivative and operations on functions

In this section we give the usual results concerning how differentiation interacts with the usual function operations.

Our first result deals with algebraic operations on functions, and for this we note that if $A \subseteq \mathbb{R}^n$, if $\boldsymbol{f}, \boldsymbol{g}\colon A \to \mathbb{R}^m$, and if $\alpha \in \mathbb{R}$ then we define $\boldsymbol{f} + \boldsymbol{g}, \alpha \boldsymbol{f}\colon U \to \mathbb{R}^m$ by

$$(\boldsymbol{f} + \boldsymbol{g})(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}) + \boldsymbol{g}(\boldsymbol{x}), \quad (\alpha \boldsymbol{f})(\boldsymbol{x}) = \alpha(\boldsymbol{f}(\boldsymbol{x})), \qquad \boldsymbol{x} \in A.$$

If, moreover, $m = 1$ and we denote the maps by $f, g\colon A \to \mathbb{R}$, then we define $fg, \frac{f}{g}\colon A \to \mathbb{R}$ by

$$(fg)(\boldsymbol{x}) = f(\boldsymbol{x})g(\boldsymbol{x}), \quad \Big(\frac{f}{g}\Big)(\boldsymbol{x}) = \frac{f(\boldsymbol{x})}{g(\boldsymbol{x})}, \qquad \boldsymbol{x} \in A.$$

With this notation we have the following result.

1.4.47 Proposition (The derivative, and addition and multiplication) *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *be open, let* $\mathbf{f}, \mathbf{g}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$ *be* r *times differentiable at* $\mathbf{x}_0 \in \mathrm{U}$*, and let* $\alpha \in \mathbb{R}$*. Then the following statements hold:*

(i) $\mathbf{f} + \mathbf{g}$ *is* r *times differentiable at* $\mathbf{x}_0$ *and* $\mathbf{D}^{\mathrm{r}}(\mathbf{f} + \mathbf{g})(\mathbf{x}_0) = \mathbf{D}^{\mathrm{r}}\mathbf{f}(\mathbf{x}_0) + \mathbf{D}^{\mathrm{r}}\mathbf{g}(\mathbf{x}_0)$*;*

(ii) $\alpha\mathbf{f}$ *is* r *times differentiable at* $\mathbf{x}_0 \in \mathrm{U}$ *and* $\mathbf{D}(\alpha\mathbf{f})(\mathbf{x}_0) = \alpha\mathbf{D}\mathbf{f}(\mathbf{x}_0)$*.*

Moreover, if $\mathrm{m} = 1$ *and if* $\mathrm{f}, \mathrm{g}\colon \mathrm{U} \to \mathbb{R}$ *are differentiable at* $\mathbf{x}_0$ *then the following statements hold:*

(iii) fg *is differentiable at* $\mathbf{x}_0$ *and*

$$\mathbf{D}(\mathrm{fg})(\mathbf{x}_0) = \mathrm{g}(\mathbf{x}_0)\mathbf{D}\mathrm{f}(\mathbf{x}_0) + \mathrm{g}(\mathbf{x}_0)\mathbf{D}\mathrm{f}(\mathbf{x}_0);$$

(iv) if $\mathrm{g}(\mathbf{x}_0) \neq 0$ *then* $\frac{\mathrm{f}}{\mathrm{g}}$ *is differentiable at* $\mathbf{x}_0$ *and*

$$\mathbf{D}\Big(\frac{\mathrm{f}}{\mathrm{g}}\Big)(\mathbf{x}_0) = \frac{\mathrm{g}(\mathbf{x}_0)\mathbf{D}\mathrm{f}(\mathbf{x}_0) - \mathrm{f}(\mathbf{x}_0)\mathbf{D}\mathrm{g}(\mathbf{x}_0)}{\mathrm{g}(\mathbf{x}_0)^2}.$$

Proof (i) We shall prove the assertion for $r = 1$, the general assertion following from this case by a simple induction. We compute

$$\begin{aligned}\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} & \frac{\|(\boldsymbol{f} + \boldsymbol{g})(\boldsymbol{x}) - (\boldsymbol{f} + \boldsymbol{g})(\boldsymbol{x}_0) - (\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}_0) + \boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}_0)) \cdot (\boldsymbol{x} - \boldsymbol{x}_0)\|_{\mathbb{R}^m}}{\|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n}} \\ &= \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \frac{\|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{f}(\boldsymbol{x}_0) - \boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{x} - \boldsymbol{x}_0)\|_{\mathbb{R}^m}}{\|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n}} \\ &\qquad + \lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \frac{\|\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{x}_0) - \boldsymbol{D}\boldsymbol{g}(\boldsymbol{x}_0) \cdot (\boldsymbol{x} - \boldsymbol{x}_0)\|_{\mathbb{R}^m}}{\|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n}} = 0,\end{aligned}$$

using Proposition 1.2.6.

(ii) Again we only prove the result for $r = 1$, the general case following by induction. We again use Proposition 1.2.6 to get

$$\lim_{x\to x_0} \frac{\|(\alpha f)(x) - (\alpha f)(x_0) - (\alpha Df(x_0)) \cdot (x - x_0)\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}} = \alpha\Big(\lim_{x\to x_0} \frac{\|f(x) - f(x_0) - Df(x_0) \cdot (x - x_0)\|_{\mathbb{R}^m}}{\|x - x_0\|_{\mathbb{R}^n}}\Big) = 0.$$

(iii) We shall simply show how this part of the result follows from Theorem 1.4.48. Define $\mathsf{B} \in \mathrm{L}^2(\mathbb{R};\mathbb{R})$ by $\mathsf{B}(a_1, a_2) = a_1 a_2$ so that $(fg)(x) = \mathsf{B}(f(x), g(x))$, and this then immediately gives this part of the result.

(iv) Since $g(x_0) \neq 0$ and since g is continuous at x_0 by Proposition 1.4.35 there exists a neighbourhood $V \subseteq U$ of x_0 such that $g(x)$ has the same sign as $g(x_0)$ for all $x \in V$. Thus the function $\iota\colon y \mapsto \frac{1}{y}$ is differentiable on $g(V)$. If we define $h\colon V \to \mathbb{R}$ by $h(x) = \frac{1}{g(x)}$ then h is differentiable at x_0 by the Chain Rule and, moreover,

$$Dh(x_0) = D\iota(g(x_0)) \circ Dg(x_0) = -\frac{Dg(x_0)}{g(x_0)^2}.$$

The result now follows from part (iii) noting that $\frac{f}{g} = hf$. ■

Part (iii) of the preceding result is the ***product rule***. Sometimes a more sophisticated version of this is useful, and so we state this here.

1.4.48 Theorem (Leibniz Rule) *Let $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ be open, let $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{r}}$ and $\mathbf{g}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{s}}$ be differentiable at $\mathbf{x}_0 \in \mathrm{U}$, and let $\mathsf{B} \in \mathrm{L}(\mathbb{R}^{\mathrm{r}}, \mathbb{R}^{\mathrm{s}}; \mathbb{R}^{\mathrm{m}})$. If $\mathbf{h}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$ is defined by $\mathbf{h}(\mathbf{x}) = \mathsf{B}(\mathbf{f}(\mathbf{x}), \mathbf{g}(\mathbf{x}))$ then $\mathbf{h}$ is differentiable at $\mathbf{x}_0$ and, moreover,*

$$\mathbf{Dh}(\mathbf{x}_0) \cdot \mathbf{v} = \mathsf{B}(\mathbf{Df}(\mathbf{x}_0) \cdot \mathbf{v}, \mathbf{g}(\mathbf{x}_0)) + \mathsf{B}(\mathbf{f}(\mathbf{x}_0), \mathbf{Dg}(\mathbf{x}_0) \cdot \mathbf{v})$$

for every $\mathbf{v} \in \mathbb{R}^{\mathrm{n}}$.

Proof By Theorem 1.4.8 the map $\mathsf{B}\colon \mathbb{R}^r \times \mathbb{R}^s \to \mathbb{R}^m$ is differentiable and

$$D\mathsf{B}(p_0, q_0) \cdot (u, w) = \mathsf{B}(u, q_0) + \mathsf{B}(p_0, w) \tag{1.28}$$

for every $(u, w) \in \mathbb{R}^r \oplus \mathbb{R}^s$. Since $h = \mathsf{B} \circ (f \times g)$ it follows from the Chain Rule below that

$$Dh(x_0) \cdot v = D\mathsf{B}((f \times g)(x_0)) \circ D(f \times g)(x_0) \cdot v.$$

By Proposition 1.4.17 we have

$$D(f \times g)(x_0) \cdot v = (Df(x_0) \cdot v, Dg(x_0) \cdot v),$$

and the result then follows from (1.28). ■

We next state the multivariable Chain Rule, this being one of the most important theorems concerning the derivative. Indeed, we have already used this result many times in this section.

1.4.49 Theorem (Chain Rule) *Let* $U \subseteq \mathbb{R}^n$ *and* $V \subseteq \mathbb{R}^m$ *be open, consider maps* $\mathbf{f}\colon U \to V$ *and* $\mathbf{g}\colon V \to \mathbb{R}^k$*, and let* $\mathbf{x}_0 \in U$*. If* $\mathbf{f}$ *is differentiable at* $\mathbf{x}_0$ *and if* $\mathbf{g}$ *is differentiable at* $\mathbf{f}(\mathbf{x}_0)$*, then* $\mathbf{g} \circ \mathbf{f}$ *is differentiable at* $\mathbf{x}_0$ *and, moreover,*

$$\mathbf{D}(\mathbf{g} \circ \mathbf{f})(\mathbf{x}_0) = \mathbf{D}\mathbf{g}(\mathbf{f}(\mathbf{x}_0)) \circ \mathbf{D}\mathbf{f}(\mathbf{x}_0).$$

Proof Let $\epsilon \in \mathbb{R}_{>0}$.

By Proposition 1.4.35 let $\delta_1, M \in \mathbb{R}_{>0}$ be such that

$$\|f(x) - f(x_0)\|_{\mathbb{R}^m} \le M\|x - x_0\|_{\mathbb{R}^n}$$

for $x \in \mathsf{B}(\delta_1, x_0)$. Since g is differentiable at $f(x_0)$ there exists $\eta \in \mathbb{R}_{>0}$ such that

$$\|g(y) - g \circ f(x_0) - Dg(f(x_0)) \cdot (y - f(x_0))\|_{\mathbb{R}^k} \le \tfrac{\epsilon}{2M}\|y - f(x_0)\|_{\mathbb{R}^m}$$

for $y \in \mathsf{B}(\eta, f(x_0))$. Since f is continuous at x_0 there exists $\delta_1 \in \mathbb{R}_{>0}$ such that

$$\|f(x) - f(x_0)\|_{\mathbb{R}^m} \le \eta$$

for $x \in \mathsf{B}(\delta_1, x_0)$. Then, letting $\delta_3 = \min\{\delta_1, \delta_2\}$, if $x \in \mathsf{B}(\delta_3, x_0)$ we have

$$\|g \circ f(x) - g \circ f(x_0) - Dg(f(x_0)) \cdot (f(x) - f(x_0))\|_{\mathbb{R}^k} \le \\ \tfrac{\epsilon}{2M}\|f(x) - f(x_0)\|_{\mathbb{R}^m} \le \tfrac{\epsilon}{2}\|x - x_0\|_{\mathbb{R}^n}.$$

By differentiability of f at x_0 let $\delta_4 \in \mathbb{R}_{>0}$ be such that

$$\|f(x) - f(x_0) - Df(x_0) \cdot (x - x_0)\|_{\mathbb{R}^m} \le \frac{\epsilon}{2\|Dg(f(x_0))\|_{\mathbb{R}^n,\mathbb{R}^m}}\|x - x_0\|_{\mathbb{R}^n}$$

for $x \in \mathsf{B}(\delta_4, x_0)$. By Proposition 1.1.16(v) we then have

$$\|Dg(f(x_0)) \cdot (f(x) - f(x_0) - Df(x_0) \cdot (x - x_0))\|_{\mathbb{R}^k} \\ \le \|Dg(f(x_0))\|_{\mathbb{R}^n,\mathbb{R}^m}\|f(x) - f(x_0) - Df(x_0) \cdot (x - x_0)\|_{\mathbb{R}^m} \le \tfrac{\epsilon}{2}\|x - x_0\|_{\mathbb{R}^n}$$

for $x \in \mathsf{B}(\delta_4, x_0)$.

Now let $\delta = \min\{\delta_3, \delta_4\}$ and note that if $x \in \mathsf{B}(\delta, x_0)$ then we have, using the triangle inequality,

$$\begin{aligned} &\|g \circ f(x) - g \circ f(x_0) - Dg(f(x_0)) \circ Df(x_0) \cdot (x - x_0)\|_{\mathbb{R}^k} \\ &\quad\le \|g \circ f(x) - g \circ f(x_0) - Dg(f(x_0)) \cdot (f(x) - f(x_0))\|_{\mathbb{R}^k} \\ &\qquad + \|Dg(f(x_0)) \cdot (f(x) - f(x_0) - Df(x_0) \cdot (x - x_0))\|_{\mathbb{R}^k} \\ &\quad\le \tfrac{\epsilon}{2}\|x - x_0\|_{\mathbb{R}^n} + \tfrac{\epsilon}{2}\|x - x_0\|_{\mathbb{R}^n} = \epsilon\|x - x_0\|_{\mathbb{R}^n}. \end{aligned}$$

This gives

$$\frac{\|g \circ f(x) - g \circ f(x_0) - Dg(f(x_0)) \circ Df(x_0) \cdot (x - x_0)\|_{\mathbb{R}^k}}{\|x - x_0\|_{\mathbb{R}^n}} < \epsilon,$$

for $x \in \mathsf{B}(\delta, x_0)$, giving differentiability of $g \circ f$ at x_0 with derivative as asserted in the theorem. ■

For completeness let us also give the higher-order versions of the Leibniz and Chain Rules. To state these results in a compact way it is convenient to borrow some of our notation concerning the symmetric group that was given preceding Proposition I-4.1.38. Let $r, r_1, \dots, r_k \in \mathbb{Z}_{\geq 0}$ have the property that $r_1 + \cdots + r_k = r$. Then we recall the subgroup $\mathfrak{S}_{r_1|\cdots|r_k}$ of $\mathfrak{S}_r$ that leaves the "slots" of length $r_1, \dots, r_k$ in $\{1, \dots, r\}$ invariant. The situation here is slightly different than that preceding the statement of Proposition I-4.1.38 in that we allow some of the numbers $r_1, \dots, r_k$ to be zero. However, this amounts to the same thing since the "slots" of length zero do not contribute materially. We also denote by $\mathfrak{S}_{r_1,\dots,r_k}$ the subset of $\mathfrak{S}_r$ having the property that $\sigma \in \mathfrak{S}_{r_1,\dots,r_k}$ satisfies

$$\sigma(r_1 + \cdots + r_j + 1) < \cdots < \sigma(r_1 + \cdots + r_j + r_{j+1}), \qquad j \in \{0, 1, \dots, k-1\}.$$

Again, this notation is in slight conflict with that preceding Proposition I-4.1.38 in that some of the numbers $r_1, \dots, r_k$ are allowed to be zero. With this notation we may state the following version of Leibniz' Rule, generalising to arbitrary derivatives and arbitrary multilinear maps.

1.4.50 Theorem (General Leibniz Rule) *Let $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ be open, let $\mathbf{f}_{\mathrm{j}}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{n}_{\mathrm{j}}}$, $\mathrm{j} \in \{1, \dots, \mathrm{k}\}$, be r times differentiable at $\mathbf{x}_0 \in \mathrm{U}$, and let $\mathsf{L} \in \mathrm{L}(\mathbb{R}^{\mathrm{n}_1}, \dots, \mathbb{R}^{\mathrm{n}_{\mathrm{k}}}; \mathbb{R}^{\mathrm{m}})$. If $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$ is defined by*

$$\mathbf{f}(\mathbf{x}) = \mathsf{L}(\mathbf{f}_1(\mathbf{x}), \dots, \mathbf{f}_{\mathrm{k}}(\mathbf{x}))$$

then $\mathbf{f}$ is r times differentiable at $\mathbf{x}_0$ and, moreover,

$$\mathbf{D}^{\mathrm{r}}\mathbf{f}(\mathbf{x}_0) \cdot (\mathbf{v}_1, \dots, \mathbf{v}_{\mathrm{r}}) = \sum_{\substack{\mathrm{r}_1,\dots,\mathrm{r}_{\mathrm{k}} \in \mathbb{Z}_{\geq 0} \\ \mathrm{r}_1 + \cdots + \mathrm{r}_{\mathrm{k}} = \mathrm{r}}} \sum_{\sigma \in \mathfrak{S}_{\mathrm{r}_1,\cdots,\mathrm{r}_{\mathrm{k}}}} \mathsf{L}(\mathbf{D}^{\mathrm{r}_1}\mathbf{f}_1(\mathbf{x}_0) \cdot (\mathbf{v}_{\sigma(1)}, \dots, \mathbf{v}_{\sigma(\mathrm{r}_1)}), \dots, \\ \mathbf{D}^{\mathrm{r}_{\mathrm{k}}}\mathbf{f}_{\mathrm{k}}(\mathbf{x}_0) \cdot (\mathbf{v}_{\sigma(\mathrm{r}_1 + \cdots + \mathrm{r}_{\mathrm{k}-1} + 1)}, \dots, \mathbf{v}_{\sigma(\mathrm{r})}))$$

for $\mathbf{v}_1, \dots, \mathbf{v}_{\mathrm{r}} \in \mathbb{R}^{\mathrm{n}}$.

Proof We prove the theorem by induction on r, noting that the case of $r = 1$ follows from the Chain Rule, Theorem 1.4.8, and Proposition 1.4.17, using the fact that $f = \mathsf{L} \circ (f_1 \times \cdots \times f_k)$.

Assume the result is true for $r \in \{1, \dots, s\}$ and suppose that $f_1, \dots, f_k$ are of class C^{s+1}. Thus, by Proposition 1.4.7, for fixed $v_1, \dots, v_s \in \mathbb{R}^n$ the function

$$\begin{aligned} x \mapsto D^s f(x) \cdot (v_2, \dots, v_{s+1}) \\ = \sum_{\substack{s_1,\dots,s_k \in \mathbb{Z}_{\geq 0} \\ s_1 + \cdots + s_k = s}} \sum_{\sigma \in \mathfrak{S}_{s_1,\cdots,s_k}} \mathsf{L}(D^{s_1} f_1(x) \cdot (v_{\sigma(2)}, \dots, v_{\sigma(s_1+1)}), \dots, \\ D^{s_k} f_k(x) \cdot (v_{\sigma(s_1 + \cdots + s_{k-1} + 2)}, \dots, v_{\sigma(s+1)})), \end{aligned}$$

is differentiable at x_0, where we think of $\sigma \in \mathfrak{S}_s$ as a permutation of the set $\{2, \dots, s+1\}$ in the obvious way.

Let us now make an observation about permutations. Let $s'_1, \dots, s'_k \in \mathbb{Z}_{>0}$ have the property that $s'_1 + \dots + s'_k = s + 1$ and let $\sigma' \in \mathfrak{S}_{s'_1,\dots,s'_k}$. For brevity denote $t'_j = s'_1 + \dots + s'_j$ for $j \in \{1, \dots, k\}$. Then there exist unique $s_1, \dots, s_k \in \mathbb{Z}_{\geq 0}$ (denote $t_j = s_1 + \dots + s_j$, $j \in \{1, \dots, k\}$), $\sigma \in \mathfrak{S}_{s_1,\dots,s_k}$, and $j_0 \in \{1, \dots, k\}$ such that

$$s_j = \begin{cases} s'_j, & j \neq j_0, \\ s'_j - 1, & j = j_0 \end{cases}$$

and

$$\begin{aligned}((\sigma'(t'_1 - s'_1 + 1), \dots, \sigma'(t'_1)), \dots, (\sigma'(t'_{j_0} - s'_{j_0} + 1), \dots, \sigma'(t'_{j_0})), \dots,& \\ (\sigma'(t'_k - s'_k + 1) + \dots + \sigma'(t'_k))) = ((\sigma(t_1 - s_1 + 1), \dots, \sigma(t_1)), \dots,& \\ (1, \sigma(t_{j_0} - s_{j_0}), \dots, \sigma(t_{j_0} + 1)), \dots,& \\ (\sigma(t_k - s_k), \dots, \sigma(t_k + 1))),& \quad (1.29)\end{aligned}$$

with the convention that σ permutes the set $\{1, \dots, t'_{j_0} - s'_{j_0}, t'_{j_0} - s'_{j_0} + 2, \dots, s + 1\}$ in the obvious way. The point is that $\sigma'(t'_{j_0} - s'_{j_0} + 1) = 1$, and by definition of $\mathfrak{S}_{s'_1,\dots,s'_k}$ this means that $\sigma'(t'_{j_0} - s'_{j_0} + 1)$ must appear at the beginning of one of the "slots" of length $s'_1, \dots, s'_k$. Conversely, let $s_1, \dots, s_k \in \mathbb{Z}_{\geq 0}$ be such that $s_1 + \dots + s_k = s \geq 2$ and let $\sigma \in \mathfrak{S}_{s_1,\dots,s_k}$. Denote $t_j = s_1 + \dots + s_j$ for $j \in \{1, \dots, k\}$. Then, for each $j_0 \in \{1, \dots, k\}$ there exist unique $s'_1, \dots, s'_k \in \mathbb{Z}_{\geq 0}$ (denote $t'_j = s'_1 + \dots + s'_j$, $j \in \{1, \dots, k\}$) such that

$$s'_j = \begin{cases} s_j, & j \neq j_0, \\ s_j + 1, & j = j_0 \end{cases}$$

and $\sigma' \in \mathfrak{S}_{s'_1,\dots,s'_k}$ such that (1.29) holds.

Using this observation, and since the result holds for $r = 1$ and $r = s$, we can apply Proposition 1.4.7 to get

$$\begin{aligned} \boldsymbol{D}^{s+1}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_1, \dots, \boldsymbol{v}_{s+1}) &= (\boldsymbol{D}(\boldsymbol{D}^s\boldsymbol{f})(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_2, \dots, \boldsymbol{v}_{s+1})) \cdot \boldsymbol{v}_1 \\ &= \Big(\sum_{\substack{s_1,\dots,s_k \in \mathbb{Z}_{\geq 0} \\ s_1 + \dots + s_k = s}} \sum_{\sigma \in \mathfrak{S}_{s_1,\dots,s_k}} \mathsf{L}(\boldsymbol{D}^{s_1+1}\boldsymbol{f}_1(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_{\sigma(2)}, \dots, \boldsymbol{v}_{\sigma(s_1+1)}), \dots, \\ &\qquad \boldsymbol{D}^{s_k}\boldsymbol{f}_k(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(s_1+\dots+s_{k-1}+2)}, \dots, \boldsymbol{v}_{\sigma(s+1)}))\Big) + \dots \\ &\qquad + \Big(\sum_{\substack{s_1,\dots,s_k \in \mathbb{Z}_{\geq 0} \\ s_1 + \dots + s_k = s}} \sum_{\sigma \in \mathfrak{S}_{s_1,\dots,s_k}} \mathsf{L}(\boldsymbol{D}^{s_1}\boldsymbol{f}_1(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(2)}, \dots, \boldsymbol{v}_{\sigma(s_1+1)}), \dots, \\ &\qquad \boldsymbol{D}^{s_k+1}\boldsymbol{f}_k(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_{\sigma(s_1+\dots+s_{k-1}+2)}, \dots, \boldsymbol{v}_{\sigma(s+1)}))\Big) \\ &= \sum_{\substack{s'_1,\dots,s'_k \in \mathbb{Z}_{\geq 0} \\ s'_1 + \dots + s'_k = s+1}} \sum_{\sigma \in \mathfrak{S}_{s'_1,\dots,s'_k}} \mathsf{L}(\boldsymbol{D}^{s'_1}\boldsymbol{f}_1(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(1)}, \dots, \boldsymbol{v}_{\sigma(s'_1+1)}), \dots, \\ &\qquad \boldsymbol{D}^{s'_k}\boldsymbol{f}_k(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(s'_1+\dots+s'_{k-1}+1)}, \dots, \boldsymbol{v}_{\sigma(s+1)})), \end{aligned}$$

as desired. ■

In Exercise 1.4.4 we ask the reader to come to grips with the formula in the theorem by writing it down explicitly in some simple cases.

Now let us consider the Chain Rule for higher-order derivatives. To conveniently state the result we introduce the following notation. Let $r \in \mathbb{Z}_{>0}$ and let $r_1, \dots, r_j \in \mathbb{Z}_{\geq 0}$ have the property that $r_1 + \cdots + r_j = r$. Let us denote by $\mathfrak{S}^{<}_{r_1,\dots,r_j}$ the subset of $\mathfrak{S}_{r_1,\dots,r_j}$ given by

$$\mathfrak{S}^{<}_{r_1,\dots,r_j} = \{\sigma \in \mathfrak{S}_{r_1,\dots,r_j} \mid \sigma(1) < \sigma(r_1+1) < \cdots < \sigma(r_{j-1}+1)\}.$$

Note, for example, that if $\sigma \in \mathfrak{S}^{<}_{r_1,\dots,r_j}$ then $\sigma(1) = 1$.

With this notation we have the following statement of the Chain Rule.

1.4.51 Theorem (General Chain Rule) *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *and* $\mathrm{V} \subseteq \mathbb{R}^{\mathrm{m}}$ *be open, consider maps* $\mathbf{f}\colon \mathrm{U} \to \mathrm{V}$ *and* $\mathbf{g}\colon \mathrm{V} \to \mathbb{R}^{\mathrm{k}}$*, and let* $\mathbf{x}_0 \in \mathrm{U}$*. If* $\mathbf{f}$ *is* r *times differentiable at* $\mathbf{x}_0$ *and if* $\mathbf{g}$ *is* r *times differentiable at* $\mathbf{f}(\mathbf{x}_0)$*, then* $\mathbf{g} \circ \mathbf{f}$ *is* r *times differentiable at* $\mathbf{x}_0$ *and, moreover,*

$$\begin{aligned}\mathbf{D}^{\mathrm{r}}(\mathbf{g}\circ\mathbf{f})(\mathbf{x}_0)\cdot(\mathbf{v}_1,\dots,\mathbf{v}_{\mathrm{r}}) &= \sum_{\mathrm{j}=1}^{\mathrm{r}} \sum_{\substack{\mathrm{r}_1,\dots,\mathrm{r}_\mathrm{j}\in\mathbb{Z}_{>0}\\ \mathrm{r}_1+\cdots+\mathrm{r}_\mathrm{j}=\mathrm{r}}} \sum_{\sigma\in\mathfrak{S}^{<}_{\mathrm{r}_1,\dots,\mathrm{r}_\mathrm{j}}} \mathbf{D}^{\mathrm{j}}\mathbf{g}(\mathbf{f}(\mathbf{x}_0))\cdot(\mathbf{D}^{\mathrm{r}_1}\mathbf{f}(\mathbf{x}_0)\cdot(\mathbf{v}_{\sigma(1)},\dots,\mathbf{v}_{\sigma(\mathrm{r}_1)}),\dots,\\ &\qquad \mathbf{D}^{\mathrm{r}_\mathrm{j}}\mathbf{f}(\mathbf{x}_0)\cdot(\mathbf{v}_{\sigma(\mathrm{r}_1+\cdots+\mathrm{r}_{\mathrm{j}-1}+1)},\dots,\mathbf{v}_{\sigma(\mathrm{r})}))\end{aligned}$$

for $\mathbf{v}_1,\dots,\mathbf{v}_{\mathrm{r}} \in \mathbb{R}^{\mathrm{n}}$.

Proof The proof is by induction on r. For $r = 1$ the result is simply Theorem 1.4.49. Assume the result is true for $r \in \{1,\dots,s\}$ and let f and g be $s+1$ times differentiable at x_0. We thus have

$$\begin{aligned}\boldsymbol{D}^{s}(\boldsymbol{g}\circ\boldsymbol{f})(\boldsymbol{x}_0)\cdot(\boldsymbol{v}_2,\dots,\boldsymbol{v}_{s+1}) &= \sum_{j=1}^{s} \sum_{\substack{s_1,\dots,s_j\in\mathbb{Z}_{>0}\\ s_1+\cdots+s_j=s}} \sum_{\sigma\in\mathfrak{S}^{<}_{s_1,\dots,s_j}} \boldsymbol{D}^{j}\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}_0))\cdot(\boldsymbol{D}^{s_1}\boldsymbol{f}(\boldsymbol{x}_0)\cdot(\boldsymbol{v}_{\sigma(2)},\dots,\boldsymbol{v}_{\sigma(s_1+1)}),\dots,\\ &\qquad \boldsymbol{D}^{s_j}\boldsymbol{f}(\boldsymbol{x}_0)\cdot(\boldsymbol{v}_{\sigma(s_1+\cdots+s_{j-1}+2)},\dots,\boldsymbol{v}_{\sigma(s+1)}))\end{aligned}$$

for every $v_2,\dots,v_{s+1} \in \mathbb{R}^n$, and where $\sigma \in \mathfrak{S}^{<}_{s_1,\dots,s_j} \subseteq \mathfrak{S}_s$ permutes the set $\{2,\dots,s+1\}$ in the obvious way.

Let us now make an observation about permutations. Let $j' \in \{1,\dots,s+1\}$, let $s'_1,\dots,s'_{j'} \in \mathbb{Z}_{>0}$ satisfy $s'_1 + \cdots + s'_{j'} = s+1$, and let $\sigma' \in \mathfrak{S}^{<}_{s'_1,\dots,s'_{j'}}$. For brevity denote $t'_l = s'_1 + \cdots + s'_l$ for $l \in \{1,\dots,j'\}$. We have two cases.

1. $s'_1 = 1$: In this case let $j = j'-1$, define $s_l = s'_{l+1}$ for $l \in \{1,\dots,j'-1\}$, and let $t_l = s_1 + \cdots + s_l$ for $l \in \{1,\dots,j\}$. We then have

$$\begin{aligned}&((1),(\sigma'(t'_2-s'_2+1),\dots,\sigma'(t'_2)),\dots,(\sigma'(t'_{j'}-s'_{j'}+1),\dots,\sigma'(t'_{j'})))\\ &\qquad = ((1),(\sigma(t_2-s_2+1),\dots,\sigma(t_2)),\dots,(\sigma(t_{j'}-s_{j'}+1),\dots,\sigma(t_{j'}))), \quad (1.30)\end{aligned}$$

where $\sigma \in \mathfrak{S}^{<}_{s'_1,\dots,s'_j} \subseteq \mathfrak{S}_s$ permutes $\{2,\dots,s+1\}$ in the obvious way. Note that this uniquely specifies $s_1,\dots,s_j$ and σ.

2. $s'_1 \neq 1$: Here we take $j = j'$, $s_1 = s'_1 - 1$, $s_l = s'_l$ for $l \in \{2,\dots,j\}$. Let us denote $t_l = s_1 + \dots + s_l$ for $l \in \{1,\dots,j\}$. Then there exist $l_0 \in \{1,\dots,j\}$ giving the corresponding cycle $\tau \in \mathfrak{S}_j$ given by $\tau = (1 \ \cdots \ l_0)$ and $\sigma \in \mathfrak{S}_{s_{\tau(1)},s_{\tau(2)},\dots,s_{\tau(j)}}$ such that

$$\begin{aligned}((\sigma'(t'_1 - s'_1 + 1),\dots,\sigma'(t'_1)),\dots,(\sigma'(t'_{j'} - s'_{j'} + 1),\dots,\sigma'(t'_{j'})))\\ = ((1,\sigma(t_{\tau(1)} - s_{\tau(1)} + 1),\dots,\sigma(t_{\tau(1)})),\dots,(\sigma(t_{\tau(j)} - s_{\tau(j)} + 1),\dots,\sigma(t_{\tau(j)}))),\end{aligned} \tag{1.31}$$

where σ permutes $\{2,\dots,s+1\}$ in the obvious way. Note that this uniquely specifies $s_1,\dots,s_j$, τ, and σ. Note that the cycle τ is necessary to ensure that $\sigma'(1) = 1$, a necessary condition that $\sigma' \in \mathfrak{S}^{<}_{s'_1,\dots,s'_{j'}}$. The cycle serves to place the slot into which the "1" is inserted at the beginning of the slot list.

Conversely, let $j \in \{1,\dots,s\}$, let $s_1,\dots,s_j \in \mathbb{Z}_{>0}$ have the property that $s_1 + \dots + s_j = s$, and let $\sigma \in \mathfrak{S}^{<}_{s_1,\dots,s_k}$. Denote $t_l = s_1 + \dots + s_l$ for $l \in \{1,\dots,j\}$. Then we have two scenarios.

1. We take $j' = j+1$, let $s'_1 = 1$ and $s'_l = s_{l-1}$ for $l \in \{2,\dots,s+1\}$. Define $t_l = s_1 + \dots + s_l$. Then there exists $\sigma' \in \mathfrak{S}^{<}_{s'_1,\dots,s'_j}$ such that (1.30) holds. Moreover, this uniquely determines $s'_1,\dots,s'_{j'}$ and σ'.
2. We take $j = j'$ and let $l_0 \in \{1,\dots,j\}$. Then take $\tau \in \mathfrak{S}_j$ to be the cycle $(1 \ \cdots \ l_0)$. We then define $s'_1 = s_{\tau(1)} + 1$ and $s'_l = s_{\tau(l)}$ for $l \in \{2,\dots,j\}$. Then there exists $\sigma' \in \mathfrak{S}^{<}_{s'_1,\dots,s'_{j'}}$ such that (1.31) holds. Note that this uniquely specifies $s'_1,\dots,s'_{j'}$ and σ'.

Using this observation, along with Proposition 1.4.7, Theorems 1.4.49 and 1.4.50, and the symmetry of the derivatives of g of order up to s, we then compute

$$\begin{aligned}
&\boldsymbol{D}^{s+1}(\boldsymbol{g} \circ \boldsymbol{f})(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_1,\dots,\boldsymbol{v}_{s+1})\\
&= \sum_{j=1}^{s} \sum_{\substack{s_1,\dots,s_j \in \mathbb{Z}_{>0}\\ s_1+\dots+s_j=s}} \sum_{\sigma \in \mathfrak{S}^{<}_{s_1,\dots,s_j}} \boldsymbol{D}^{j+1}\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}_0)) \cdot (\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}_0) \cdot \boldsymbol{v}_1,\\
&\quad \boldsymbol{D}^{s_1}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(2)},\dots,\boldsymbol{v}_{\sigma(s_1+1)}),\dots,\\
&\quad \boldsymbol{D}^{s_j}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(s_1+\dots+s_{j-1}+2)},\dots,\boldsymbol{v}_{\sigma(s+1)}))\\
&\quad + \boldsymbol{D}^{j}\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}_0)) \cdot (\boldsymbol{D}^{s_1+1}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_1,\boldsymbol{v}_{\sigma(2)},\dots,\boldsymbol{v}_{\sigma(s_1+1)}),\dots,\\
&\quad \boldsymbol{D}^{s_j}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(s_1+\dots+s_{j-1}+2)},\dots,\boldsymbol{v}_{\sigma(s+1)})) + \dots\\
&\quad + \boldsymbol{D}^{j}\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}_0)) \cdot (\boldsymbol{D}^{s_1}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma(2)},\dots,\boldsymbol{v}_{\sigma(s_1+1)}),\dots,\\
&\quad \boldsymbol{D}^{s_j}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_1,\boldsymbol{v}_{\sigma(s_1+\dots+s_{j-1}+2)},\dots,\boldsymbol{v}_{\sigma(s+1)}))\\
&= \sum_{j'=1}^{s+1} \sum_{\substack{s'_1,\dots,s'_{j'} \in \mathbb{Z}_{>0}\\ s'_1+\dots+s'_{j'}=s+1}} \sum_{\sigma' \in \mathfrak{S}^{<}_{s'_1,\dots,s'_{j'}}} \boldsymbol{D}^{j'}\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}_0)) \cdot (\boldsymbol{D}^{s'_1}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma'(1)},\dots,\boldsymbol{v}_{\sigma'(s'_1)}),\\
&\quad \dots,\boldsymbol{D}^{s'_{j'}}\boldsymbol{f}(\boldsymbol{x}_0) \cdot (\boldsymbol{v}_{\sigma'(s'_1+\dots+s'_{j'-1}+1)},\dots,\boldsymbol{v}_{\sigma'(s+1)})),
\end{aligned}$$

as desired. ■

Let us parse the formula of the preceding result in the case where $r = 2$. We denote the components of $\boldsymbol{f}$ by $f_1, \dots, f_m$ and the components of $\boldsymbol{g}$ by $g_1, \dots, g_k$. The components of $\boldsymbol{D}^2(\boldsymbol{g} \circ \boldsymbol{f})(\boldsymbol{x})$ are

$$\sum_{a,b=1}^{m} \frac{\partial^2 g_\alpha(\boldsymbol{f}(\boldsymbol{x}))}{\partial y_a \partial y_b} \frac{\partial f_a(\boldsymbol{x})}{\partial x^i} \frac{\partial f_b(\boldsymbol{x})}{\partial x_j} + \sum_{a=1}^{m} \frac{\partial g_\alpha(\boldsymbol{f}(\boldsymbol{x}))}{\partial y_a} \frac{\partial^2 f_a(\boldsymbol{x})}{\partial x_i \partial x_j},$$
$$\alpha \in \{1, \dots, k\},\ i, j \in \{1, \dots, n\}.$$

Of course, if you are familiar with how the Chain Rule and the product rule work, this is exactly the formula you would produce. In Exercise 1.4.5 we ask the reader to directly parse the formula in the theorem in the case when $r = 3$.

1.4.10 Notes

We refer to [Abraham, Marsden, and Ratiu 1988, Chapter 2] for a thorough presentation of multivariable calculus, including definitions of higher-order derivatives, the general version of Taylor's Theorem, the Inverse Function Theorem in the multivariable case, the multivariable Chain Rule, and much more. We comment that the approach in [Abraham, Marsden, and Ratiu 1988] also extends the presentation from the multivariable case to the infinite-dimensional case, and that this is important in some applications.

The proof we give of Theorem 1.4.44 follows the excellent presentation of McShane [1973]. The companion second-derivative test, Theorem 1.4.46, has hypotheses the checking of which has caused many papers to be written. The most common technique is that of "bordered Hessians" introduced by Mann [1943] and reiterated, for example, by Spring [1985].

Exercises

1.4.1 Let $\mathsf{L} \in S^k(\mathbb{R}^n; \mathbb{R}^m)$ and define $\boldsymbol{f}_{\mathsf{L}} \colon \mathbb{R}^n \to \mathbb{R}^m$ by $\boldsymbol{f}_{\mathsf{L}}(\boldsymbol{x}) = \mathsf{L}(\boldsymbol{x}, \dots, \boldsymbol{x})$. Show that for $r \in \{1, \dots, k\}$ we have

$$\boldsymbol{D}^r \boldsymbol{f}_{\mathsf{L}}(\boldsymbol{x}) \cdot (\boldsymbol{v}_1, \dots, \boldsymbol{v}_r) = \frac{k!}{(k-r)!} \mathsf{L}(\boldsymbol{x}, \dots, \boldsymbol{x}, \boldsymbol{v}_1, \dots, \boldsymbol{v}_r).$$

1.4.2 Consider the map $f \colon \mathbb{R}^n \to \mathbb{R}$ given by $f(\boldsymbol{x}) = \|\boldsymbol{x}\|^2_{\mathbb{R}^n}$.

(a) Give explicit and attractive formulae for $\boldsymbol{D}f(\boldsymbol{x}) \cdot \boldsymbol{v}$ and $\boldsymbol{D}^2 f(\boldsymbol{x}) \cdot (\boldsymbol{v}_1, \boldsymbol{v}_2)$ for $\boldsymbol{x}, \boldsymbol{v}, \boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^n$.

(b) Show that $\boldsymbol{D}^j f(\boldsymbol{x}) = \boldsymbol{0}$ for $\boldsymbol{x} \in \mathbb{R}^n$ and $j \geq 3$.

1.4.3 Let $U \subseteq \mathbb{R}^n$ be an open set and let $\boldsymbol{f} \colon U \to \mathbb{R}^m$ be differentiable at $\boldsymbol{x}_0 \in U$. Show that $\boldsymbol{D}_j \boldsymbol{f}(\boldsymbol{x}_0) = \boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}_0; \boldsymbol{e}_j)$ for each $j \in \{1, \dots, n\}$.

1.4.4 Expand the formula of Theorem 1.4.50 in the case of $r = k = 3$.

1.4.5 Expand the formula of Theorem 1.4.51 in the case of $r = 3$.

Section 1.5

The rôle of the rank of the derivative

One of the most important features of the derivative is that its rank can often be used to give make surprisingly strong conclusions about the character of the function. In this section we shall explore this in some depth. The central result here is the Inverse Function Theorem, which can thus be thought of as one of the more important results in analysis.

Do I need to read this section? The material in this section is important, but maybe not so important for a lot of what we shall do. Thus it can be skipped until needed if the reader is taking the streamlined approach. •

1.5.1 Critical and regular points and values

To understand the local behaviour of maps between Euclidean spaces it is important to understand the rôle of the rank of the derivative. In this section we organise the language needed to do this and prove a simple version of an important result known as Sard's Theorem.

1.5.1 Definition (Regular point, critical point, regular value, critical value) Let $U \subseteq \mathbb{R}^n$ be open and let $f\colon U \to \mathbb{R}^m$ be of class C^1.

(i) A point $x \in U$ is a ***regular point*** for f if $Df(x)$ is surjective.

(ii) A point $x \in U$ is a ***critical point*** for f if it is not a regular point.

(iii) A point $y \in \mathbb{R}^m$ is a ***regular value*** for f if the set $f^{-1}(y)$ is comprised solely of regular points.

(iv) A point $y \in \mathbb{R}^m$ is a ***critical value*** for f if it is not a regular value. •

1.5.2 Inverse Function Theorem and consequences

1.5.2 Theorem (Inverse Function Theorem)

1.5.3 Theorem (Implicit Function Theorem)

1.5.3 Constant Rank Theorem

1.5.4 Sard's Theorem

1.5.5 Notes

Section 1.6

The multivariable Riemann integral

In this section we generalise the construction of the Riemann integral given in Section I-3.4 to functions defined on subsets of $\mathbb{R}^n$. Many of the constructions and results follow in a manner similar to those for the Riemann integral for compact intervals, so we do not spend as much time with motivation or explanation. Also, we may occasionally abbreviate some of the arguments in proofs since they are straightforward generalisations of those given in Section I-3.4.

One of the less trivial things we do in this section is give a complete statement and proof of Stokes' Theorem in $\mathbb{R}^n$ for regions with piecewise smooth boundary. This statement and proof are not so easily located in the literature, many texts sticking to proofs, and sometimes statements, valid only in special cases. However, in applications one often wants the general case, and so we give this here.

Do I need to read this section? This material in this section is not much used, and so can certainly be omitted on a first reading. •

1.6.1 Step functions

As with the Riemann integral for functions on the real line, the key idea in defining the Riemann integral over regions in $\mathbb{R}^n$ is to give a class of functions we use to approximate arbitrary functions. We call these step functions, even thought they are a little less like "step" functions than one might like. But what's in a name?

First let us define the notion of a step function. We recall from Section 1.2.3 the notion of a rectangle in $\mathbb{R}^n$.

1.6.1 Definition (Step function) Let $R = [a_1, b_1]\times\cdots\times[a_n, b_n] \subseteq \mathbb{R}^n$ be a compact rectangle. A function $f\colon R \to \mathbb{R}$ is a ***step function*** if there exists a partition $\boldsymbol{P} = (P_1, \dots, P_n)$ of R such that

(i) $f|\operatorname{int}(R_{l_1,\dots,l_n})$ is a constant function for each $l_j \in \{1, \dots, k_j\}$, $k_j \in \{1, \dots, n\}$,

(ii) $f(x_1, \dots, x_{j-1}, a_j+, x_{j+1}, \dots, x_n) = f(x_1, \dots, x_{j-1}, a_j, x_{j+1}, \dots, x_n)$ for $x_l \in [a_l, b_l]$, $l \in \{1, \dots, n\} \setminus \{j\}$, $j \in \{1, \dots, n\}$,

(iii) $f(x_1, \dots, x_{j-1}, b_j-, x_{j+1}, \dots, x_n) = f(x_1, \dots, x_{j-1}, b_j, x_{j+1}, \dots, x_n)$ for $x_l \in [a_l, b_l]$, $l \in \{1, \dots, n\} \setminus \{j\}$, $j \in \{1, \dots, n\}$, and

(iv) for each $j \in \{1, \dots, n\}$ and $x \in \operatorname{EP}(P_j)$, either

  (a) $f(x_1, \dots, x_{j-1}, x+, x_{j+1}, \dots, x_n) = f(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n)$ or

  (b) $f(x_1, \dots, x_{j-1}, x-, x_{j+1}, \dots, x_n) = f(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n)$

  for each $x_l \in [a_l, b_l]$, $l \in \{1, \dots, n\} \setminus \{j\}$. •

The idea is the same as for step functions on $\mathbb{R}$ in that we do not care what value is assumed by the function at the boundaries of the sets defining the partitions. To define the integral of a step function we need the following simple idea.

1.6.2 Definition (Volume of a rectangle) The ***volume*** of a compact rectangle $R = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$ is

$$\mathrm{vol}(R) = (b_1 - a_1) \cdot \cdots \cdot (b_n - a_n). \qquad \bullet$$

The integral of a step function is then obviously as follows.

1.6.3 Definition (Riemann integral of a step function) Let $R \subseteq \mathbb{R}^n$ be a fat rectangle and let $\boldsymbol{P} = (P_1, \dots, P_n)$ be a partition of R with $P_j = (I_1, \dots, I_{k_j})$. If $f\colon R \to \mathbb{R}$ is a step function such that $f|R_{l_1,\dots,l_n} = c_{l_1,\dots,l_n}$, then the ***Riemann integral*** of f is

$$A(f) = \sum_{l_1=1}^{k_1} \cdots \sum_{l_n=1}^{k_n} c_{l_1,\dots,c_n} \mathrm{vol}(R_{l_1,\dots,l_n}). \qquad \bullet$$

1.6.2 The Riemann integral on bounded sets

The initial definition of the Riemann integral is made for bounded functions $f\colon A \to \mathbb{R}$ defined on a bounded subset $A \subseteq \mathbb{R}^n$. Since A is bounded there exists a rectangle R such that $A \subseteq R$. Therefore, by taking f to be zero outside A, we may consider f as a function from R to $\mathbb{R}$. For this reason we consider in this section the Riemann integral of bounded functions $f\colon R \to \mathbb{R}$ defined on a rectangle $R \subseteq \mathbb{R}^n$. defined? defined?

We begin by associating a step function to a function on a rectangle and a partition of that rectangle.

1.6.4 Definition (Lower and upper step functions) Let $R = [a_1, b_1] \times \cdots \times [a_n, b_n] \subseteq \mathbb{R}^n$ be a fat rectangle, let $f\colon R \to \mathbb{R}$ be a bounded function, and let $\boldsymbol{P} = (P_1, \dots, P_n)$ be a partition of R.

(i) The ***lower step function*** associated to f and $\boldsymbol{P}$ is the function $s_-(f, \boldsymbol{P})\colon R \to \mathbb{R}$ defined according to the following:

(a) if $\boldsymbol{x} \in R$ lies in the interior of a subrectangle $R_{l_1,\dots,l_n}$, $l_j \in \{1, \dots, k_j\}$, $j \in \{1, \dots, n\}$, then $s_-(f, P)(\boldsymbol{x}) = \inf\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathrm{cl}(R_{l_1,\dots,l_n})\}$;

(b) $s_-(f, \boldsymbol{P})(x_1, \dots, x_{j-1}, a_j+, x_{j+1}, \dots, x_n) = s_-(f, \boldsymbol{P})(x_1, \dots, x_{j-1}, a_j, x_{j+1}, \dots, x_n)$ for $x_l \in [a_l, b_l]$, $l \in \{1, \dots, n\} \setminus \{j\}$, $j \in \{1, \dots, n\}$;

(c) $s_-(f, \boldsymbol{P})(x_1, \dots, x_{j-1}, b_j-, x_{j+1}, \dots, x_n) = s_-(f, \boldsymbol{P})(x_1, \dots, x_{j-1}, b_j, x_{j+1}, \dots, x_n)$ for $x_l \in [a_l, b_l]$, $l \in \{1, \dots, n\} \setminus \{j\}$, $j \in \{1, \dots, n\}$;

(d) for each $j \in \{1, \dots, n\}$ and $x \in \mathrm{EP}(P_j)$,
$s_-(f, \boldsymbol{P})(x_1, \dots, x_{j-1}, x, x_{j+1}, \dots, x_n) = s_-(f, \boldsymbol{P})(x_1, \dots, x_{j-1}, x+, x_{j+1}, \dots, x_n)$ for each $x_l \in [a_l, b_l]$, $l \in \{1, \dots, n\} \setminus \{j\}$.

(ii) The ***upper step function*** associated to f and $\boldsymbol{P}$ is the function $s_+(f, \boldsymbol{P})\colon R \to \mathbb{R}$ defined according to the following:

(a) if $x \in R$ lies in the interior of a subrectangle $R_{l_1,\dots,l_n}$, $l_j \in \{1,\dots,k_j\}$, $j \in \{1,\dots,n\}$, then $s_+(f,P)(x) = \sup\{f(x) \mid x \in \mathrm{cl}(R_{l_1,\dots,l_n})\}$;

(b) $s_+(f,\boldsymbol{P})(x_1,\dots,x_{j-1},a_j+,x_{j+1},\dots,x_n) = s_+(f,\boldsymbol{P})(x_1,\dots,x_{j-1},a_j,x_{j+1},\dots,x_n)$ for $x_l \in [a_l,b_l]$, $l \in \{1,\dots,n\} \setminus \{j\}$, $j \in \{1,\dots,n\}$;

(c) $s_+(f,\boldsymbol{P})(x_1,\dots,x_{j-1},b_j-,x_{j+1},\dots,x_n) = s_+(f,\boldsymbol{P})(x_1,\dots,x_{j-1},b_j,x_{j+1},\dots,x_n)$ for $x_l \in [a_l,b_l]$, $l \in \{1,\dots,n\} \setminus \{j\}$, $j \in \{1,\dots,n\}$;

(d) for each $j \in \{1,\dots,n\}$ and $x \in \mathrm{EP}(P_j)$, $s_+(f,\boldsymbol{P})(x_1,\dots,x_{j-1},x,x_{j+1},\dots,x_n) = s_+(f,\boldsymbol{P})(x_1,\dots,x_{j-1},x+,x_{j+1},\dots,x_n)$ for each $x_l \in [a_l,b_l]$, $l \in \{1,\dots,n\} \setminus \{j\}$. •

The tedium required to make the construction unambiguous does nothing to hide the fact that the step functions $s_-(f,\boldsymbol{P})$ and $s_+(f,\boldsymbol{P})$ are, in essence, the "smallest" and "largest" step functions that bound f from above and below, respectively. Thus

$$s_-(f,\boldsymbol{P})(\boldsymbol{x}) \le f(\boldsymbol{x}) \le s_+(f,\boldsymbol{P})(\boldsymbol{x})$$

for every $\boldsymbol{x} \in R$.

Then we have the following definitions.

1.6.5 Definition (Lower and upper Riemann sums) Let $R \subseteq \mathbb{R}^n$ be a fat compact rectangle, let $f\colon R \to \mathbb{R}$ be a bounded function, and let $\boldsymbol{P} = (P_1,\dots,P_n)$ be a partition of R.

(i) The ***lower Riemann sum*** associated to f and $\boldsymbol{P}$ is $A_-(f,\boldsymbol{P}) = A(s_-(f,\boldsymbol{P}))$.

(ii) The ***upper Riemann sum*** associated to f and $\boldsymbol{P}$ is $A_+(f,\boldsymbol{P}) = A(s_+(f,\boldsymbol{P}))$. •

1.6.6 Definition (Lower and upper Riemann integral) Let $R \subseteq \mathbb{R}^n$ be a fat compact rectangle, let $f\colon R \to \mathbb{R}$ be a bounded function.

(i) The ***lower Riemann integral*** of f is

$$I_-(f) = \sup\{A_-(f,\boldsymbol{P}) \mid \boldsymbol{P} \in \mathrm{Part}(I)\}.$$

(ii) The ***upper Riemann integral*** of f is

$$I_+(f) = \inf\{A_+(f,\boldsymbol{P}) \mid \boldsymbol{P} \in \mathrm{Part}(I)\}.$$ •

As with the Riemann integral in $\mathbb{R}$, the lower and upper Riemann integrals exist since f is assumed to be bounded. The interesting case is when they agree.

1.6.7 Definition (Riemann integrable of a bounded function on a compact rectangle) A bounded function $f\colon R \to \mathbb{R}$ on a fat compact rectangle is ***Riemann integrable*** if $I_-(f) = I_+(f)$. We denote

$$\int_R f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = I_-(f) = I_+(f),$$

which is the ***Riemann integral*** of f. The function f is called the ***integrand***. •

It is also possible to adapt this notion to define the integral over general subsets.

1.6.8 Definition (Riemann integral for bounded functions defined on bounded sets) Let $A \subseteq \mathbb{R}^n$ be bounded and let $f: A \to \mathbb{R}$ be bounded. The function f is ***Riemann integrable*** if, for some fat rectangle R such that $A \subseteq R$ and for $f_R: R \to \mathbb{R}$ defined by

$$f_R(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}), & \boldsymbol{x} \in A, \\ 0, & \boldsymbol{x} \notin A, \end{cases}$$

the function f_R is Riemann integrable. The ***Riemann integral*** of f is

$$\int_A f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_R f_R(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}. \qquad \bullet$$

Of course, a possible problem arises here that the definition may depend on the rectangle R one chooses to contain the subset A. We do not show here that this is not the case; it is a straightforward application of the definition of the Riemann integral, possibly (depending on how one goes about the proof) along with Theorem 1.6.10 below.

1.6.9 Notation (Riemann integral of restrictions) In the construction of Definition 1.6.8 we extended f from A to a fat compact rectangle R, denoting the extension by f_R. We would like to not have to repeat this notation involving the extension. Therefore, we adopt the convention that if $A \subseteq \mathbb{R}^n$ is a bounded set, if $f: A \to \mathbb{R}$ is a Riemann integrable function, and if $B \subseteq A$, then we denote by

$$\int_B f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

the Riemann integral of the function that has the value of f on B and zero otherwise, when this integral exists (it may not, for some choices of B). We may also say that this is the Riemann integral of $f|B$. The point is that we do not want to always introduce tedious notation for the function that is equal to f on B and is zero otherwise. This also makes it easier to simply define functions with domain equal to a fat compact rectangle containing the set A on which the function is actually defined. •

1.6.3 Characterisations of Riemann integrable functions on compact sets

As with the Riemann integral of a single variable, it is possible to give multiple equivalent interpretations of Riemann integrable functions. Some of these are more or less useful, depending on the context.

The first theorem gives four equivalent "$\epsilon - \delta$" style characterisations of a Riemann integrable function. To state the result we need the following terminology. Let $\boldsymbol{P} = (P_1, \dots, P_n)$ be a partition of a fat compact rectangle $R \subseteq \mathbb{R}^n$ and denote $P_j = (I_{j1}, \dots, I_{jk_j})$, $j \in \{1, \dots, n\}$. A ***selection*** from $\boldsymbol{P}$ is an family

$$(\boldsymbol{\xi}_{l_1,\dots,l_n} \mid \; l_j \in \{1, \dots, k_j\},\ j \in \{1, \dots, n\})$$

of points from R such that $\boldsymbol{\xi}_{l_1,\dots,l_n} \in \mathrm{cl}(R_{l_1,\dots,l_n})$.

1.6.10 Theorem (Riemann, Darboux, and Cauchy characterisations of Riemann integrable functions) *For a fat compact rectangle* $R \subseteq \mathbb{R}^n$ *and a bounded function* $f\colon R \to \mathbb{R}$*, the following statements are equivalent:*

(i) f *is Riemann integrable;*

(ii) for every $\epsilon \in \mathbb{R}_{>0}$*, there exists a partition* $\mathbf{P}$ *such that* $A_+(f,\mathbf{P}) - A_-(f,\mathbf{P}) < \epsilon$ *(**Riemann's condition**);*

(iii) there exists $I(f) \in \mathbb{R}$ *such that, for every* $\epsilon \in \mathbb{R}_{>0}$ *there exists* $\delta \in \mathbb{R}_{>0}$ *such that, if* $\mathbf{P} = (P_1,\dots,P_k)$ *is a partition for which* $|P| < \delta$ *and if* $(\xi_{l_1,\dots,l_n} \mid l_j \in \{1,\dots,k_j\},\ j \in \{1,\dots,n\})$ *is a selection from* $\mathbf{P}$*, then*

$$\left|\sum_{l_1=1}^{k_1}\cdots\sum_{l_n=1}^{k_n} f(\xi_{l_1,\dots,l_n})\mathrm{vol}(R_{l_1,\dots,l_n}) - I(f)\right| < \epsilon,$$

where $P_j = (I_{j1},\dots,I_{nk_n})$*,* $j \in \{1,\dots,n\}$ *(**Darboux' condition**);*

(iv) for each $\epsilon \in \mathbb{R}_{>0}$ *there exists* $\delta \in \mathbb{R}_{>0}$ *such that, for any partitions* $\mathbf{P} = (P_1,\dots,P_n)$ *and* $\mathbf{P}' = (P'_1,\dots,P'_n)$ *with* $|\mathbf{P}|, |\mathbf{P}'| < \delta$ *and for any selections* $(\xi_{l_1,\dots,l_n} \mid l_j \in \{1,\dots,k_j\},\ j \in \{1,\dots,n\})$ *and* $(\xi'_{l_1,\dots,l_n} \mid l_j \in \{1,\dots,k_j\},\ j \in \{1,\dots,n\})$ *from* $\mathbf{P}$ *and* $\mathbf{P}'$*, respectively, we have*

$$\left|\sum_{l_1=1}^{k_1}\cdots\sum_{l_n=1}^{k_n} f(\xi_{l_1,\dots,l_n})\mathrm{vol}(R_{l_1,\dots,l_n}) - \sum_{l_1=1}^{k'_1}\cdots\sum_{l_n=1}^{k'_n} f(\xi'_{l_1,\dots,l_n})\mathrm{vol}(R'_{l_1,\dots,l_n})\right| < \epsilon,$$

where $P_j = (I_{j1},\dots,I_{nk_n})$ *and* $P'_j = (I'_{1k'_1},\dots,I'_{nk'_n})$ *(**Cauchy's condition**).*

Proof We begin by noting that Lemmata I-1 and I-2 of Theorem I-3.4.9 hold for rectangles, and their proofs are identical, *mutatis mutandis*, to those for intervals.

(i) $\Longrightarrow$ (ii) The argument here is identical to that for Theorem I-3.4.9.

(ii) $\Longrightarrow$ (i) Again, the argument is identical to that in Theorem I-3.4.9.

(i) $\Longrightarrow$ (iii) Here we prove a lemma analogous to Lemma I-3 from Theorem I-3.4.9.

1 Lemma *If* $\mathbf{P} = (P_1,\dots,P_n)$ *is a partition of a fat compact rectangle* $R \subseteq \mathbb{R}^n$ *and if* $\epsilon \in \mathbb{R}_{>0}$*, then there exists* $\delta \in \mathbb{R}_{>0}$ *such that, if* $\mathbf{P} = (P'_1,\dots,P'_n)$ *is a partition of* R *with* $|\mathbf{P}| < \delta$ *and if*

$$L_1 = \{(l'_1,\dots,l'_n) \in \{1,\dots,k'_1\}\times\cdots\times\{1,\dots,k'_n\} \mid \mathrm{cl}(R_{l'_1,\dots,l'_n}) \not\subset \mathrm{cl}(R_{l_1,\dots,l_n}) \text{ for any } (l_1,\dots,l_n) \in \{1,\dots,k_1\}\times\cdots\times\{1,\dots,k_n\}\},$$

then

$$\sum_{(l'_1,\dots,l'_n)\in L_1} \mathrm{vol}(R_{l'_1,\dots,l'_n}) < \epsilon,$$

where $P_j = (I_{j1},\dots,I_{jk_j})$ *and* $P'_j = (I'_{j1},\dots,I'_{jk'_n}))$ *for* $j \in \{1,\dots,n\}$*.*

Proof Suppose that $R = [a_1,b_1]\times\cdots\times[a_n,b_n]$ and define

$$A_{\mathbf{P}} = (k_1+1)(b_2-a_2)(b_3-a_3)\cdots(b_n-a_n)\cdot(k_2+1)(b_1-a_1)(b_3-a_3)\cdots\cdot(b_n-a_n)\cdots\cdot(k_n+1)(b_1-a_1)(b_2-a_2)\cdots(b_{n-1}-a_{n-1}),$$

noting that $A_{\boldsymbol{P}}$ is the "area" of all of the $(n-1)$-dimensional faces of the subrectangles of $\boldsymbol{P}$. Let $\epsilon \in \mathbb{R}_{>0}$ and define $\delta = \frac{\epsilon}{A_{\boldsymbol{P}}}$. Let $\boldsymbol{P}' = (P'_1, \dots, P'_n)$ be a partition with $P'_j = (I'_{j1}, \dots, I'_{jk'_j})$ and satisfying $|\boldsymbol{P}'| < \delta$. Define L_1 as in the statement of the lemma and let $(l'_1, \dots, l'_n) \in L_1$. Since $R_{l'_1,\dots,l'_n}$ is not contained in any subrectangle of $\boldsymbol{P}$ it must be the case that $R_{l'_1,\dots,l'_n}$ intersects the $(n-1)$-dimensional face of at least one of the subrectangles of $\boldsymbol{P}$. This means that the total volume of such subrectangles from $\boldsymbol{P}'$ will be at most δ times the total area of the $(n-1)$-dimensional faces of the subrectangles from $\boldsymbol{P}$. That is,

$$\sum_{(l'_1,\dots,l'_n)\in L_1} \mathrm{vol}(R_{l'_1,\dots,l'_n}) \le A_{\boldsymbol{P}}\delta \le \epsilon,$$

as desired. ▼

Now let $\epsilon \in \mathbb{R}_{>0}$ and define $M = \sup\{|f(x)| \mid \boldsymbol{x} \in R\}$. Denote by $I(f)$ the Riemann integral of f. Choose partitions $\boldsymbol{P}_- = (P_{-,1}, \dots, P_{-,n})$ and $\boldsymbol{P}_+ = (P_{+,1}, \dots, P_{+,n})$ such that

$$I(f) - A_-(f, \boldsymbol{P}_-) < \tfrac{\epsilon}{2}, \quad A_+(f, \boldsymbol{P}_+) - I(f) < \tfrac{\epsilon}{2}.$$

If $\boldsymbol{P} = (P_1, \dots, P_n)$ with $P_j = (I_{j1}, \dots, I_{jk_j})$, $j \in \{1, \dots, n\}$, is chosen such that $\mathrm{EP}(P_j) = \mathrm{EP}(P_{-,j}) \cup \mathrm{EP}(P_{+,j})$, $j \in \{1, \dots, n\}$, then

$$I(f) - A_-(f, \boldsymbol{P}) < \tfrac{\epsilon}{2}, \quad A_+(f, \boldsymbol{P}) - I(f) < \tfrac{\epsilon}{2}.$$

By the lemma above choose $\delta \in \mathbb{R}_{>0}$ such that if $\boldsymbol{P}'$ is any partition for which $|\boldsymbol{P}'| < \delta$ then the sum of the volumes of the subrectangles of $\boldsymbol{P}'$ not contained in some subrectangle of $\boldsymbol{P}$ does not exceed $\frac{\epsilon}{2M}$. Let $\boldsymbol{P}' = (P'_1, \dots, P'_n)$ be a partition with $P'_j = (I'_{j1}, \dots, I'_{jk'_j})$, $j \in \{1, \dots, n\}$, and satisfying $|\boldsymbol{P}'| < \delta$. Denote by L_1 the subset defined in the lemma and take

$$L_2 = (\{1, \dots, k'_1\} \times \dots \times \{1, \dots, k'_n\}) \setminus L_1.$$

Let $(\xi_{l_1,\dots,l_n} \mid \; l_j \in \{1, \dots, k'_j\},\; j \in \{1, \dots, n\})$ be a selection of P'. Then we compute

$$\begin{aligned}\sum_{l_1=1}^{k'_1} \cdots \sum_{l_n=1}^{k'_n} f(\xi_{l_1,\dots,l_n}) \mathrm{vol}(R_{l_1,\dots,l_n}) &= \sum_{(l_1,\dots,l_n)\in L_1} f(\xi_{l_1,\dots,l_n}) \mathrm{vol}(R_{l_1,\dots,l_n}) \\ &\quad + \sum_{(l_1,\dots,l_n)\in L_2} f(\xi_{l_1,\dots,l_n}) \mathrm{vol}(R_{l_1,\dots,l_n}) \\ &\le A_+(f, P) + M\frac{\epsilon}{2M} < I(f) + \epsilon.\end{aligned}$$

In like manner we show that

$$\sum_{l_1=1}^{k'_1} \cdots \sum_{l_n=1}^{k'_n} f(\xi_{l_1,\dots,l_n}) \mathrm{vol}(R_{l_1,\dots,l_n}) > I(f) - \epsilon.$$

This gives

$$\Big|\sum_{l_1=1}^{k'_1} \cdots \sum_{l_n=1}^{k'_n} f(\xi_{l_1,\dots,l_n}) \mathrm{vol}(R_{l_1,\dots,l_n}) - I(f)\Big| < \epsilon,$$

as desired.

(iii) $\Longrightarrow$ (ii) Let $\epsilon \in \mathbb{R}_{>0}$ and let $\boldsymbol{P} = (P_1, \dots, P_n)$ be a partition with $P_j = (I_{j1}, \dots, I_{jk_j})$, $j \in \{1, \dots, n\}$, and for which

$$\Big|\sum_{l_1=1}^{k_1} \cdots \sum_{l_n=1}^{k_n} f(\xi_{l_1,\dots,l_n})\mathrm{vol}(R_{l_1,\dots,l_n}) - I(f)\Big| < \tfrac{\epsilon}{4}$$

for every selection $(\xi_{l_1,\dots,l_n} \mid l_j \in \{1, \dots, k_j\},\ j \in \{1, \dots, n\})$ from $\boldsymbol{P}$. Now particularly choose a selection such that

$$|f(\xi_{l_1,\dots,l_n}) - \sup\{f(x) \mid x \in \mathrm{cl}(R_{l_1,\dots,l_n})\}| < \frac{\epsilon}{4k_1 \cdots k_n \mathrm{vol}(R_{l_1,\dots,l_n})}.$$

Then

$$\begin{aligned}|A_+(f,\boldsymbol{P}) - I(f)| \le{}& \Big|A_+(f,\boldsymbol{P}) - \sum_{l_1=1}^{k_1} \cdots \sum_{l_n=1}^{k_n} f(\xi_{l_1,\dots,l_n})\mathrm{vol}(R_{l_1,\dots,l_n})\Big| \\ &+ \Big|\sum_{l_1=1}^{k_1} \cdots \sum_{l_n=1}^{k_n} f(\xi_{l_1,\dots,l_n})\mathrm{vol}(R_{l_1,\dots,l_n}) - I(f)\Big| \\ <{}& \sum_{l_1=1}^{k_1} \cdots \sum_{l_n=1}^{k_n} \frac{\epsilon}{4k_1 \cdots k_n \mathrm{vol}(R_{l_1,\dots,l_n})}\mathrm{vol}(R_{l_1,\dots,l_n}) + \tfrac{\epsilon}{4} < \tfrac{\epsilon}{2}.\end{aligned}$$

In like manner one shows that $|A_-(f,\boldsymbol{P}) - I(f)| < \frac{\epsilon}{2}$. Therefore,

$$|A_+(f,\boldsymbol{P}) - A_-(f,\boldsymbol{P})| \le |A_+(f,\boldsymbol{P}) - I(f)| + |I(f) - A_-(f,\boldsymbol{P})| < \epsilon,$$

as desired.

(iii) $\Longrightarrow$ (iv) The proof goes like the analogous part of the proof of Theorem I-3.4.9.

(iv) $\Longrightarrow$ (iii) The proof here is a straightforward adaptation of the notation from the corresponding part of Theorem I-3.4.9. ■

The next characterisation of Riemann integrable functions we give is the more deep one, and generalises Theorem I-3.4.11.

1.6.11 Theorem (Riemann integrable functions are continuous almost everywhere, and vice versa) *For a bounded set $A \subseteq \mathbb{R}^n$, a bounded function $f\colon A \to \mathbb{R}$ is Riemann integrable if and only if the set*

$$D_f = \{\mathbf{x} \in \mathbb{R}^n \mid f \text{ is discontinuous at } \mathbf{x}\}$$

has measure zero.

Proof Recall the definition of the oscillation of a function from Definition 1.3.10 and the fact, from Proposition 1.3.11, that a function is continuous at $\boldsymbol{x}$ if and only if its oscillation at $\boldsymbol{x}$ is zero.

We let $R \subseteq \mathbb{R}^n$ be a fat compact rectangle containing A and let f denote the extension of f from A to R by asking that f take value zero off A.

Suppose that D_f has measure zero and for $\epsilon \in \mathbb{R}_{>0}$ define

$$D_{f,\epsilon} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \omega_f(\boldsymbol{x}) \geq \epsilon\}.$$

We claim that $D_{f,\epsilon}$ is compact. Indeed, suppose that $\boldsymbol{x} \in \mathbb{R}^n$ is an accumulation point for $D_{\epsilon,f}$. By Proposition III-3.6.8 it follows that $\mathsf{B}^n(r, \boldsymbol{x}) \cap D_{\epsilon,f} \neq \emptyset$. Thus, by definition of $D_{f,\epsilon}$,

$$\sup\{|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)| \mid \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathsf{B}^n(r, \boldsymbol{x})\} \geq \epsilon.$$

Therefore, $\omega_f(\boldsymbol{x}) \geq \epsilon$ and so $\boldsymbol{x} \in D_{f,\epsilon}$. Therefore, $D_{f,\epsilon}$ is closed. Since $D_{f,\epsilon} \subseteq R$ and since R is compact, it follows from that $D_{f,\epsilon}$ is compact for any $\epsilon \in \mathbb{R}_{>0}$. what

Now let $\epsilon \in \mathbb{R}_{>0}$ and take $\epsilon' = \frac{\epsilon}{2\mathrm{vol}(R)}$. Since $D_{f,\epsilon'} \subseteq D_f$ and since D_f has measure zero, $D_{f,\epsilon'}$ has measure zero. From this, and since open balls and open rectangles with equal length sides have volumes that differ only by a multiplicative constant, it follows that there exists a family $(R_j)_{j\in\mathbb{Z}_{>0}}$ of open rectangles such that $D_{f,\epsilon'} \subseteq \cup_{j\in\mathbb{Z}_{>0}} R_j$ and

$$\sum_{j=1}^{\infty} \mathrm{vol}(R_j) < \frac{\epsilon}{4M},$$

where $M = \sup\{|f(\boldsymbol{x})| \mid \boldsymbol{x} \in R\}$. Since $D_{f,\epsilon'}$ is compact there exists a finite subset $j_1, \dots, j_r \in \mathbb{Z}_{>0}$ such that $D_{f,\epsilon'} = \cup_{k=1}^r R_{j_k}$.

Now let $\boldsymbol{P} = (P_1, \dots, P_n)$ be a partition of R and denote $P_j = (I_{j1}, \dots, I_{jk_j})$, $j \in \{1, \dots, n\}$. Denote

$$L_1 = \{(l_1, \dots, l_n) \in \{1, \dots, k_1\} \times \dots \times \{1, \dots, k_n\} \mid R_{l_1,\dots,l_n} \subseteq R_{j_k} \text{ for some } k \in \{1, \dots, r\}\}$$

and

$$L_2 = (\{1, \dots, k_1\} \times \dots \times \{1, \dots, k_n\}) \setminus L_1.$$

If $(l_1, \dots, l_n) \in L_2$ then $\omega_f(\boldsymbol{x}) < \epsilon'$ for each $\boldsymbol{x} \in R_{l_1,\dots,l_n}$. Thus, for $\boldsymbol{x} \in R_{l_1,\dots,l_n}$ there exists a $r_{\boldsymbol{x}} \in \mathbb{R}_{>0}$ such that

$$\sup\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathsf{B}^n(r_{\boldsymbol{x}}, \boldsymbol{x})\} - \inf\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \mathsf{B}^n(r_{\boldsymbol{x}}, \boldsymbol{x})\} < \epsilon'.$$

The balls $(\mathsf{B}^n(r_{\boldsymbol{x}}, \boldsymbol{x}))_{\boldsymbol{x}\in R_{l_1,\dots,l_n}}$ cover $R_{l_1,\dots,l_n}$. Since $R_{l_1,\dots,l_n}$ is compact there exists a finite set of points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in R_{l_1,\dots,l_n}$ such that $(\mathsf{B}^n(r_{\boldsymbol{x}_j}, \boldsymbol{x}_j))_{j=1}^m$ covers $R_{l_1,\dots,l_n}$. Now one can make a refinement of the partition R which gives rise to a partition of the rectangle $R_{l_1,\dots,l_n}$ such that each subrectangle of this partition lies within one of the balls $(\mathsf{B}^n(r_{\boldsymbol{x}_j}, \boldsymbol{x}_j))_{j=1}^m$. Doing this for each $(l_1, \dots, l_n) \in L_2$ gives a partition $\boldsymbol{P}' = (P'_1, \dots, P'_n)$ of R with $P'_j = (I'_{j1}, \dots, I'_{jk'_j})$. For this partition define

$$L'_1 = \{(l_1, \dots, l_n) \in \{1, \dots, k'_1\} \times \dots \times \{1, \dots, k'_n\} \mid R_{l_1,\dots,l_n} \subseteq R_{j_k} \text{ for some } k \in \{1, \dots, r\}\}$$

and

$$L'_2 = (\{1, \dots, k'_1\} \times \dots \times \{1, \dots, k'_n\}) \setminus L'_1.$$

For this partition we have

$$\begin{aligned}
&A_+(f,\boldsymbol{P}') - A_-(f,\boldsymbol{P}') = \\
&\quad \sum_{(l_1,\dots,l_n)\in L_1'} (\sup\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in R_{l_1,\dots,l_n}\} - \inf\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in R_{l_1,\dots,l_n}\})\mathrm{vol}(R_{l_1,\dots,l_n}) \\
&\quad + \sum_{(l_1,\dots,l_n)\in L_2'} (\sup\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in R_{l_1,\dots,l_n}\} - \inf\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in R_{l_1,\dots,l_n}\})\mathrm{vol}(R_{l_1,\dots,l_n}) \\
&\leq \sum_{(l_1,\dots,l_n)\in L_1'} 2\sup\{|f(\boldsymbol{x})| \mid \boldsymbol{x} \in R\}\mathrm{vol}(R_{l_1,\dots,l_n}) + \epsilon'\mathrm{vol}(R) \\
&\leq \frac{\epsilon}{4M}2M + \frac{\epsilon}{\mathrm{vol}(R)}\mathrm{vol}(R) = \epsilon,
\end{aligned}$$

where we have used the fact that

$$\sum_{k=1}^{r} \mathrm{vol}(R_{j_k}) < \frac{\epsilon}{4M}$$

and, for $(l_1,\dots,l_n) \in L_1'$, $R_{l_1,\dots,l_n} \subseteq R_{j_k}$ for some $k \in \{1,\dots,r\}$. Thus $A_+(f,\boldsymbol{P}') - A_-(f,\boldsymbol{P}')$ can be made arbitrarily small by a suitable choice of partition $\boldsymbol{P}'$, and so f is Riemann integrable.

Now suppose that f is Riemann integrable. For $k \in \mathbb{Z}_{>0}$ define

$$D_{f,k} = \left\{\boldsymbol{x} \in R \;\middle|\; \omega_f(\boldsymbol{x}) \geq \tfrac{1}{k}\right\}.$$

Then the analogue of Proposition I-3.1.11 in $\mathbb{R}^n$ implies that $D_f = \cup_{k\in\mathbb{Z}_{>0}} D_{f,k}$. By Exercise 1.2.12 we can assert that D_f has measure zero if and only if each of the sets $D_{f,k}$ has measure zero, $k \in \mathbb{Z}_{>0}$. Let $\boldsymbol{P} = (P_1,\dots,P_n)$ be a partition of R and denote $P_j = (I_{j1},\dots,I_{jk_j})$ for $j \in \{1,\dots,n\}$. Since f is Riemann integrable we can choose $\boldsymbol{P}$ such that

$$A_+(f,\boldsymbol{P}) - A_-(f,\boldsymbol{P}) < \frac{\epsilon}{k}.$$

Write $D_{f,k} = A_{f,k} \cup B_{f,k}$ where

$$A_{f,k} = \{\boldsymbol{x} \in D_{f,k} \mid \ \boldsymbol{x} \in \mathrm{bd}(R_{l_1,\dots,l_n}) \quad \text{for some}(l_1,\dots,l_n) \in \{1,\dots,k_1\} \times \dots \times \{1,\dots,k_n\}\}$$

and where $B_{f,k} = D_{f,k} \setminus A_{f,k}$. Since the boundaries of the subrectangles of $\boldsymbol{P}$ each have measure zero (by Exercise 1.2.13), and since there are finitely many such boundaries, it follows that $A_{f,k}$ has measure zero. Denote

$$L_1 = \{(l_1,\dots,l_n) \in \{1,\dots,k_1\} \times \dots \times \{1,\dots,k_n\} \mid \ \mathrm{int}(R_{l_1,\dots,l_n}) \cap D_{f,k} \neq \emptyset\}.$$

For $(l_1,\dots,l_n) \in L_1$ is follows that

$$\sup\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in R_{l_1,\dots,l_n}\} - \inf\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in R_{l_1,\dots,l_n}\} < \frac{1}{k}.$$

Therefore,

$$\begin{aligned}\frac{1}{k}\sum_{(l_1,\dots,l_n)\in L_1}\operatorname{vol}(R_{l_1,\dots,l_n}) \le \sum_{(l_1,\dots,l_n)\in L_1} &(\sup\{f(\boldsymbol{x}) \mid \boldsymbol{x}\in R_{l_1,\dots,l_n}\} \\ &- \inf\{f(\boldsymbol{x}) \mid \boldsymbol{x}\in R_{l_1,\dots,l_n}\})\operatorname{vol}(R_{l_1,\dots,l_n}) \\ &\le A_+(f,\boldsymbol{P}) - A_-(f,\boldsymbol{P}) < \frac{\epsilon}{k}.\end{aligned}$$

Thus the rectangles $R_{l_1,\dots,l_n}$, $(l_1,\dots,l_n)\in L_1$, cover $B_{f,k}$ and have total volume bounded above by ϵ. Since $\epsilon\in\mathbb{R}_{>0}$ is arbitrary it follows that $B_{f,k}$ has measure zero. Since $A_{f,k}$ has measure zero it then follows from Exercise 1.2.12 that $D_{f,k}$ has measure zero. ■

1.6.4 The Riemann integral and subsets of $\mathbb{R}^n$

For the Riemann integral on the real line there is not much of interest that one can derive concerning the domains of definition of Riemann integrable functions. However, for functions defined on $\mathbb{R}^n$ the domain of definition of the function becomes of great interest, and it is not at all obvious, at least immediately, what are the viable domains of interest for Riemann integrable functions. We comment that we will examine the topic of this section in more detail in Section III-2.1.1.

We begin by defining the notion of a Jordan measurable set.

1.6.12 Definition (Jordan measurable[5] set) A subset $A\subseteq\mathbb{R}^n$ is ***Jordan measurable*** if $\operatorname{bd}(A)$ is a set of measure zero. •

Sometimes the notion of being Jordan measurable is referred to as ***having volume***, ***having content***, or ***having Jordan content***.

The definition of Jordan measurable sets as given seems to come from nowhere. However, for bounded Jordan measurable sets, the following characterisation brings us back to something reasonable. We recall from Example I-1.3.3–5 the notion of the characteristic function χ_A of a set A as the function that takes value 1 on the set and 0 off the set.

1.6.13 Theorem (Characterisation 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) χ_A *is Riemann integrable.*

Proof By Theorem 1.6.11 it follows that A is Jordan measurable if and only if the set of discontinuities of χ_A has measure zero. Thus we need only show that the set of discontinuities of χ_A is equal to $\operatorname{bd}(A)$.

First let $\boldsymbol{x}_0$ be a point of discontinuity of χ_A. Thus there exists $\epsilon\in\mathbb{R}_{>0}$ such that, for every $\delta\in\mathbb{R}_{>0}$, $|f(\boldsymbol{x})-f(\boldsymbol{x}_0)|\ge\epsilon$ for $\boldsymbol{x}\in\mathsf{B}^n(\delta,\boldsymbol{x}_0)$. This means that in any neighbourhood

[5] The term "measurable" seems to come from nowhere at this juncture. The expression will not have context for us until Chapter III-2.

of $\boldsymbol{x}_0$ the function χ_A takes values 0 and 1. By Proposition III-3.6.8 it follows that $\boldsymbol{x}_0 \in \mathrm{bd}(A)$.

Now let $\boldsymbol{x}_0 \in \mathrm{bd}(A)$. Then, by Proposition III-3.6.8, for every $\delta \in \mathbb{R}_{>0}$ the ball $\mathsf{B}^n(\delta, \boldsymbol{x}_0)$ contains points for which χ_A takes values 0 and 1. However, this precludes χ_A from being continuous at $\boldsymbol{x}_0$. ■

The following simple example shows that not all sets are Jordan measurable. Further examples will be considered in Section III-2.1.1.

1.6.14 Example (A set that is not Jordan measurable) Let $A = \mathbb{Q} \cap [0,1]$. Then $\mathrm{bd}(A) = [0,1]$ and so, by the previous result, A is not Jordan measurable. •

Now having a useful characterisation of bounded Jordan measurable sets, let us assign some notation to these.

1.6.15 Definition (Volume of a bounded Jordan measurable set) If A is a bounded Jordan measurable set then the ***volume*** of A is $\mathrm{vol}(A) = \int_A \chi_A(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$. •

We can now give an further characterisation of bounded Jordan measurable sets which shows that they are quite special.

1.6.16 Theorem (Another characterisation 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) $\mathrm{vol}(\mathrm{bd}(A)) = 0$.

Proof Suppose that R is a fat compact rectangle containing A.

(i) $\implies$ (ii) Let $\epsilon \in \mathbb{R}_{>0}$. Since A is Jordan measurable let $\boldsymbol{P}$ be partition of R such that

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

The subrectangles of $\boldsymbol{P}$ come in three sorts: (1) the first sort are subrectangles contained in A; (2) the second sort are rectangles that intersect A but are not contained in A; (3) the third sort are rectangles that do not intersect A. Clearly, by Proposition 1.2.26, $\mathrm{bd}(A)$ is contained in the union of the subrectangles of the second sort. Moreover, since $s_-(\chi_A, \boldsymbol{P})$ is equal to $s_+(\chi_A, \boldsymbol{P})$ at all points in subrectangles of the first and third sort, and since $s_-(\chi_A, \boldsymbol{P})$ has value 0 and $s_+(\chi_A, \boldsymbol{P})$ has value 1 at all points in subrectangles of the second sort, it follows that the total volume of the subrectangles of the second sort is exactly

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

which means that $\mathrm{bd}(A)$ is covered by rectangles whose total volume is bounded above by ϵ. More or less by definition (but see Proposition 1.6.18 below) this implies that $\mathrm{bd}(A)$ has volume zero.

(ii) $\implies$ (i) Let $\epsilon \in \mathbb{R}_{>0}$. If $\mathrm{bd}(A)$ has volume zero, by Proposition 1.6.18 below it follows that there are finitely many rectangles covering $\mathrm{bd}(A)$ whose total volume is bounded above by ϵ. While these rectangles may overlap, it is easy to see that these rectangles can be written as a union of finitely many rectangles that *do not* overlap (why is this?). Now let $\boldsymbol{P}$ be a partition for which these nonoverlapping rectangles are subrectangles. The argument in the preceding part of the proof can now be reversed to show that

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

An interesting corollary is the following.

1.6.17 Corollary (Characterisation of the boundary of a bounded set) *The boundary of a bounded set* A *has zero measure if and only if it has zero volume.*

Thus boundaries of bounded sets have the property that zero measure and zero volume agree. This is not the usual situation; it is saying that boundaries of bounded sets are rather special. Indeed, zero volume and zero measure are generally not the same. In fact, we have the following characterisation of zero volume which makes clear the distinction.

1.6.18 Proposition (Characterisation of sets of zero volume) *For a bounded subset* $A \subseteq \mathbb{R}^n$ *the following statements are equivalent:*

(i) A *has zero volume;*

(ii) for every $\epsilon \in \mathbb{R}_{>0}$ *there exists a finite collection* $(R_j)_{j=1}^k$ *of rectangles such that* $A \subseteq \cup_{j=1}^k R_j$ *and* $\sum_{j=1}^k \mathrm{vol}(R_j) < \epsilon$*;*

(iii) for every $\epsilon \in \mathbb{R}_{>0}$ *there exists a finite collection* $(\mathsf{B}^n(r_j, \mathbf{x}_j))_{j=1}^k$ *of open balls such that* $A \subseteq \cup_{j=1}^k \mathsf{B}^n(r_j, \mathbf{x}_j)$ *and* $\sum_{j=1}^k \mathrm{vol}(\mathsf{B}^n(r_k, \mathbf{x}_j)) < \epsilon$*.*

Proof (i) $\Longrightarrow$ (ii) Let $\epsilon \in \mathbb{R}_{>0}$. By definition of zero volume there exists a partition $\boldsymbol{P}$ such that

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

The subrectangles of A that intersect A form a cover of A by rectangles whose volumes sum to at most ϵ.

(i) $\Longrightarrow$ (ii) Let $\epsilon \in \mathbb{R}_{>0}$. Cover A by rectangles $R_1, \dots, R_k$ having total volume at most ϵ. Since finite intersections of rectangles are rectangles (why?) it follows that there exists nonoverlapping rectangles $\tilde{R}_1, \dots, \tilde{R}_{\tilde{k}}$ such that

$$\cup_{j=1}^k R_j = \cup_{\tilde{j}=1}^{\tilde{k}} \tilde{R}_{\tilde{k}}.$$

If $\boldsymbol{P}$ is a partition for which the rectangles $\tilde{R}_1, \dots, \tilde{R}_{\tilde{k}}$ are subrectangles (such a partition exists since the rectangles $\tilde{R}_1, \dots, \tilde{R}_{\tilde{k}}$ do not overlap), then

$$A_+(\chi_A, \boldsymbol{P}) \le \sum_{j=1}^{k} \mathrm{vol}(R_k) < \epsilon,$$

showing that A has zero volume.

(i) $\Longrightarrow$ (iii) First suppose that A has zero volume. Let R be a fat compact rectangle containing A and let $\boldsymbol{P}$ be a partition of R such that

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

where $\alpha \in \mathbb{R}_{>0}$ is the constant for which the volume of the ball of radius $2\sqrt{n}r$ has volume α times the volume of the rectangle all of whose sides have length r (cf. (1.10)). By Theorem 1.6.10 we can suppose, without loss of generality, that the partition $\boldsymbol{P}$ has the property that all subrectangles have sides of equal length δ. For each subrectangle of $\boldsymbol{P}$ which contains a point from A place an open ball of radius $2\sqrt{n}\delta$ at the centre of the subrectangle. Since the distance from the centre of a rectangle whose sides have length δ to the furthest point in the rectangle is $\sqrt{n}\delta$ (why?), it follows that the resulting balls cover A and have volume

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

as desired.

(iii) $\Longrightarrow$ (i) Now suppose that, for each $\epsilon \in \mathbb{R}_{>0}$, A can be covered by a finite collection of open balls whose total volume does not exceed ϵ. Then let $(\mathsf{B}^n(r_j, \boldsymbol{x}_j))_{j=1}^k$ be a collection of balls covering A and having total volume less that $\frac{\epsilon}{\beta}$ where $\beta \in \mathbb{R}_{>0}$ is the constant for which the volume of the rectangle whose sides all have length 2δ is β times the volume of the ball of radius δ (again, cf. (1.10)). For each of the balls $\mathsf{B}^n(r_j, \boldsymbol{x}_j)$, $j \in \{1, \dots, k\}$, place a rectangle R_j with centre at $\boldsymbol{x}_j$ whose sides have length $2r_j$. Then the rectangles $R_1, \dots, R_k$ cover A. The total volume of the rectangles $R_1, \dots, R_k$ is then less than $\beta\frac{\epsilon}{\beta} = \epsilon$. Now let R be a fat compact rectangle containing A and let $\boldsymbol{P}$ be a partition having the property that all $(n-1)$-dimensional faces of the rectangles $R_1, \dots, R_k$ that lie in R are contained in the $(n-1)$-dimensional faces of the subrectangles of $\boldsymbol{P}$. The subrectangles from $\boldsymbol{P}$ which intersect A will be contained in the rectangles $R_1, \dots, R_k$. Therefore, by definition of χ_A,

$$A_+(\chi_A, \boldsymbol{P}) \leq \sum_{j=1}^k \mathrm{vol}(R_j) < \epsilon,$$

giving χ_A as Riemann integrable with Riemann integral zero since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary. ■

This result then gives the following relationship between volume zero and measure zero.

1.6.19 Corollary (Volume zero implies measure zero) *If* $A \subseteq \mathbb{R}^n$ *has volume zero then it has measure zero.*

Proof Let $(\mathsf{B}^n(r_j, \boldsymbol{x}_j))_{j=1}^k$ be a finite collection of balls covering A and having total volume less that ϵ. Then this gives a countable collection of balls covering A and having total volume less than ϵ, and this means A has measure zero. ■

The converse of the corollary is not generally true.

1.6.20 Example (A set that is not Jordan measurable and has zero measure) Let $A = \mathbb{Q} \cap [0, 1]$. Then A is not Jordan measurable by Example I-3.4.10. However, A has measure zero by Exercises I-2.1.3 and I-2.5.10. •

Now let us investigate the relationship between functions with Riemann integral zero and sets with volume zero.

1.6.21 Proposition (Functions with zero integral and sets of zero measure) *Let* $A \subseteq \mathbb{R}^n$ *be bounded and let* $f\colon A \to \mathbb{R}$ *be bounded. Then the following statements hold:*

(i) *if* A *has measure zero and if* f *is Riemann integrable then the Riemann integral of* f *is zero;*

(ii) *if* $\operatorname{image}(f) \subseteq \mathbb{R}_{\geq 0}$ *and if the Riemann integral of* f *is zero, then* $\{\mathbf{x} \in A \mid f(\mathbf{x}) = 0\}$ *has measure zero.*

Proof (i) Let R be a fat compact rectangle containing A and think of f as being defined on R by asking it to be zero off A, let $\boldsymbol{P} = (P_1, \dots, P_n)$ be a partition of R with $P_j = (I_{j1}, \dots, I_{jk_j})$, $j \in \{1, \dots, n\}$, and denote $M = \sup\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in A\}$. We claim that $A_-(f, \boldsymbol{P}) \leq 0$. Indeed, we have

$$\begin{aligned} A_-(f, \boldsymbol{P}) &= \sum_{l_1=1}^{k_1} \cdots \sum_{l_n=1}^{k_n} \inf\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \operatorname{cl}(R_{l_1,\dots,l_n})\} \operatorname{vol}(R_{l_1,\dots,l_n}) \\ &\leq M \sum_{l_1=1}^{k_1} \cdots \sum_{l_n=1}^{k_n} \inf\{\chi_A(\boldsymbol{x}) \mid \boldsymbol{x} \in \operatorname{cl}(R_{l_1,\dots,l_n})\} \operatorname{vol}(R_{l_1,\dots,l_n}). \end{aligned}$$

If

$$\inf\{\chi_A(\boldsymbol{x}) \mid \boldsymbol{x} \in \operatorname{cl}(R_{l_1,\dots,l_n})\} \operatorname{vol}(R_{l_1,\dots,l_n}) \neq 0$$

for some $(l_1, \dots, l_n)$ then this implies that $R_{l_1,\dots,l_n} \subseteq A$ and $\operatorname{vol}(R_{l_1,\dots,l_n} > 0$. However, this cannot happen since A has measure zero. Therefore, each term in the sum defining $A_-(f, \boldsymbol{P})$ is nonpositive. Thus $A_-(f, \boldsymbol{P}) \leq 0$. A similarly styled argument gives $A_+(f, \boldsymbol{P}) \geq 0$. Therefore,

$$I_-(f) \leq 0 \leq I_+(f)$$

which gives $I_-(f) = I_+(f) = 0$ since f is Riemann integrable.

(ii) For $k \in \mathbb{Z}_{>0}$ define

$$A_k = \left\{\boldsymbol{x} \in A \;\middle|\; f(\boldsymbol{x}) > \tfrac{1}{k}\right\}.$$

We claim that A_k has zero volume. Let R be a fat compact rectangle containing A and extend f to R b making it zero off A. For $\epsilon \in \mathbb{R}_{>0}$ let $\boldsymbol{P} = (P_1, \dots, P_n)$ be a partition of R such that $A_+(f, \boldsymbol{P}) < \frac{\epsilon}{k}$. Denote $P_j = (I_{j1}, \dots, I_{k_j})$, $j \in \{1, \dots, n\}$. Let

$$L_1 = \{(l_1, \dots, l_n) \in \{1, \dots, k_1\} \times \cdots \times \{1, \dots, k_n\} \mid R_{l_1,\dots,l_n} \cap A_k \neq \emptyset\}.$$

Then

$$\sum_{(l_1,\dots,l_n) \in L_1} \operatorname{vol}(R_{l_1,\dots,l_n}) \leq \sum_{(l_1,\dots,l_n) \in L_1} k \sup\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in \operatorname{cl}(R_{l_1,\dots,l_n})\} \operatorname{vol}(R_{l_1,\dots,l_n}) < \epsilon,$$

using the definition of A_k and L_1. Since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary, it follows that A_k has zero volume by Proposition 1.6.18. Thus, by Corollary 1.6.19, A_k has zero measure. Since $A = \cup_{k=1}^{\infty} A_k$ it follows from Exercise 1.2.12 that A has measure zero. ■

1.6.5 The Riemann integral for unbounded functions on unbounded domains

For the Riemann integral on the real line, the extension of the integral from compact intervals to general intervals is achieved in a more or less direct way (see Definition I-3.4.14). For the integral over noncompact subsets of $\mathbb{R}^n$ it is not so obvious how this should be done. The procedure we give generalises the construction of the indefinite integral of Section I-3.4.4. In that section, for functions defined on $\mathbb{R}$, we considered two notions of integrability, Riemann integrability and conditional Riemann integrability. For the Riemann integral on $\mathbb{R}^n$ we do not bother with this level of generality since it actually poses certain problems with respect to how the definition should work.

The definition is made in stages. For $r \in \mathbb{R}_{>0}$ denote by

$$R_r = [-r, r] \times \cdots \times [-r, r]$$

the rectangle all of sides have length $2r$.

1.6.22 Definition (The Riemann integral for unbounded functions and unbounded domains) Let $A \subseteq \mathbb{R}^n$, let $f\colon A \to \mathbb{R}$, and also denote by $f\colon \mathbb{R}^n \to \mathbb{R}$ the extension of f made by asking that $f(\boldsymbol{x}) = 0$ for $\boldsymbol{x} \notin A$. For $M \in \mathbb{R}_{>0}$ denote

$$f_M(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}), & |f(\boldsymbol{x})| \leq M, \\ 0, & |f(\boldsymbol{x})| > M. \end{cases}$$

Also denote by

$$f_+(\boldsymbol{x}) = \max\{0, f(\boldsymbol{x})\}, \quad f_-(\boldsymbol{x}) = -\min\{0, f(\boldsymbol{x})\}$$

the positive and negative parts of f.

(i) If $\mathrm{image}(f) \subseteq \mathbb{R}_{\geq 0}$ and if $f|R_r$ is Riemann integrable (in the sense of Definition 1.6.7) for every $r \in \mathbb{R}_{>0}$, then f is ***Riemann integrable*** if the limit

$$\lim_{r\to\infty} \int_{R_r} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

exists. This limit is denoted by

$$\int_A f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

when it exists, and is called the ***Riemann integral*** of f.

(ii) If $\mathrm{image}(f) \subseteq \mathbb{R}_{\geq 0}$ and if f_M is Riemann integrable (as in part (i)) for each $M \in \mathbb{R}_{>0}$, then f is ***Riemann integrable*** if the limit

$$\lim_{M\to\infty} \int_A f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

exists. This limit is denoted by

$$\int_A f(x)\,dx$$

when it exists, and is called the ***Riemann integral*** of f.

(iii) If f_+ and f_- are Riemann integrable (as in part (ii)) then f is ***Riemann integrable*** and we denote by

$$\int_A f(x)\,dx = \int_A f_+(x)\,dx - \int_A f_-(x)\,dx$$

the ***Riemann integral*** of f. •

As was the case after Definition I-3.4.15, we now have at hand two possibly competing definitions for the Riemann integral of a function on a bounded set A. First we can extend A to a fat compact rectangle R and then apply Definition 1.6.7. Second we can break up f into its positive and negative parts and then apply Definition 1.6.22. Let us now make sure these two definitions are equivalent.

1.6.23 Proposition (Consistency of definition of Riemann integral on bounded subsets) *Let* $A \subseteq \mathbb{R}^n$ *be a bounded set, let* $R \subseteq \mathbb{R}^n$ *be a fat compact rectangle for which* $A \subseteq R$*, let* $f\colon A \to \mathbb{R}$ *be bounded, and let* $f_+, f_-\colon A \to \mathbb{R}_{\geq 0}$ *be the positive and negative parts of* f*. Denote by* f*,* f_+*, and* f_- *the extensions of the functions from* A *to* R *by asking that they have value zero off* A*. Then the following two statements are equivalent:*

(i) f *is integrable as per Definition 1.6.7 with Riemann integral* $I(f)$*;*

(ii) f_+ *and* f_- *are Riemann integrable as per Definition 1.6.7 with Riemann integrals* $I(f_+)$ *and* $I(f_-)$*.*

Moreover, if one, and therefore both, of parts (i) and (ii) hold, then $I(f) = I(f_+) - I(f_-)$*.*

Proof We refer ahead to . what?

(i) $\Longrightarrow$ (ii) Define continuous functions $g_+, g_-\colon \mathbb{R} \to \mathbb{R}$ by Make sure this proof holds up

$$g_+(x) = \max\{0, x\}, \quad g_-(x) = -\min\{0, x\}$$

so that $f_+ = g_+ \circ f$ and $f_- = g_- \circ f$. By Proposition 1.6.29 (noting that the proof of that result is valid for the Riemann integral as per Definition 1.6.7) it follows that f_+ and f_- are Riemann integrable as per Definition 1.6.7.

(ii) $\Longrightarrow$ (i) Note that $f = f_+ - f_-$. Also note that the proof of Proposition 1.6.28 is valid for the Riemann integral as per Definition 1.6.7. Therefore, f is Riemann integrable as per Definition 1.6.7.

Now we show that $I(f) = I(f_+) - I(f_-)$. This, however, follows immediately from Proposition 1.6.28. ■

Let us give some equivalent characterisations of Riemann integrable functions. First some notation. A sequence $(A_j)_{j\in\mathbb{Z}_{>0}}$ of bounded Jordan measurable subsets of $\mathbb{R}^n$ is called ***space filling*** if $A_j \subseteq A_{j+1}$ for each $j \in \mathbb{Z}_{>0}$ and if, for any rectangle $R \subseteq \mathbb{R}^n$, we have $R \subseteq A_j$ for some sufficiently large $j \in \mathbb{Z}_{>0}$.

1.6.24 Proposition (Characterisation of locally bounded Riemann integrable functions) *Let $A \subseteq \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}$ have the property that $f|R_r$ is Riemann integrable (in the sense of Definition 1.6.7). Then the following statements are equivalent:*

(i) f is Riemann integrable;

(ii) for any space filling sequence $(A_j)_{j\in\mathbb{Z}_{>0}}$ of subsets of $\mathbb{R}^n$ the limit

$$\lim_{j\to\infty}\int_{A_j} f(\mathbf{x})\,d\mathbf{x}$$

exists.

Moreover, if one, and therefore both, of parts (i) and (ii) hold then the limit in part (ii) is equal to the Riemann integral of f.

Proof (i) $\Longrightarrow$ (ii) Let f be Riemann integrable and let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a space filling sequence of subsets of $\mathbb{R}^n$. First let us suppose that image$(f) \subseteq \mathbb{R}^n$. For $j \in \mathbb{Z}_{>0}$ let $c_j, C_j \in \mathbb{R}_{>0}$ be such that $R_{c_j} \subseteq A_j \subseteq R_{C_j}$. Note that for small j this may not be possible, but for sufficiently large j this is always possibly by the definition of space filling. Since we are interested in the limit as $j \to \infty$, we suppose, without loss of generality, that this can be done for all $j \in \mathbb{Z}_{>0}$. This being the case, we have

$$\int_{R_{c_j}} f(\mathbf{x})\,d\mathbf{x} \le \int_{A_j} f(\mathbf{x})\,d\mathbf{x} \le \int_{R_{C_j}} f(\mathbf{x})\,d\mathbf{x}. \tag{1.32}$$

Then, since

$$\lim_{j\to\infty}\int_{R_{c_j}} f(\mathbf{x})\,d\mathbf{x} = \lim_{j\to\infty}\int_{R_{C_j}} f(\mathbf{x})\,d\mathbf{x}$$

it follows that

$$\lim_{j\to\infty}\int_{A_j} f(\mathbf{x})\,d\mathbf{x}$$

exists and is equal to the Riemann integral of f. The case when f is not necessarily nonnegative follows from Proposition 1.6.23.

(ii) $\Longrightarrow$ (i) First suppose that image$(f) \subseteq \mathbb{R}_{\ge 0}$ and let $(A_j)_{j\in\mathbb{Z}_{>0}}$ be a space filling sequence. Then the sequence of numbers whose jth term is $\int_{A_j} f(\mathbf{x})\,d\mathbf{x}$ is nondecreasing (since f is nonnegative and since $A_j \subseteq A_{j+1}$ for $j \in \mathbb{Z}_{>0}$) and is also convergent. Let us denote the limit by $L(f)$ so that $\int_{A_j} f(\mathbf{x})\,d\mathbf{x} \le L(f)$ for all $j \in \mathbb{Z}_{>0}$. Since the sequence $(A_j)_{j\in\mathbb{Z}_{>0}}$ is space filling, for each $r \in \mathbb{R}_{>0}$ there exists $j_r \in \mathbb{Z}_{>0}$ such that $R_r \subseteq A_{j_r}$. Therefore $\int_{R_r} f(\mathbf{x})\,d\mathbf{x} \le L(f)$ for each r. Thus the function $r \mapsto \int_{R_r} f(\mathbf{x})\,d\mathbf{x}$ is monotonically increasing and bounded, and so

$$\lim_{r\to\infty}\int_{R_r} f(\mathbf{x})\,d\mathbf{x}$$

exists. That $L(f)$ is equal to the Riemann integral of f now follows since the inequalities (1.32) hold, and by taking the limit as $j \to \infty$ as we did in the preceding part of the proof. ■

Next we turn to an alternative characterisation of the Riemann integral for possibly unbounded functions on bounded domains. For a bounded subset $A \subseteq \mathbb{R}^n$ a sequence of compact Jordan measurable subsets $(K_j)_{j\in\mathbb{Z}_{>0}}$ is ***A-filling*** if $K_j \subseteq K_{j+1}$ for each $j \in \mathbb{Z}_{>0}$ and if $A = \cup_{j=1}^{\infty} K_j$.

1.6.25 Proposition (Characterisation of Riemann integrable function on a bounded domain) *Let $A \subseteq \mathbb{R}^n$ be a bounded set for which there exists an A-filling sequence $(K_j)_{j\in\mathbb{Z}_{>0}}$ and let $f\colon A \to \mathbb{R}_{\geq 0}$ have the property that f_M is Riemann integrable for each $M \in \mathbb{R}_{>0}$. Then the following statements are equivalent:*

(i) f is Riemann integrable;

(ii) for any A-filling sequence $(K_j)_{j\in\mathbb{Z}_{>0}}$ of subsets of $\mathbb{R}^n$ for which $f|K_j$ is bounded and Riemann integrable for each $j \in \mathbb{Z}_{>0}$, the limit

$$\lim_{j\to\infty}\int_{K_j} f(x)\,dx$$

exists.

Moreover, if either statement is true, then the limit in part (ii) is equal to the Riemann integral of f.

Proof First we prove a lemma.

1 Lemma *Let $A \subseteq \mathbb{R}^n$ be a bounded Jordan measurable set and let $(K_j)_{j\in\mathbb{Z}_{>0}}$ be an A-filling sequence of subsets of $\mathbb{R}^n$. Then $\lim_{j\to\infty}\mathrm{vol}(K_j) = \mathrm{vol}(A)$.*

Proof This follows from the Monotone Convergence Theorem, stated as Theorem 1.7.6 below. ▼

(i) $\implies$ (ii) Suppose that f is Riemann integrable and let $(K_j)_{j\in\mathbb{Z}_{>0}}$ be an A-filling sequence such that $f|K_j$ is Riemann integrable for each $j \in \mathbb{Z}_{>0}$. Let $M_j = \sup\{f(\boldsymbol{x}) \mid \boldsymbol{x} \in K_j\}$. We then have

$$\int_{K_j} f(\boldsymbol{x})\,d\boldsymbol{x} \leq \int_A f_{M_j}(\boldsymbol{x})\,d\boldsymbol{x} \leq I(f) \triangleq \int_A f(\boldsymbol{x})\,d\boldsymbol{x}.$$

Note that the sequence whose jth term is $\int_{K_j} f(\boldsymbol{x})\,d\boldsymbol{x}$ is increasing since $K_{j+1} \subseteq K_j$ for each $j \in \mathbb{Z}_{>0}$ and since f is assumed to take nonnegative values. The inequalities above show that the sequence is also bounded above by $I(f)$ and so, by Theorem I-2.3.8, the sequence converges to a limit bounded above by $I(f)$.

(ii) $\implies$ (i) Let $(K_j)_{j\in\mathbb{Z}_{>0}}$ be an A-filling sequence and denote

$$L(f) = \lim_{j\to\infty}\int_{K_j} f(\boldsymbol{x})\,d\boldsymbol{x}.$$

For $M \in \mathbb{R}_{>0}$ define $g_M\colon \mathbb{R}_{\geq 0} \to \mathbb{R}$ by $g_M(x) = \max\{x, M\}$ and note that $g_M$is continuous. From this and from Proposition 1.6.29 it follows that $f_M|K_j$ is Riemann integrable. We then have, for each $j \in \mathbb{Z}_{>0}$ and $M \in \mathbb{R}_{>0}$,

$$\int_{K_j} f_M(\boldsymbol{x})\,d\boldsymbol{x} \leq \int_{K_j} f(\boldsymbol{x})\,d\boldsymbol{x} \leq L(f).$$

Therefore, since $K_{j+1} \subseteq K_j$ for each $j \in \mathbb{Z}_{>0}$ and since f is nonnegative, the sequence whose jth element is $\int_{K_j} f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ is monotonically increasing. The above computation shows that it is bounded. Therefore, the limit

$$\lim_{j\to\infty} \int_{K_j} f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

exists and is bounded above by $L(f)$. Moreover, the same computations that we shall use below to prove the final assertion of the proposition show that this limit is the Riemann integral of f_M. That is,

$$\int_A f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \le L(f)$$

for each $M \in \mathbb{R}_{>0}$. Therefore, since $\int_A f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ is a monotonically increasing function of M, it follows that the limit

$$\lim_{M\to\infty} \int_A f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

exists. Thus f is Riemann integrable.

For the final assertion of the proof we need to show that

$$\int_A f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \lim_{j\to\infty} \int_{K_j} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

Let $\epsilon \in \mathbb{R}_{>0}$. Suppose that for $\epsilon \in \mathbb{R}_{>0}$ we choose $M \in \mathbb{R}_{>0}$ sufficiently large that

$$\int_A f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_A f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} < \frac{\epsilon}{3}$$

Now let N be sufficiently large that $\mathrm{vol}(A) - \mathrm{vol}(K_j) < \frac{\epsilon}{3M}$, this being possible by the lemma above. Then, for $j \ge N$,

$$\begin{aligned}
\int_A f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_{K_j} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \left|\int_A f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_{K_j} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \\
&\le \left|\int_A f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_A f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \\
&\quad + \left|\int_A f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_{K_j} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \\
&\le \frac{\epsilon}{3} + \left|\int_{A\setminus K_j} f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int_{K_j} f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int_{K_j} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \\
&\le \frac{\epsilon}{3} + \int_{A\setminus K_j} f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int_{K_j} (f(\boldsymbol{x}) - f_M(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x} \\
&\le \frac{\epsilon}{3} + M\frac{\epsilon}{3M} + \frac{\epsilon}{3} = \epsilon.
\end{aligned}$$

This gives us $\lim_{j\to\infty} \int_{K_j} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_A f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ as desired. ■

Let us give an example that illustrates that Definition 1.6.22 actually generalises the definition of the Riemann integral in Section I-3.4.

1.6.26 Example (An unbounded Riemann integrable function) On $I = [-1, 1]$ define $f\colon [-1, 1] \to \mathbb{R}$ by

$$f(x) = \begin{cases} \frac{1}{\sqrt{x}}, & x \in (0, 1], \\ \frac{1}{\sqrt{-x}}, & x \in [-1, 0), \\ 0, & x = 0. \end{cases}$$

Note that this function is *not* Riemann integrable by the definition of the Riemann integral in Section I-3.4. Indeed, the definition of the integral in that section requires that the function be bounded on any compact subset of I, which is not the case for f. However, f *is* Riemann integrable according to Definition 1.6.22. Let us verify this. According to the construction, for $M \in \mathbb{R}_{>0}$ we have

$$f_M(x) = \begin{cases} \frac{1}{\sqrt{x}}, & x \in [\frac{1}{M^2}, 1],\ M > 1, \\ \frac{1}{\sqrt{-x}}, & x \in [-1, -\frac{1}{M^2}],\ M > 1, \\ M, & x \in (0, \frac{1}{M^2}) \cup (-\frac{1}{M^2}, 0),\ M > 1, \\ 0, & x = 0, \\ M, & x \in (0, 1] \cup [-1, 0),\ M \leq 1. \end{cases}$$

A direct computation then gives

$$\int_{-1}^{1} f_M(x)\,\mathrm{d}x = \begin{cases} 4 - \frac{2}{M}, & M > 1, \\ 2M, & M \leq 1. \end{cases}$$

Thus the definition of the Riemann integral in Definition 1.6.22 gives

$$\int_{-1}^{1} f(x)\,\mathrm{d}x = \lim_{M\to\infty} \int_{-1}^{1} f_M(x)\,\mathrm{d}x = 4.$$

This shows the increased flexibility of Definition 1.6.22 over the construction in Section I-3.4. •

1.6.27 Remark (Now what do we mean when we say "Riemann integrable") Even without introducing the complication of conditional integrability as introduced in Section I-3.4.4, our extension of the Riemann integral in this section to possibly unbounded functions defined on possibly unbounded domains can lead to confusion over whether "Riemann integrable" refers to Definition 1.6.8 or Definition 1.6.22. Sometimes we will (as we have done in some of the results above) simply say which definition we intend to apply. But even in absence of this, it will be possible to deduce from the context which definition is intended. If both the domain and the function are explicitly given as being bounded, then we intend Definition 1.6.8. Otherwise we intend Definition 1.6.22. The reader should also keep in mind that the more general Definition 1.6.22 is derived as a limit of integrals defined using Definition 1.6.8. •

1.6.6 The Riemann integral and operations on functions

In this section we give the results that relate the Riemann integral to the various algebraic and other operations one can perform on $\mathbb{R}$-valued functions.

First let us indicate how the Riemann integral behaves with respect to the usual algebraic operations.

1.6.28 Proposition (Algebraic operations and the Riemann integral) *Let* $A \subseteq \mathbb{R}$*, let* $f, g\colon A \to \mathbb{R}$ *be Riemann integrable functions (as per Definition 1.6.22), and let* $c \in \mathbb{R}$*. Then the following statements hold:*

(i) $f + g$ *is Riemann integrable and*

$$\int_A (f+g)(\mathbf{x})\,d\mathbf{x} = \int_A f(\mathbf{x})\,d\mathbf{x} + \int_A g(\mathbf{x})\,d\mathbf{x};$$

(ii) cf *is Riemann integrable and*

$$\int_A (cf)(\mathbf{x})\,d\mathbf{x} = c\int_A f(\mathbf{x})\,d\mathbf{x};$$

(iii) if A *is additionally bounded and if* f *and* g *are bounded, then* fg *is Riemann integrable;*

(iv) if A *is additionally bounded, if* f *and* g *are bounded, and if there exists* $\alpha \in \mathbb{R}_{>0}$ *such that* $g(\mathbf{x}) \geq \alpha$ *for each* $\mathbf{x} \in A$*, then* $\frac{f}{g}$ *is Riemann integrable.*

Proof The proof goes essentially like the proof of Proposition I-3.4.22, with appropriate changes of notation. ■

We can also consider composing Riemann integrable functions with continuous functions on the left. Compositions on the right are more complicated and are the topic of Section 1.6.8.

1.6.29 Proposition (Function composition and the Riemann integral) *If* $A \subseteq \mathbb{R}^n$ *is bounded, if* $f\colon A \to \mathbb{R}$ *is a bounded Riemann integrable function satisfying* $\operatorname{image}(f) \subseteq [c, d]$*, and if* $g\colon [c, d] \to \mathbb{R}$ *is continuous, then* $g \circ f$ *is Riemann integrable.*

Proof The proof closely follows the proof of Proposition I-3.4.23 with only changes of notation necessary. ■

The Riemann integral also behaves in the expected manner with respect to the total order on $\mathbb{R}$ and the absolute value function.

1.6.30 Proposition (Riemann integral and total order on $\mathbb{R}$) *Let* $A \subseteq \mathbb{R}^n$ *and let* $f, g\colon A \to \mathbb{R}$ *be Riemann integrable functions for which* $f(\mathbf{x}) \leq g(\mathbf{x})$ *for each* $\mathbf{x} \in A$*. Then*

$$\int_A f(\mathbf{x})\,d\mathbf{x} \leq \int_A g(\mathbf{x})\,d\mathbf{x}.$$

Proof The proof here can easily be adapted from the proof of Proposition I-3.4.24. ■

1.6.31 Proposition (Absolute value and Riemann integral) *Let* $A \subseteq \mathbb{R}^n$*, let* $f\colon A \to \mathbb{R}$*, and define* $|f|\colon A \to \mathbb{R}$ *by* $|f|(\mathbf{x}) = |f(\mathbf{x})|$*. Then* $|f|$ *is Riemann integrable if* f *is Riemann integrable, and, in this case,*

$$\left|\int_A f(\mathbf{x})\,d\mathbf{x}\right| \le \int_A |f|(\mathbf{x})\,d\mathbf{x}.$$

Proof The proof here is just like the proof of Proposition I-3.4.25 with suitable changes of notation. ∎

The final result we state in this section is a generalisation of Proposition I-3.4.26 in the single-variable case. For the multivariable integral the statement is more complicated since the sets over which one integrates are more complicated. But the idea is similar.

1.6.32 Proposition (Breaking the Riemann integral in two) *Let* $A, B \subseteq \mathbb{R}^n$ *be such that* $A \cap B$ *has measure zero and let* $f\colon A \cup B \to \mathbb{R}$*. Then the following statements hold:*

(i) if $f|A$*,* $f|B$*, and* $f|(A \cap B)$ *are Riemann integrable then* f *is Riemann integrable;*

(ii) if f *is Riemann integrable and if* A*,* B*, and* $A \cap B$ *are Jordan measurable then* $f|A$ *and* $f|B$ *is Riemann integrable.*

Moreover, if the hypotheses of either of the above statements hold then

$$\int_{A\cup B} f(\mathbf{x})\,d\mathbf{x} = \int_A f(\mathbf{x})\,d\mathbf{x} + \int_B f(\mathbf{x})\,d\mathbf{x}.$$

Proof Let us first suppose that A and B are bounded and that f is bounded. Then define

$$f_A = f\chi_A, \quad f_B = f\chi_B, \quad f_{A\cap B} = f\chi_{A\cap B}. \tag{1.33}$$

The functions f_A, f_B, and $f_{A\cap B}$ are Riemann integrable by hypothesis. Note that $f = f_A + f_B - f_{A\cap B}$. By Proposition 1.6.28 we have

$$\begin{aligned}\int_{A\cup B} f(x)\,dx &= \int_{A\cup B} f_A(x)\,dx + \int_{A\cup B} f_B(x)\,dx - \int_{A\cup B} f_{A\cap B}(x)\,dx\\ &= \int_A f(x)\,dx + \int_B f(x)\,dx - \int_{A\cap B} f(x)\,dx.\end{aligned}$$

By Proposition 1.6.21(i) the third integral is zero and so this part of the result follows. Now, if A, B, and f are not assumed to be bounded, then the result follows from Proposition I-2.3.23 since the integrals in this case are limits of integrals of bounded functions over bounded domains.

Again, first assume that A, B, and f are bounded and define f_A, f_B, and $f_{A\cap B}$ as in (1.33). By Exercise 1.6.2 it follows that f_A, f_B, and $f_{A\cap B}$ are Riemann integrable. The same computation as in the preceding part of the proof now gives

$$\int_{A\cup B} f(x)\,dx = \int_A f(x)\,dx + \int_B f(x)\,dx. \tag{1.34}$$

If A, B, and f are not bounded then it follows from Definition 1.6.22 that f_A and f_B are Riemann integrable. The equality (1.34) then follows in this case from Proposition I-2.3.23. ∎

1.6.7 Fubini's Theorem

In our discussion of the single-variable Riemann integral in Section I-3.4 we investigated the basic properties of the Riemann integral. While we did not spend a lot of time discussing this, the fact is that one can explicitly compute single-variable integrals in many cases, and it is valuable to be able to do this. However, the explicit computation of multivariable integrals is more difficult. It would be helpful, therefore, were we able to somehow reduce the computation of multivariable integrals to the single-variable case. This provides some of the motivation for Fubini's Theorem which we discuss here. Another motivation for Fubini's Theorem is that it describes when one can switch the order of integration in multivariable integrals. This is something we will subsequently wish to do often.

Fubini's Theorem has to do with integrals over products. Thus the first thing we do is consider $\mathbb{R}^n$ to be a product: $\mathbb{R}^n = \mathbb{R}^m \times \mathbb{R}^{n-m}$ for some $m \in \{1, \dots, n-1\}$. With this product in mind, we write points in $\mathbb{R}^n$ as $\boldsymbol{x} = (\boldsymbol{y}, \boldsymbol{z})$ for $\boldsymbol{y} \in \mathbb{R}^m$ and $\boldsymbol{z} \in \mathbb{R}^{n-m}$. If $A \subseteq \mathbb{R}^m \times \mathbb{R}^{n-m}$ and if $f\colon A \to \mathbb{R}$ then we think of f as being defined on all of $\mathbb{R}^m \times \mathbb{R}^{n-m}$ is the usual manner, by asking that it be zero off A. We shall, therefore and without loss of generality, think of all functions be being defined on all of $\mathbb{R}^m \times \mathbb{R}^{n-m}$. For $\boldsymbol{y} \in \mathbb{R}^m$ define $f_{\boldsymbol{y}}\colon \mathbb{R}^{n-m} \to \mathbb{R}$ by $f_{\boldsymbol{y}}(\boldsymbol{z}) = f(\boldsymbol{y}, \boldsymbol{z})$. Thus $f_{\boldsymbol{y}}$ is the restriction of f to the set $\{\boldsymbol{y}\} \times \mathbb{R}^{n-m}$. Similarly we denote by $f^{\boldsymbol{z}}$ the restriction of f to $\mathbb{R}^m \times \{\boldsymbol{z}\}$.

Now we can state the result.

1.6.33 Theorem (Fubini's Theorem) *Let $n, m \in \mathbb{Z}_{>0}$ satisfy $n \geq 2$ and $m \in \{1, \dots, n-1\}$. Suppose that $f\colon \mathbb{R}^m \times \mathbb{R}^{n-m} \to \mathbb{R}$ has the following properties:*

(i) f is Riemann integrable in the sense of Definition 1.6.22;

(ii) for each $\mathbf{y} \in \mathbb{R}^m$, $f_{\mathbf{y}}$ is Riemann integrable in the sense of Definition 1.6.22.

Then the function

$$\mathbf{y} \mapsto \int_{\mathbb{R}^{n-m}} f_{\mathbf{y}}(\mathbf{z})\, d\mathbf{z}$$

is Riemann integrable in the sense of Definition 1.6.22 and

$$\int_{\mathbb{R}^n} f(\mathbf{x})\, d\mathbf{x} = \int_{\mathbb{R}^m} \Big(\int_{\mathbb{R}^{n-m}} f_{\mathbf{y}}(\mathbf{z})\, d\mathbf{z} \Big) d\mathbf{y}.$$

Similarly, if instead of (ii) we ask that

(iii) for each $\mathbf{z} \in \mathbb{R}^{n-m}$, $f^{\mathbf{z}}$ is Riemann integrable in the sense of Definition 1.6.22,

then the function

$$\mathbf{z} \mapsto \int_{\mathbb{R}^m} f^{\mathbf{z}}(\mathbf{y})\, d\mathbf{y}$$

is Riemann integrable in the sense of Definition 1.6.22 and

$$\int_{\mathbb{R}^n} f(\mathbf{x})\, d\mathbf{x} = \int_{\mathbb{R}^{n-m}} \Big(\int_{\mathbb{R}^m} f^{\mathbf{z}}(\mathbf{y})\, d\mathbf{y} \Big) d\mathbf{z}.$$

Proof We first consider the case where f is the extension to $\mathbb{R}^n$ of a bounded function defined on a bounded subset $A \subseteq \mathbb{R}^n$. In this case we can use Definition 1.6.8 for the Riemann integral. And in this case we have the following lemma.

1 Lemma *Let* $\mathrm{n}, \mathrm{m} \in \mathbb{Z}_{>0}$ *satisfy* $\mathrm{n} \geq 2$ *and* $\mathrm{m} \in \{1, \dots, \mathrm{n}-1\}$ *and let* $\mathrm{R} \subseteq \mathbb{R}^{\mathrm{m}}$ *and* $\mathrm{S} \subseteq \mathbb{R}^{\mathrm{n}-\mathrm{m}}$ *be fat compact rectangles. Let* $\mathrm{f}\colon \mathrm{R} \times \mathrm{S} \to \mathbb{R}$ *be Riemann integrable and define*

$$\underline{\mathrm{F}}(\mathbf{y}) = \mathrm{I}_-(\mathrm{f}_{\mathbf{y}}), \quad \overline{\mathrm{F}}(\mathbf{y}) = \mathrm{I}_+(\mathrm{f}_{\mathbf{y}}).$$

Then $\underline{\mathrm{F}}$ *and* $\overline{\mathrm{F}}$ *are Riemann integrable on* R *and*

$$\int_{\mathrm{R}\times\mathrm{S}} \mathrm{f}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\mathrm{R}} \underline{\mathrm{F}}(\mathbf{y})\,\mathrm{d}\mathbf{y} = \int_{\mathrm{R}} \overline{\mathrm{F}}(\mathbf{y})\,\mathrm{d}\mathbf{y}.$$

Proof Let $\boldsymbol{P} = (P_1, \dots, P_m)$ be a partition of R and let $\boldsymbol{Q} = (Q_1, \dots, Q_{n-m})$ be a partition of S. Note that

$$\boldsymbol{P} \times \boldsymbol{Q} \triangleq (P_1, \dots, P_m, Q_1, \dots, Q_{n-m})$$

is then a partition of $R \times S$. For simplicity, let us denote by $\mathscr{R}_{\boldsymbol{P}}$, $\mathscr{R}_{\boldsymbol{Q}}$, and $\mathscr{R}_{\boldsymbol{P}\times\boldsymbol{Q}}$ the subrectangles of $\boldsymbol{P}$, $\boldsymbol{Q}$, and $\boldsymbol{P} \times \boldsymbol{Q}$, respectively. Note that a typical element of $\mathscr{R}_{\boldsymbol{P}\times\boldsymbol{Q}}$ has the form $R_{\boldsymbol{P}} \times S_{\boldsymbol{Q}}$ where $R_{\boldsymbol{P}}$ is a subrectangle of $\boldsymbol{P}$ and $S_{\boldsymbol{Q}}$ is a subrectangle of $\boldsymbol{Q}$. Let us denote

$$m_{R_{\boldsymbol{P}}\times S_{\boldsymbol{Q}}}(f) = \inf\{f(\boldsymbol{y}, \boldsymbol{z}) \mid \boldsymbol{y} \in \mathrm{cl}(R_{\boldsymbol{P}}),\ \boldsymbol{z} \in \mathrm{cl}(S_{\boldsymbol{Q}})\}$$

and

$$m_{S_{\boldsymbol{Q}}}(f_{\boldsymbol{y}}) = \inf\{f_{\boldsymbol{y}}(\boldsymbol{z}) \mid \boldsymbol{z} \in \mathrm{cl}(S_{\boldsymbol{Q}})\}.$$

Since

$$m_{R_{\boldsymbol{P}}\times S_{\boldsymbol{Q}}}(f) \leq m_{S_{\boldsymbol{Q}}}(f_{\boldsymbol{y}})$$

for each $\boldsymbol{y} \in R_{\boldsymbol{P}}$ we compute

$$\sum_{S_{\boldsymbol{Q}}\in\mathscr{R}_{\boldsymbol{Q}}} m_{R_{\boldsymbol{P}}\times S_{\boldsymbol{Q}}}(f)\mathrm{vol}(S_{\boldsymbol{Q}}) \leq \sum_{S_{\boldsymbol{Q}}\in\mathscr{R}_{\boldsymbol{Q}}} m_{S_{\boldsymbol{Q}}}(f_{\boldsymbol{y}})\mathrm{vol}(S_{\boldsymbol{Q}}) \leq \underline{F}(\boldsymbol{y}).$$

Therefore,

$$\sum_{S_{\boldsymbol{Q}}\in\mathscr{R}_{\boldsymbol{Q}}} m_{R_{\boldsymbol{P}}\times S_{\boldsymbol{Q}}}(f)\mathrm{vol}(S_{\boldsymbol{Q}}) \leq \inf\{\underline{F}(\boldsymbol{y}) \mid \boldsymbol{y} \in \mathrm{cl}(R_{\boldsymbol{P}})\} \triangleq m_{R_{\boldsymbol{P}}}(\underline{F}).$$

$$\begin{aligned} A_-(f, \boldsymbol{P}\times\boldsymbol{Q}) &= \sum_{R_{\boldsymbol{P}}\times S_{\boldsymbol{Q}}\in\mathscr{R}_{\boldsymbol{P}\times\boldsymbol{Q}}} m_{R_{\boldsymbol{P}}\times S_{\boldsymbol{Q}}}(f)\mathrm{vol}(R_{\boldsymbol{P}}\times S_{\boldsymbol{Q}}) \\ &= \sum_{R_{\boldsymbol{P}}\in\mathscr{R}_{\boldsymbol{P}}} \Big(\sum_{S_{\boldsymbol{Q}}\in\mathscr{R}_{\boldsymbol{Q}}} m_{R_{\boldsymbol{P}}\times S_{\boldsymbol{Q}}}(f)\mathrm{vol}(S_{\boldsymbol{Q}})\Big)\mathrm{vol}(R_{\boldsymbol{P}}) \\ &\leq \sum_{R_{\boldsymbol{P}}\in\mathscr{R}_{\boldsymbol{P}}} m_{R_{\boldsymbol{P}}}(\underline{F})\mathrm{vol}(R_{\boldsymbol{P}}) = A_-(\underline{G}, \boldsymbol{P}). \end{aligned}$$

An entirely similarly styled computation yields

$$A_+(\overline{F}, \boldsymbol{P}) \leq A_+(f, \boldsymbol{P}\times\boldsymbol{Q}).$$

Thus we have

$$A_-(f, \boldsymbol{P} \times \boldsymbol{Q}) \leq A_-(\underline{F}, \boldsymbol{P}) \leq A_+(\underline{F}, \boldsymbol{P}) \leq A_+(\overline{F}, \boldsymbol{P}) \leq A_+(f, \boldsymbol{P} \times \boldsymbol{Q}).$$

Since f is Riemann integrable, for $\epsilon \in \mathbb{R}_{>0}$ there exists partitions $\boldsymbol{P}$ of R and $\boldsymbol{Q}$ of S such that

$$A_+(f, \boldsymbol{P} \times \boldsymbol{Q}) - A_-(f, \boldsymbol{P} \times \boldsymbol{Q}) < \epsilon$$

But this implies that

$$A_+(\underline{F}, \boldsymbol{P}) - A_-(\underline{F}, \boldsymbol{P}) < \epsilon,$$

and so $\underline{F}$ is Riemann integrable. Moreover, choosing $\boldsymbol{P}$ and $\boldsymbol{Q}$ such that

$$I(f) - A_-(f, \boldsymbol{P} \times \boldsymbol{Q}) < \tfrac{\epsilon}{3}, \quad A_-(f, \boldsymbol{P} \times \boldsymbol{Q}) - A_-(\underline{F}, \boldsymbol{P}) < \tfrac{\epsilon}{3}, \quad A_-(\underline{F}, \boldsymbol{P}) - I(f) < \tfrac{\epsilon}{3}$$

we have

$$I(f) - I(\underline{F}) = I(f) - A_-(f, \boldsymbol{P} \times \boldsymbol{Q}) + A_-(f, \boldsymbol{P} \times \boldsymbol{Q}) - A_-(\underline{F}, \boldsymbol{P}) + A_-(\underline{F}, \boldsymbol{P}) - I(f) < \epsilon.$$

This gives $I(f) = I(\underline{F})$, as desired.

The same argument, *mutatis mutandis*, can be used to show that $\overline{F}$ is Riemann integrable and that $I(f) = I(\overline{F})$. ▼

Now let us assume (ii) which implies that $\overline{F} = \underline{F}$. This then immediately gives the theorem in the case when f is bounded and is nonzero only on a bounded subset.

Now suppose that f is bounded, nonnegative, but possibly nonzero on an unbounded subset of $\mathbb{R}^n$. For $r \in \mathbb{R}_{>0}$ let

$$R_r = \underbrace{[-r, r] \times \cdots \times [-r, r]}_{m \text{ times}} \subseteq \mathbb{R}^m, \quad S_r = \underbrace{[-r, r] \times \cdots \times [-r, r]}_{n-m \text{ times}} \subseteq \mathbb{R}^{n-m}.$$

so that $R_r \times S_\rho$ is a rectangle in $\mathbb{R}^n$ for each $r, \rho \in \mathbb{R}_{>0}$. Then the hypotheses of the lemma are satisfied for $f|R_r \times S_\rho$. Therefore, the conclusions of the lemma ensure that the function

$$\boldsymbol{y} \mapsto \int_{S_\rho} f_{\boldsymbol{y}}(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z}$$

is Riemann integrable on R_r and that

$$\int_{R_r \times S_\rho} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{R_r} \Big(\int_{S_\rho} f_{\boldsymbol{y}}(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z} \Big) \mathrm{d}\boldsymbol{y}.$$

Since f is nonnegative this gives

$$\lim_{\rho \to \infty} \int_{R_r \times S_\rho} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{R_r \times \mathbb{R}^{n-m}} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{R_r} \Big(\int_{\mathbb{R}^{n-m}} f_{\boldsymbol{y}}(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z} \Big) \mathrm{d}\boldsymbol{y}.$$

This gives the Riemann integrability of the function

$$\boldsymbol{y} \mapsto \int_{\mathbb{R}^{n-m}} f_{\boldsymbol{y}}(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z}$$

defined on R_r. Again since f is nonnegative this gives

$$\lim_{r\to\infty}\int_{R_r\times\mathbb{R}^{n-m}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^n} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^m}\Big(\int_{\mathbb{R}^{n-m}} f_{\boldsymbol{y}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}\Big)\,\mathrm{d}\boldsymbol{y},$$

giving the Riemann integrability of the function

$$\boldsymbol{y}\mapsto\int_{\mathbb{R}^{n-m}} f_{\boldsymbol{y}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$$

defined on $\mathbb{R}^m$, and moreover the conclusions of the theorem in this case.

Next suppose that f is nonnegative, but possibly unbounded. For each $M\in\mathbb{R}_{>0}$ the function f_M defined by $f_M(\boldsymbol{x}) = \min\{f(\boldsymbol{x}), M\}$ is then bounded and so by our previous computations the function

$$\boldsymbol{y}\mapsto\int_{\mathbb{R}^{n-m}} f_{M,\boldsymbol{y}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$$

is Riemann integrable and

$$\int_{\mathbb{R}^n} f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}^m}\Big(\int_{\mathbb{R}^{n-m}} f_{M,\boldsymbol{y}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}\Big)\mathrm{d}\boldsymbol{y}.$$

We then have

$$\lim_{M\to\infty}\int_{\mathbb{R}^n} f_M(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \lim_{M\to\infty}\int_{\mathbb{R}^m}\Big(\int_{\mathbb{R}^{n-m}} f_{M,\boldsymbol{y}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}\Big)\mathrm{d}\boldsymbol{y}.$$

This gives the Riemann integrability of the function

$$\boldsymbol{y}\mapsto\int_{\mathbb{R}^{n-m}} f_{\boldsymbol{y}}(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$$

and the theorem in this case.

The final case is the general one where f takes on both positive and negative values and is possibly bounded. In this case, however, the theorem applied separately to both the positive and negative parts of f, and so also applied to f itself.

The assertion of the theorem where condition (ii) is replaced with (iii) is the same as the assertion we have proved with the order of $\boldsymbol{y}$ and $\boldsymbol{z}$ swapped. Thus it follows from what we have already proved by considering the function g on $\mathbb{R}^{n-m}\times\mathbb{R}^m$ defined by $g(\boldsymbol{z},\boldsymbol{y}) = f(\boldsymbol{y},\boldsymbol{z})$. ■

A commonly encountered situation where the hypotheses of Fubini's Theorem are satisfied is the following.

1.6.34 Corollary (Fubini's Theorem for continuous functions) *Let* $n, m\in\mathbb{Z}_{>0}$ *satisfy* $n\geq 2$ *and* $m\in\{1,\dots,n-1\}$ *and let* $f\colon\mathbb{R}^m\times\mathbb{R}^{n-m}\to\mathbb{R}$ *be continuous and Riemann integrable in the sense of Definition 1.6.22. Then the functions*

$$\mathbf{y}\mapsto\int_{\mathbb{R}^{n-m}} f_{\mathbf{y}}(\mathbf{z})\,\mathrm{d}\mathbf{z},\qquad \mathbf{z}\mapsto\int_{\mathbb{R}^m} f^{\mathbf{z}}(\mathbf{y})\,\mathrm{d}\mathbf{y}$$

are Riemann integrable in the sense of Definition 1.6.22 and

$$\int_{\mathbb{R}^n} f(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{\mathbb{R}^m}\Big(\int_{\mathbb{R}^{n-m}} f_{\mathbf{y}}(\mathbf{z})\,\mathrm{d}\mathbf{z}\Big)\mathrm{d}\mathbf{y} = \int_{\mathbb{R}^{n-m}}\Big(\int_{\mathbb{R}^m} f^{\mathbf{z}}(\mathbf{y})\,\mathrm{d}\mathbf{y}\Big)\mathrm{d}\mathbf{z}.$$

Proof By the functions f_y and f^z are Riemann integrable, and with this in mind the corollary immediately follows from the theorem. ■ what

One must take some care with Fubini's Theorem to understand what are its hypotheses and what are its conclusions. For example, Fubini's Theorem requires the Riemann integrability of f; this is not a conclusion. The following example illustrates that one cannot expect a general version of Fubini's Theorem where the Riemann integrability of the restriction of f to $\mathbb{R}^m$ or $\mathbb{R}^{n-m}$ ensures the Riemann integrability of f.

1.6.35 Example (Integrability of f is required in Fubini's Theorem) Let us define a subset $B \subseteq \mathbb{R}^2$ by

$$B = \cup_{l \in \mathbb{Z}_{>0}} \left\{ \left(\tfrac{j}{2^l}, \tfrac{k}{2^l}\right) \;\middle|\; j,k \in \mathbb{Z} \text{ odd} \right\}$$

and take $A = B \cap [0,1] \times [0,1]$. Let us record some facts about the set A.

1 Lemma *For each* $y_0, z_0 \in [0,1]$ *the sets*

$$\{y \in [0,1] \mid (y, z_0) \in A\}, \quad \{z \in [0,1] \mid (y_0, z) \in A\}$$

are finite.

Proof Define

$$B_l = \left\{ \left(\tfrac{j}{2^l}, \tfrac{k}{2^l}\right) \;\middle|\; j,k \in \mathbb{Z} \text{ odd} \right\}$$
$$I_l = \left\{ \tfrac{j}{2^l} \;\middle|\; j \in \mathbb{Z} \text{ odd} \right\}.$$

For $j \in \mathbb{Z}$ odd and for $l \in \mathbb{Z}_{>0}$ note that j and 2^l are necessarily coprime. Thus $\frac{j}{2^l}$ is the coprime fractional representative of this rational number. For this reason, we have

$$\{y \mid (y, z_0) \in B\} = \begin{cases} I_l \times \{z_0\}, & z_0 \in I_l,\ l \in \mathbb{Z}_{>0}, \\ \emptyset, & \text{otherwise.} \end{cases}$$

Thus horizontal lines in $\mathbb{R}^2$ intersect B in sets of the form $I_l \times \{z_0\}$ for some $l \in \mathbb{Z}_{>0}$ when this intersection is nonempty. Since at most finitely many points in $I_l \times \{z_0\}$ lie in $[0,1]$ it immediately follows that $\{y \in [0,1] \mid (y, z_0) \in A\}$ is finite. Similarly one shows that $\{z \in [0,1] \mid (y_0, z) \in A\}$ is finite. ▼

2 Lemma $\mathrm{cl}(A) = [0,1] \times [0,1]$.

Proof We shall show that B is dense in $\mathbb{R}^2$. Let $(y_0, z_0) \in \mathbb{R}^2$ and let U be a neighbourhood of (y_0, z_0). Let $\epsilon \in \mathbb{R}_{>0}$ be such that $(y_0 - \epsilon, y_0 + \epsilon) \times (z_0 - \epsilon, z_0 + \epsilon) \subseteq U$. (That such an ϵ exists is clear geometrically. Since U contains an open disk about (y_0, z_0) of nonzero radius, it will also contain an open rectangle whose sides have nonzero length, cf. Exercise III-3.1.6.) Let $l \in \mathbb{Z}_{>0}$ be sufficiently large that $2^l > \frac{1}{2\epsilon}$. Now let $j, k \in \mathbb{Z}_{>0}$ be such that

$$j \le 2^l (y_0 - \epsilon) < j + 1, \quad k \le 2^l (z_0 - \epsilon) < k + 1.$$

Then

$$y_0 - \epsilon < \tfrac{j+1}{2^l} = \tfrac{j}{2^l} + \tfrac{1}{2^l} < y_0 - \epsilon + 2\epsilon = y_0 + \epsilon$$

and similarly

$$z_0 - \epsilon < \tfrac{k}{2^l} < z_0 + \epsilon.$$

Thus $(\frac{j}{2^l}, \frac{k}{2^l}) \subseteq U$, which shows, by Proposition 1.2.26, that $(y_0, z_0) \in \mathrm{cl}(B)$. ▼

Now let $f\colon [0,1] \times [0,1] \to \mathbb{R}$ be defined by

$$f(y,z) = \begin{cases} 1, & (y,z) \in A, \\ 0, & (y,z) \notin A. \end{cases}$$

By the first lemma it follows that f_y and f^z are Riemann integrable for each $y, z \in [0,1]$ (why?). By the second lemma it follows that f is not Riemann integrable (the reason being entirely analogous to that for Example I-3.4.10). •

Another subtlety in applying Fubini's Theorem can arise by assuming that the Riemann integrability of f *implies* the Riemann integrability of f_y or f^z. The next example shows that this is not generally true.

1.6.36 Example (Integrability of f does not imply integrability of f_y) We consider the function $f\colon [0,1] \times [0,1] \to \mathbb{R}$ defined by

$$f(y,z) = \begin{cases} 1, & y = \frac{1}{2},\ z \in \mathbb{Q}, \\ 0, & \text{otherwise.} \end{cases}$$

We claim first that f is Riemann integrable. To see this, for $\epsilon \in \mathbb{R}_{>0}$ sufficiently small that it makes sense, define a partition $\boldsymbol{P}_\epsilon = (P_{1,\epsilon}, P_{2,\epsilon})$ of $[0,1] \times [0,1]$ by

$$P_{1,\epsilon} = ([0, \tfrac{1}{2} - \tfrac{\epsilon}{2}), [\tfrac{1}{2} - \tfrac{\epsilon}{2}, \tfrac{1}{2} + \tfrac{\epsilon}{2}], (\tfrac{1}{2} + \tfrac{\epsilon}{2}, 1]), \quad P_{2,\epsilon} = [0,1].$$

We directly compute $I_+(f, \boldsymbol{P}_\epsilon) = \epsilon$ and $I_-(f, \boldsymbol{P}_\epsilon) = 0$. Thus f is indeed Riemann integrable with Riemann integral 0.

However, $f_{\frac{1}{2}}$ is not Riemann integrable as we saw in Example I-3.4.10. •

Let us give an illustration of Fubini's Theorem via a nontrivial but interesting example.

1.6.37 Example (Volume of an n-dimensional ball) For $r \in \mathbb{R}_{>0}$ we shall show that

$$\mathrm{vol}(\mathsf{B}^n(r, \mathbf{0})) = \begin{cases} \frac{\pi^{n/2}}{(\frac{n}{2})!} r^n, & n \text{ even}, \\ \frac{2^n \pi^{(n-1)/2} (\frac{n-1}{2})!}{n!} r^n, & n \text{ odd}, \end{cases} \tag{1.35}$$

where $\mathsf{B}^n(r, \mathbf{0})$ is the open ball of radius r in $\mathbb{R}^n$. We should first indicate that the open ball is a set whose volume exists.

1 **Lemma** *The set* $\mathsf{B}^n(r, \mathbf{0})$ *is Jordan measurable.*

Proof This follows from Theorem 1.9.45 below, but let us give an independent proof in this case.

According to Theorem 1.6.13 we must show that $\mathrm{bd}(\mathsf{B}^n(r, \mathbf{0}))$ has measure zero. Note that

$$\mathrm{bd}(\mathsf{B}^n(r, \mathbf{0})) = \{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_{\mathbb{R}^n} = r\}.$$

For $j \in \{1, \dots, n\}$ define

$$A_{j,+} = \{\boldsymbol{x} \in \mathrm{bd}(\mathsf{B}^n(r, \mathbf{0})) \mid x_j \geq (x_1^2 + \cdots + x_{j-1}^2 + x_{j+1}^2 + \cdots + x_n^2)^{1/2}\},$$
$$A_{j,-} = \{\boldsymbol{x} \in \mathrm{bd}(\mathsf{B}^n(r, \mathbf{0})) \mid x_j \leq -(x_1^2 + \cdots + x_{j-1}^2 + x_{j+1}^2 + \cdots + x_n^2)^{1/2}\}.$$

In Figure 1.11 we show these sets in the case when $n = 2$. In this case it is clear that

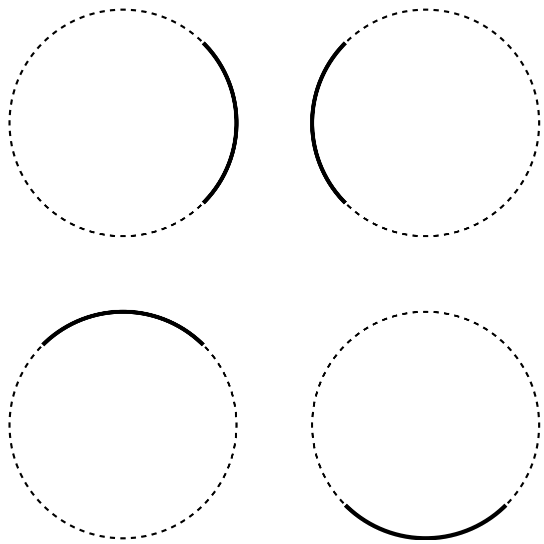

Figure 1.11 The sets $A_{1,+}$ (top left), $A_{1,-}$ (top right), $A_{2,+}$ (bottom left), and $A_{2,-}$ (bottom right)

the boundary of the ball is the union of the sets $A_{j,+}$ and $A_{j,-}$, $j \in \{1, 2\}$. Moreover, some elementary trigonometry shows that this is true for general n.

Now, by Exercise 1.6.3 it suffices to show that each of the sets $A_{j,+}$ and $A_{j,-}$, $j \in \{1, \dots, n\}$ has zero volume. For concreteness, let us consider in detail the set $A_{1,+}$; the argument for the other sets is the same. Let

$$\mathsf{V}_1 = \{\boldsymbol{x} \in \mathbb{R}^n \mid x_1 = 0\}$$

and define

$$B_1 = \left\{\boldsymbol{x} \in \mathsf{V}_1 \;\middle|\; \|\boldsymbol{x}\|_{\mathbb{R}^n} \leq \tfrac{1}{\sqrt{2}} r\right\}.$$

Thus B_1 is the closed ball of radius $\frac{1}{\sqrt{2}} r$ in the subspace V_1. Define $\phi_{1,+} \colon \mathsf{B}^n(r, \mathbf{0}) \to \mathbb{R}^n$ by

$$\phi_{1,+}(\boldsymbol{x}) = (x_1 - (r^2 - (x_2^2 + \cdots + x_n^2))^{1/2}, x_2, \dots, x_n)$$

and note that $\phi_{1,+}(B_1) = A_{1,+}$. One can verify that $\phi_{1,+}$ satisfies the conditions of the change of variables formula, Theorem 1.6.39, on the interior of any compact ball contained in $\mathsf{B}^n(r, \mathbf{0})$. Since B_1 has volume zero by Exercise 1.6.4, it follows from that $A_{1,+}$ has volume zero. ▼ what

Since we plan to prove our formula for the volume of the open ball by induction on n, let us denote by $V_n(r)$ the volume of the ball of radius r in $\mathbb{R}^n$. Since there can be differing conventions on what n means in this formula, let us be explicit:

$$V_n(r) = \mathrm{vol}(\{\boldsymbol{x} \in \mathbb{R}^n \mid \|\boldsymbol{x}\|_{\mathbb{R}^n} < r\}).$$

Note that $V_1(r) = 2r$.

To simplify things a little we prove the following "obvious" lemma.

2 Lemma $V_n(r) = V_n(1)r^n$.

Proof For $r \in \mathbb{R}_{>0}$ the map $\boldsymbol{x} \mapsto r\boldsymbol{x}$ maps $\mathsf{B}^n(1, \mathbf{0})$ bijectively onto $\mathsf{B}^n(r, \mathbf{0})$. Moreover, the hypotheses of the change of variables formula, Theorem 1.6.39, apply and so, since the Jacobian determinant of the map is obviously r^n,

$$\int_{\mathsf{B}^n(r,\mathbf{0})} \mathrm{d}\boldsymbol{x} = \int_{\mathsf{B}^n(1,\mathbf{0})} r^n \, \mathrm{d}\boldsymbol{x},$$

which is the result. ▼

With the lemma in mind let us define $v_n = V_n(1)$. It then suffices to determine v_n. As we have mentioned, this is done by induction, and the following lemma is key.

3 Lemma $\frac{v_{n+1}}{v_n} = 2\int_0^{\frac{\pi}{2}} (\cos\theta)^{n+1} \, \mathrm{d}\theta$ *for each* $n \in \mathbb{Z}_{>0}$.

Proof Let f be the characteristic function of the ball of unit radius in $\mathbb{R}^{n+1}$. Note that f is Riemann integrable, i.e., $\mathsf{B}^n(1, \mathbf{0})$ is Jordan measurable. Thus $v_{n+1} = \int_{\mathbb{R}^{n+1}} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}$. Let us write $\mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R}$ and denote a point in $\mathbb{R}^n \times \mathbb{R}$ as $(\boldsymbol{y}, z)$. For $z \in \mathbb{R}$, using the notation of Fubini's Theorem, note that

$$f^z(\boldsymbol{y}) = \begin{cases} 1, & \|\boldsymbol{y}\|_{\mathbb{R}^n} < \sqrt{1 - z^2},\ |z| < 1, \\ 0, & \text{otherwise.} \end{cases}$$

This function is clearly Riemann integrable for every $\boldsymbol{y} \in \mathbb{R}^n$ (it is a step function) and so the hypotheses of Fubini's Theorem apply. Thus we have

$$\begin{aligned} v_{n+1} &= \int_{\mathbb{R}^{n+1}} f(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} = \int_{\mathbb{R}} \Big(\int_{\mathbb{R}^n} f^z(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y} \Big) \mathrm{d}z \\ &= \int_{-1}^{1} V_n(\sqrt{1 - z^2}) \, \mathrm{d}z = v_n \int_{-1}^{1} (1 - z^2)^{n/2} \, \mathrm{d}z \end{aligned}$$

Thus the result will follow if we can show that

$$\int_{-1}^{1}(1-z^2)^{n/2}\,\mathrm{d}z = 2\int_0^{\frac{\pi}{2}}(\cos\theta)^{n+1}\,\mathrm{d}\theta. \tag{1.36}$$

By symmetry considerations,

$$\int_{-1}^{1}(1-z^2)^{n/2}\,\mathrm{d}z = 2\int_0^{1}(1-z^2)^{n/2}\,\mathrm{d}z.$$

Now we make a change of variable as per Proposition I-3.4.27. We let $f\colon [0,1]\to\mathbb{R}$ be defined by $f(z)=(1-z^2)^{n/2}$ and define $u\colon [0,\frac{\pi}{2}]\to[0,1]$ by $u(\theta)=\sin\theta$. Then

$$\int_0^{\frac{\pi}{2}} f\circ u(\theta)u'(\theta)\,\mathrm{d}\theta = \int_0^{\frac{\pi}{2}}(\cos\theta)^{n+1}\,\mathrm{d}\theta$$

and

$$\int_{u(0)}^{u(\frac{\pi}{2})} f(z)\,\mathrm{d}z = \int_0^{1}(1-z^2)^{n/2}\,\mathrm{d}z,$$

and so (1.36), and thus the lemma, follow. ▼

Now we have reduced ourselves to the computation of an integral. Let us do this computation.

4 Lemma $\displaystyle\int_0^{\frac{\pi}{2}}(\cos\theta)^{\mathrm{n}+1}\,\mathrm{d}\theta = \begin{cases}\frac{\pi}{2^{\mathrm{n}+1}}\frac{\mathrm{n}!}{(\frac{\mathrm{n}+1}{2})!(\frac{\mathrm{n}-1}{2})!}, & \mathrm{n}\ \textit{odd},\\ 2^{\mathrm{n}}\frac{(\frac{\mathrm{n}}{2}!)^2}{(\mathrm{n}+1)!}, & \mathrm{n}\ \textit{even}.\end{cases}$

Proof We use induction on n. Note that for $n\in\{1,2\}$ we have

$$\int_0^{\frac{\pi}{2}}(\cos\theta)^2\,\mathrm{d}\theta = \frac{\pi}{4},\quad \int_0^{\frac{\pi}{2}}(\cos\theta)^3\,\mathrm{d}\theta = \frac{2}{3},$$

as may be verified using the trigonometric identities

$$(\cos\theta)^2 = \tfrac{1}{2}(1+\cos(2\theta)),\quad (\cos\theta)^3 = \tfrac{3}{4}\cos\theta + \tfrac{1}{4}\cos(3\theta).$$

These expressions may be verified to agree with the asserted formulae of the lemma in this case. Using integration by parts we compute

$$\begin{aligned}\int_0^{\frac{\pi}{2}}(\cos\theta)^{n+1}\,\mathrm{d}\theta &= \sin\theta(\cos\theta)^n\Big|_{\theta=0}^{\theta=\frac{\pi}{2}} + n\int_0^{\frac{\pi}{2}}(\sin\theta)^2(\cos\theta)^{n-1}\,\mathrm{d}\theta\\ &= n\int_0^{\frac{\pi}{2}}(\cos\theta)^{n-1}\,\mathrm{d}\theta - n\int_0^{\frac{\pi}{2}}(\cos\theta)^{n+1}\,\mathrm{d}\theta.\end{aligned}$$

From this we conclude that

$$\int_0^{\frac{\pi}{2}} (\cos\theta)^{n+1}\,\mathrm{d}\theta = \frac{n}{n+1}\int_0^{\frac{\pi}{2}} (\cos\theta)^{n-1}\,\mathrm{d}\theta.$$

Now we use the induction hypothesis which gives

$$\int_0^{\frac{\pi}{2}} (\cos\theta)^{n-1}\,\mathrm{d}\theta = \begin{cases} \frac{\pi}{2^{n-1}}\frac{(n-2)!}{(\frac{n-1}{2})!(\frac{n-3}{2})!}, & n \text{ odd},\\ 2^{n-2}\frac{(\frac{n-2}{2}!)^2}{(n-1)!}, & n \text{ even}. \end{cases}$$

The lemma then follows by a direct computation for the even and odd cases. ▼

Now let us complete the computation of the constants v_n, $n \in \mathbb{Z}_{>0}$.

Let us consider the case when n is even. First of all we have

$$\frac{v_{n+1}}{v_{n-1}} = \frac{v_{n+1}}{v_n}\frac{v_n}{v_{n-1}} = 2^{n+1}\frac{(\frac{n}{2}!)^2}{(n+1)!}\frac{\pi}{2^n}\frac{(n-1)!}{(\frac{n}{2})!(\frac{n-2}{2})!} = \frac{2\pi}{n+1}.$$

Thus, for n even,

$$\begin{aligned} v_{n+1} &= v_1\frac{v_3}{v_1}\frac{v_5}{v_3}\cdots\frac{v_{n-1}}{v_{n-3}}\frac{v_{n+1}}{v_{n-1}} = 2\frac{2\pi}{3}\frac{2\pi}{5}\cdots\frac{2\pi}{n-1}\frac{2\pi}{n+1}\\ &= 2^{n/2+1}\pi^{n/2}\frac{2}{2\cdot 3}\frac{4}{4\cdot 5}\cdots\frac{n-2}{(n-2)(n-1)}\frac{n}{n(n+1)}\\ &= 2^{n/2+1}\pi^{n/2}2^{n/2}\frac{(\frac{n}{2})!}{(n+1)!} = 2^{n+1}\pi^{n/2}\frac{(\frac{n}{2})!}{(n+1)!}, \end{aligned}$$

using the fact that $v_1 = 2$. This gives the claimed answer in this case.

Now, when n is odd we have

$$\frac{v_{n+1}}{v_n} = \frac{v_{n+1}}{v_n}\frac{v_n}{v_{n-1}} = \frac{\pi}{2^n}\frac{n!}{(\frac{n+1}{2})!(\frac{n-1}{2})!}2^n\frac{(\frac{n-1}{2}!)^2}{n!} = \frac{2\pi}{n+1}.$$

Then

$$\begin{aligned} v_{n+1} &= v_2\frac{v_4}{v_2}\frac{v_6}{v_4}\cdots\frac{v_{n-1}}{v_{n-3}}\frac{v_{n+1}}{v_{n-1}} = \pi\frac{2\pi}{4}\frac{2\pi}{6}\cdots\frac{2\pi}{n-1}\frac{2\pi}{n+1}\\ &= \pi^{(n+1)/2}\frac{1}{2}\frac{1}{3}\cdots\frac{1}{\frac{n-1}{2}}\frac{1}{\frac{n+1}{2}} = \frac{\pi^{(n+1)/2}}{(\frac{n+1}{2})!}, \end{aligned}$$

using the fact that $v_2 = \pi$, as the reader can show in Exercise 1.6.6. This gives (1.35) when n is odd.

The formula for the volume of the n-dimensional ball can be simplified using so-called Γ-function,

$$\Gamma(x) = \int_0^\infty \mathrm{e}^{-y}y^{x-1}\,\mathrm{d}y.$$

By knowing some stuff about the Γ-function which we do not care to get into, one can show that

$$\mathrm{vol}(\mathsf{B}^n(r,\mathbf{0})) = \frac{\pi^{n/2} r^n}{\Gamma(\frac{n}{2}+1)}.$$

This is amusing, but only gives insight after one reduces back to the form involving factorials that we give.

See Exercise 1.6.7 for an interesting fact about the volume of balls as n gets large. •

1.6.8 The change of variables formula

In this section we give a quite general version of the change of variables formula for the multidimensional Riemann integral. This result is notoriously difficult to prove. It is also somewhat nonintuitive on a first encounter. It turns out that the best way to get insight into the formula is to first consider a special case. Since this result is actually used in the proof of the change of variables formula, it is worth recording it here.

1.6.38 Theorem (Volume and determinant) *If* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *is bounded and Jordan measurable, and if* $\mathbf{L} \in \mathrm{End}_{\mathbb{R}}(\mathbb{R}^{\mathrm{n}})$ *then* $\mathrm{vol}(\mathbf{L}(\mathrm{A})) = \det \mathbf{L}\, \mathrm{vol}(\mathrm{A})$*, i.e.,*

$$\int_{\mathbf{L}(\mathrm{A})} 1\, \mathrm{d}\mathbf{y} = \int_{\mathrm{A}} |\det \mathbf{L}|\, \mathrm{d}\mathbf{x}.$$

Proof We first consider the case where $A = R$ with R a compact rectangle:

$$R = [a_1, b_1] \times \cdots \times [a_n, b_n].$$

If $\det \boldsymbol{L} = 0$ then $\boldsymbol{L}(\mathbb{R}^n)$ is a subspace whose dimension is less than n by Theorem I-5.3.10. Thus $\boldsymbol{L}(R)$ is contained in this subspace and so has volume zero by Exercise 1.6.4. Thus the result holds in this case. When $\det \boldsymbol{L} \neq 0$ then L is invertible by Theorem I-5.3.10 and so $\boldsymbol{L}$ (as a matrix) is a product of elementary matrices by Theorem I-5.1.33. Let us then first consider the case when $\boldsymbol{L}$ is an elementary matrix. As described in Section I-5.1.5 there are three possibilities.

1. L *is the elementary matrix obtained from the identity matrix by swapping the ith and jth rows:* Let us suppose that $i < j$. In this case, by two uses of Fubini's Theorem, we have

$$\begin{aligned}\int_R \mathrm{d}\boldsymbol{x} &= \int_{a_1}^{b_1} \cdots \int_{a_i}^{b_i} \cdots \int_{a_j}^{b_j} \cdots \int_{a_n}^{b_n} \mathrm{d}x_n \cdots \mathrm{d}x_j \cdots \mathrm{d}x_i \cdots \mathrm{d}x_1 \\ &= \int_{a_1}^{b_1} \cdots \int_{a_j}^{b_j} \cdots \int_{a_i}^{b_i} \cdots \int_{a_n}^{b_n} \mathrm{d}x_n \cdots \mathrm{d}x_i \cdots \mathrm{d}x_j \cdots \mathrm{d}x_1 \\ &= \int_{\boldsymbol{L}(R)} \mathrm{d}\boldsymbol{x}.\end{aligned}$$

Since $|\det \boldsymbol{L}| = 1$ this gives

$$\int_R |\det \boldsymbol{L}| \mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{L}(R)} \mathrm{d}\boldsymbol{y},$$

as desired.

2. L *is the elementary matrix obtained from the identity matrix by multiplying the ith row by* $c \in \mathbb{R}^*$*:* When $c \in \mathbb{R}_{>0}$ we have

$$\begin{aligned}\int_{\boldsymbol{L}(R)} \mathrm{d}\boldsymbol{y} &= \int_{a_1}^{b_1} \cdots \int_{ca_i}^{cb_i} \cdots \int_{a_n}^{b_n} \mathrm{d}y_n \cdots \mathrm{d}y_i \cdots \mathrm{d}y_1 \\ &= (b_1 - a_1) \cdots (cb_i - ca_i) \cdots (b_n - a_n) \\ &= c(b_1 - a_1) \cdots (b_i - a_i) \cdots (b_n - a_n) \\ &= \int_{a_1}^{b_1} \cdots \int_{a_i}^{b_i} \cdots \int_{a_n}^{b_n} c\mathrm{d}x_n \cdots \mathrm{d}x_i \cdots \mathrm{d}x_1 \\ &= \int_R |\det \boldsymbol{L}|\, \mathrm{d}\boldsymbol{x},\end{aligned}$$

as claimed, where we have used the fact that $\det \boldsymbol{L} = c$ by Exercise I-5.3.8. When $c \in \mathbb{R}_{<0}$ the computation is similar:

$$\begin{aligned}\int_{\boldsymbol{L}(R)} \mathrm{d}\boldsymbol{y} &= \int_{a_1}^{b_1} \cdots \int_{cb_i}^{ca_i} \cdots \int_{a_n}^{b_n} \mathrm{d}y_n \cdots \mathrm{d}y_i \cdots \mathrm{d}y_1 \\ &= (b_1 - a_1) \cdots (ca_i - cb_i) \cdots (b_n - a_n) \\ &= -c(b_1 - a_1) \cdots (b_i - a_i) \cdots (b_n - a_n) \\ &= \int_{a_1}^{b_1} \cdots \int_{a_i}^{b_i} \cdots \int_{a_n}^{b_n} (-c)c\mathrm{d}x_n \cdots \mathrm{d}x_i \cdots \mathrm{d}x_1 \\ &= \int_R |\det \boldsymbol{L}|\, \mathrm{d}\boldsymbol{x},\end{aligned}$$

since we now have $|\det \boldsymbol{L}| = -c$.

3. L *is the elementary matrix obtained from the identity matrix by adding c times the ith row to the jth row:* We may suppose, without loss of generality, that $c = 1$. Indeed, if $c \neq 1$ then we can write the corresponding elementary matrix as a product of an elementary matrix of the form in the preceding case with one for which $c = 1$. Let us first consider the case when $n = 2$ (the case $n = 1$ is vacuous). In this case we have either

$$\boldsymbol{L} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad \boldsymbol{L} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}.$$

Let us consider the first case, as the second follows in a similar spirit. In this case,

$$\boldsymbol{L}(R) = \{(x_1, x_1 + x_2) \mid (x_1, x_2) \in [a_1, b_1] \times [a_2, b_2]\}.$$

In Figure 1.12 we show $\boldsymbol{L}(R)$ in the case when $b_1 - a_1 \geq b_2 - a_2$. The case when

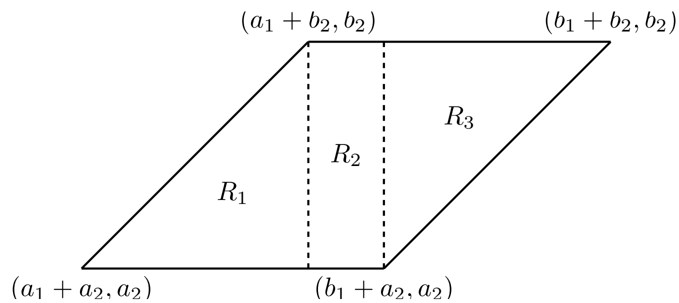

Figure 1.12 $\boldsymbol{L}(R)$

$b_2 - a_2 \geq b_1 - a_1$ is similar, so we shall only consider the case in the figure. In this case, $\boldsymbol{L}(R)$ can be divided into three parts, R_1, R_2, and R_3 as shown in the figure. In this case we have, using Fubini's Theorem,

$$\begin{aligned}\int_{\boldsymbol{L}(R)} \mathrm{d}\boldsymbol{x} &= \int_{R_1} \mathrm{d}\boldsymbol{x} + \int_{R_2} \mathrm{d}\boldsymbol{x} + \int_{R_3} \mathrm{d}\boldsymbol{x} \\ &= \int_{a_1+a_2}^{a_1+b_2} \int_{a_2}^{x_1-a_1} \mathrm{d}x_2 \mathrm{d}x_1 + \int_{a_1+b_2}^{b_1+a_2} \int_{a_2}^{b_2} \mathrm{d}x_2 \mathrm{d}x_1 + \int_{b_1+a_2}^{b_1+b_2} \int_{a_2}^{x_1-b_1} \mathrm{d}x_2 \mathrm{d}x_1 \\ &= \tfrac{1}{2}(b_2 - a_2)^2 + (b_2 - a_2)(b_1 - a_1 - (b_2 - a_2)) + \tfrac{1}{2}(b_2 - a_2)^2 \\ &= (b_1 - a_1)(b_2 - a_2) = \int_R \mathrm{d}\boldsymbol{y}.\end{aligned}$$

Since $\det \mathsf{L} = 1$ this gives our claim for $n = 2$. For general n, where the ith row is added to the jth row, let us define

$$\begin{aligned}A_1 &= \boldsymbol{L}(R) \cap \{\boldsymbol{x} \in \mathbb{R}^n \mid x_i = x_j = 0\} \\ A_2 &= \boldsymbol{L}(R) \cap \{\boldsymbol{x} \in \mathbb{R}^n \mid x_k = 0,\ k \in \{1, \dots, n\} \setminus \{i, j\}\}.\end{aligned}$$

Note that A_1 is a rectangle in $\mathbb{R}^{n-2}$ and that A_2 is a deformation of a rectangle in $\mathbb{R}^2$, like we considered above. Let us write a point in $\mathbb{R}^n$ as $(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathbb{R}^{n-2} \times \mathbb{R}^2$. By Fubini's Theorem,

$$\begin{aligned}\int_{\boldsymbol{L}(R)} \mathrm{d}\boldsymbol{x} = \int_{A_1} \int_{A_2} \mathrm{d}\boldsymbol{x}_2 \mathrm{d}\boldsymbol{x}_1 &= \int_{A_1} (b_i - a_1)(b_j - a_j)\, \mathrm{d}\boldsymbol{x}_1 \\ &= (b_1 - a_1) \cdots (b_{i-1} - a_{i-1})(b_{i+1} - a_{i+1})(b_{j-1} - a_{j-1}) \\ &\quad \cdot (b_{j+1} - a_{j+1}) \cdots (b_n - a_n)(b_i - a_i)(b_j - a_j) = \int_R \mathrm{d}\boldsymbol{y},\end{aligned}$$

giving the result in this case since $\det \boldsymbol{L} = 1$.

Having established the theorem in the case when A is a rectangle and $\boldsymbol{L}$ is an elementary matrix, let us establish the theorem in the case when A is an arbitrary bounded Jordan measurable set and $\boldsymbol{L}$ is an elementary matrix. Let R be a fat compact rectangle such that $A \subseteq R$. Since A is Jordan measurable choose a partition $\boldsymbol{P}_\epsilon$ of R for which

$$\operatorname{vol}(A) - A_-(\chi_A, \boldsymbol{P}_\epsilon) < \frac{\epsilon}{2|\det \boldsymbol{L}|}, \qquad A_+(\chi_A, \boldsymbol{P}_\epsilon) - \operatorname{vol}(A) < \frac{\epsilon}{2|\det \boldsymbol{L}|}.$$

Note that $A_-(\chi_A, \boldsymbol{P}_\epsilon)$ is, by definition of χ_A, the sum of the volumes of the subrectangles of $\boldsymbol{P}_\epsilon$ that lie entirely within A. Let $\underline{A}_\epsilon$ be the union of such rectangles. Similarly, $A_+(\chi_A, \boldsymbol{P}_\epsilon)$ is the sum of the volumes of the subrectangles of $\boldsymbol{P}_\epsilon$ that intersect A. Let $\overline{A}_\epsilon$ be the union of such rectangles. For each subrectangle R' of R we have $\mathrm{vol}(\boldsymbol{L}(R')) = |\det \boldsymbol{L}|\,\mathrm{vol}(R')$ by our computations above for elementary matrices. Thus

$$\mathrm{vol}(\boldsymbol{L}(\underline{A}_\epsilon)) = |\det \boldsymbol{L}|\, A_-(\chi_A, \boldsymbol{P}_\epsilon), \qquad \mathrm{vol}(\boldsymbol{L}(\overline{A}_\epsilon)) = |\det \boldsymbol{L}|\, A_+(\chi_A, \boldsymbol{P}_\epsilon).$$

Thus

$$\begin{aligned}\mathrm{vol}(\boldsymbol{L}(\overline{A}_\epsilon)) - \mathrm{vol}(\boldsymbol{L}(\underline{A}_\epsilon)) &= |\det \boldsymbol{L}|(A_+(\chi_A, \boldsymbol{P}_\epsilon) - A_-(\chi_A, \boldsymbol{P}_\epsilon))\\ &= |\det \boldsymbol{L}|(A_+(\chi_A, \boldsymbol{P}_\epsilon) - \mathrm{vol}(A) + \mathrm{vol}(A) - A_-(\chi_A, \boldsymbol{P}_\epsilon))\\ &< |\det \boldsymbol{L}|\Big(\frac{\epsilon}{2|\det \boldsymbol{L}|} + \frac{\epsilon}{2|\det \boldsymbol{L}|}\Big) = \epsilon.\end{aligned}$$

That is, the functions $\chi_{\boldsymbol{L}(\overline{A}_\epsilon)}$ and $\chi_{\boldsymbol{L}(\underline{A}_\epsilon)}$ satisfy

$$\chi_{\boldsymbol{L}(\underline{A}_\epsilon)}(\boldsymbol{x}) \le \chi_{\boldsymbol{L}(A)}(\boldsymbol{x}) \le \chi_{\boldsymbol{L}(\overline{A}_\epsilon)}(\boldsymbol{x}), \qquad \boldsymbol{x} \in \boldsymbol{L}(R)$$

and

$$\int_{\boldsymbol{L}(R)} \chi_{\boldsymbol{L}(\overline{A}_\epsilon)}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_{\boldsymbol{L}(R)} \chi_{\boldsymbol{L}(\underline{A}_\epsilon)}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} < \epsilon.$$

By Exercise 1.6.1 it follows that $\chi_{\boldsymbol{L}(A)}$ is Riemann integrable, or equivalently that $\boldsymbol{L}(A)$ is Jordan measurable. By Exercise 1.6.1 we also have

$$\mathrm{vol}(\boldsymbol{L}(A)) = \lim_{\epsilon\to 0} \mathrm{vol}(\boldsymbol{L}(\underline{A}_\epsilon)) = |\det \boldsymbol{L}| \lim_{\epsilon\to 0} A_-(\chi_A, \boldsymbol{P}_\epsilon) = |\det \boldsymbol{L}|\,\mathrm{vol}(A),$$

as desired.

Now we complete the proof of the theorem by letting A be an arbitrary bounded Jordan measurable set and letting $\boldsymbol{L}$ be an arbitrary invertible matrix. In this case we write $\boldsymbol{L} = \boldsymbol{L}_1 \cdots \boldsymbol{L}_k$ for elementary matrices $\boldsymbol{L}_1, \dots, \boldsymbol{L}_k$. We then have that $\boldsymbol{L}_k(A)$ is Jordan measurable, then that $\boldsymbol{L}_{k-1}\boldsymbol{L}_k(A)$ is Jordan measurable, and, continuing, that $\boldsymbol{L}(A)$ is Jordan measurable. Moreover,

$$\mathrm{vol}(\boldsymbol{L}_k(A)) = |\det \boldsymbol{L}_k|\,\mathrm{vol}(A), \quad \mathrm{vol}(\boldsymbol{L}_{k-1}\boldsymbol{L}_k(A)) = |\det \boldsymbol{L}_{k-1}||\det \boldsymbol{L}_k|\,\mathrm{vol}(A),$$

and, carrying on, we have

$$\mathrm{vol}(\boldsymbol{L}(A)) = |\det \boldsymbol{L}_1| \cdots |\det \boldsymbol{L}_k|\,\mathrm{vol}(A) = |\det \boldsymbol{L}|\,\mathrm{vol}(A),$$

using Proposition I-5.3.3(ii). ■

With this somewhat simple situation in hand, let us see if we can derive a plausibility argument for the change of variables formula before we formally state and prove it. A plausibility argument is the following. Suppose that we have an open subset $U \subseteq \mathbb{R}^n$ and a continuously differentiable map $\boldsymbol{\phi}\colon U \to \mathbb{R}^n$. Let $f\colon \boldsymbol{\phi}(U) \to \mathbb{R}$ be Riemann integrable. From Proposition 1.4.1 we know that, around $\boldsymbol{x}_0 \in U$, $\boldsymbol{\phi}$ is well approximated (in some sense) by its derivative. Moreover, if

$\det \boldsymbol{D\phi}(\boldsymbol{x}_0) \neq 0$, by this approximation is very good if one uses suitable coordinates. what Furthermore, the Riemann integrable function f is well approximated (in some sense) by a step function, i.e., by a function that is locally constant. Motivated by this, we propose that we can take $\boldsymbol{\phi}$ to be linear in some small rectangle R around $\boldsymbol{x}_0$. Precisely,

$$\boldsymbol{\phi}(\boldsymbol{x}) \approx \boldsymbol{D\phi}(\boldsymbol{x}_0) \cdot (\boldsymbol{x} - \boldsymbol{x}_0) = \underbrace{\boldsymbol{D\phi}(\boldsymbol{x}_0) \cdot \boldsymbol{x}}_{\text{linear}} - \underbrace{\boldsymbol{D\phi}(\boldsymbol{x}_0) \cdot \boldsymbol{x}_0}_{\text{constant}}$$

for $\boldsymbol{x} \in R$. We also take f to be constant on the set $\boldsymbol{\phi}(R)$ of $\boldsymbol{\phi}(\boldsymbol{x}_0)$ (i.e., $f(\boldsymbol{y}) = f(\boldsymbol{\phi}(\boldsymbol{x}_0))$ for $\boldsymbol{y} \in \boldsymbol{\phi}(U'))$. (Alert: $\boldsymbol{\phi}(R)$ may not contain a neighbourhood of $\boldsymbol{\phi}(\boldsymbol{x}_0$.) In this case,

$$\begin{aligned}\int_R f \circ \boldsymbol{\phi}(\boldsymbol{x}) |\det \boldsymbol{D\phi}(\boldsymbol{x}_0)| \, \mathrm{d}\boldsymbol{x} &= f(\boldsymbol{\phi}(\boldsymbol{x}_0)) |\det \boldsymbol{D\phi}(\boldsymbol{x}_0)| \int_R \mathrm{d}\boldsymbol{x} \\ &= f(\boldsymbol{\phi}(\boldsymbol{x}_0)) |\det \boldsymbol{D\phi}(\boldsymbol{x}_0)| \mathrm{vol}(R) \\ &= f(\boldsymbol{\phi}(\boldsymbol{x}_0)) \mathrm{vol}(\boldsymbol{\phi}(R)) \\ &= f(\boldsymbol{\phi}(\boldsymbol{x}_0)) \int_{\boldsymbol{\phi}(R)} \mathrm{d}\boldsymbol{y}.\end{aligned}$$

In the third step we have used Theorem 1.6.38 along with the plausible (and true, but as yet unproven) fact that the volume of a set is invariant under translations of the set. Now one can imagine applying the above procedure over the rectangles in a sequence of increasingly partitions of a rectangle that contains U and arriving at a formula

$$\int_U f \circ \boldsymbol{\phi}(\boldsymbol{x}) |\det \boldsymbol{D\phi}(\boldsymbol{x})| \, \mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{\phi}(U)} f(\boldsymbol{y}) \, \mathrm{d}\boldsymbol{y}.$$

This is the formula which we set out to verify under appropriate hypotheses. The proof of the following theorem is exceedingly long and detailed. However, since it is difficult to find statements and proofs of this theorem with the joint properties that (1) the statement is correct, (2) the statement is sufficiently general to cover the situations of common interest, and (3) the proof is correct, we give all of the details. But the reader not wanting to spend a week understanding the proof is advised to skip it.

1.6.39 Theorem (The change of variables formula for the Riemann integral) *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$, *let* $\mathrm{A} \subseteq \mathrm{U}$, *let* $\boldsymbol{\phi}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$, *and let* $\mathrm{f}\colon \boldsymbol{\phi}(\mathrm{A}) \to \mathbb{R}$ *have the following properties:*

(i) U *is open;*

(ii) A *is Jordan measurable with* $\mathrm{cl}(\mathrm{A}) \subseteq \mathrm{U}$*;*

(iii) $\boldsymbol{\phi}$ *has the following properties:*

(a) $\boldsymbol{\phi}$ *is continuously differentiable;*

(b) $\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(\mathrm{A}))$ *is injective;*

(c) $(\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(\mathrm{A})))^{-1}|\boldsymbol{\phi}(\mathrm{int}(\mathrm{cl}(\mathrm{A})))$ *is continuously differentiable;*

(iv) f *is Riemann integrable.*

Then

$$\int_{A} f\circ\boldsymbol{\phi}(\mathbf{x})|\det \mathbf{D}\boldsymbol{\phi}(\mathbf{x})|\,\mathrm{d}\mathbf{x} = \int_{\boldsymbol{\phi}(A)} f(\mathbf{y})\,\mathrm{d}\mathbf{y}.$$

Proof We shall prove the theorem under a sequence of hypotheses on U, A, $\boldsymbol{\phi}$, and f that get weaker as we go along, until in the end we arrive at a proof valid for the hypotheses of the theorem.

Let us introduce some notation. For $\boldsymbol{x}\in\mathbb{R}^n$ let us recall from Definition 1.1.9 that we denote

$$\|\boldsymbol{x}\|_\infty = \max\{|x_j| \mid j\in\{1,\dots,n\}\}.$$

Note that, for $\boldsymbol{x}_0\in\mathbb{R}^n$ and $r\in\mathbb{R}_{>0}$, the set

$$\{\boldsymbol{x}\in\mathbb{R}^n \mid \|\boldsymbol{x}-\boldsymbol{x}_0\| < r\}$$

is an open rectangle centred at $\boldsymbol{x}_0$ with sides of length $2r$. For $\boldsymbol{L}\in\mathrm{End}_{\mathbb{R}}(\mathbb{R}^n)$ we recall from Theorem 1.1.14 that the norm of L induced by the ∞-norm is

$$\|\boldsymbol{L}\|_{\infty,\infty} = \max\Big\{\sum_{j=1}^{n}|\boldsymbol{L}(k,j)| \,\Big|\, k\in\{1,\dots,n\}\Big\}.$$

Note that $\|\boldsymbol{L}\boldsymbol{x}\|_\infty \le \|\boldsymbol{L}\|_{\infty,\infty}\|\boldsymbol{x}\|_\infty$. In the proof we will occasionally implicitly use the fact that the norm $\|\cdot\|_\infty$ is equivalent to the norm $\|\cdot\|_{\mathbb{R}^n}$ in the sense of Definition III-3.1.13. For example, we will use the fact that continuity with respect to the norm $\|\cdot\|_{\mathbb{R}^n}$ is equivalent to continuity with respect to the norm $\|\cdot\|_\infty$. We will also use the fact that $\|\cdot\|_{\infty,\infty}$ is a norm on $\mathrm{End}_{\mathbb{R}}(\mathbb{R}^n)$, and that continuity of the derivative is equivalent to continuity with respect to this norm. All in all, a reader would benefit from understanding some basic material from Chapter III-3 in understanding this proof.

Case 1

The simplest case is the following:

1. U is bounded and Jordan measurable;
2. $A = U$;
3. $\boldsymbol{\phi}$ is linear, i.e., there exists $\boldsymbol{L}\in\mathrm{End}_{\mathbb{R}}(\mathbb{R}^n)$ such that $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{L}\boldsymbol{x}$ for each $\boldsymbol{x}\in U$;
4. f is constant, i.e., there exists $\alpha\in\mathbb{R}$ such that $f(\boldsymbol{y}) = \alpha$ for all $\boldsymbol{y}\in\boldsymbol{\phi}(A)$.

When $\alpha = 1$ this is exactly Theorem 1.6.38. For arbitrary α this case then follows from the linearity of the integral, Proposition 1.6.28.

Case 2

We suppose the following conditions hold:

1. U is open;
2. A is a compact rectangle (now denoted by R for the remainder of this case) with $R\subseteq U$;

3. ϕ has the following properties:
   (a) ϕ is continuously differentiable;
   (b) ϕ is injective;
   (c) $\phi^{-1}\colon \phi(U) \to U$ is continuously differentiable;
4. f is constant, i.e., there exists $\alpha \in \mathbb{R}$ such that $f(y) = \alpha$ for all $y \in \phi(U)$.

We prove a series of lemmata. Note that in the hypotheses of the lemmata the conditions on U, R, and ϕ may differ from the hypotheses above. This is a notational convenience, but the reader should be careful to understand what conditions are in effect at what place.

First we prove a lemma which shows that a bounded rectangle is mapped inside another bounded rectangle.

1 Lemma *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$, $\mathrm{R} \subseteq \mathrm{U}$, *and* $\phi\colon \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$ *have the following properties:*

(i) U *is open;*
(ii) $\mathrm{R} \subseteq \mathrm{U}$ *is a compact rectangle;*
(iii) ϕ *is continuously differentiable.*

Then $\phi(\mathrm{R})$ *is contained in a rectangle* R′ *satisfying* $\mathrm{vol}(\mathrm{R}') \leq \mathrm{M}^{\mathrm{n}}\,\mathrm{vol}(\mathrm{R})$, *where*

$$\mathrm{M} = \sup\left\{\|\mathbf{D}\phi(\mathbf{x})\|_{\infty,\infty} \;\middle|\; \mathbf{x} \in \mathrm{R}\right\}.$$

Proof First note that since R is compact and since $\boldsymbol{D}\phi$ is continuous, M is finite. Let x_0 be the centre of R and let us denote

$$\begin{aligned} x &= (x_1, \dots, x_n),\\ x_0 &= (x_{0,1}, \dots, x_{0,n}),\\ \phi(x) &= (\phi_1(x), \dots, \phi_n(x)). \end{aligned}$$

what? Applying the Mean Value Theorem, , we have

$$\begin{aligned} & \phi(x) - \phi(x_0) = \boldsymbol{D}\phi(x') \cdot (x - x_0) \text{ for some}\\ & x' \in \{s x_0 + (1-s)x \mid s \in [0,1]\},\\ \implies\quad & \|\phi(x) - \phi(x_0)\|_\infty \leq \|\boldsymbol{D}\phi(x')\|_{\infty,\infty}\|x - x_0\|_\infty \text{ for some}\\ & x' \in \{s x_0 + (1-s)x \mid s \in [0,1]\},\\ \implies\quad & |\phi_j(x) - \phi_j(x_0)| \leq M|x_j - x_{0,j}|, \qquad j \in \{1, \dots, n\}. \end{aligned}$$

Thus $\phi(R)$ is contained a rectangle whose edge parallel to the jth coordinate axis is at most M times the length of the corresponding edge for R. From this we have the lemma. ▼

Now we prove a lemma of a converse flavour, telling us that the image of a bounded rectangle contains a fat rectangle.

2 Lemma *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$*,* $\mathrm{R} \subseteq \mathrm{U}$*,* $\mathbf{x}_0 \in \mathrm{U}$*,* $\boldsymbol{\phi}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$*, and* $\sigma \in (0,1)$ *have the following properties:*

(i) U *is open;*

(ii) $\mathrm{R} \subseteq \mathrm{U}$ *is a fat compact rectangle with centre* $\mathbf{x}_0$*;*

(iii) $\boldsymbol{\phi}$ *has the following properties:*

(a) $\boldsymbol{\phi}$ *is continuously differentiable;*

(b) $\mathbf{D}\boldsymbol{\phi}(\mathbf{x}_0) = \mathbf{I}_{\mathrm{n}}$*;*

(c) $\|\mathbf{D}\boldsymbol{\phi}(\mathbf{x}) - \mathbf{D}\boldsymbol{\phi}(\mathbf{x}')\|_{\infty,\infty} \le \sigma$ *for all* $\mathbf{x}, \mathbf{x}' \in \mathrm{R}$.

Then $\boldsymbol{\phi}(\mathrm{R})$ *contains a rectangle* R' *centred at* $\boldsymbol{\phi}(\mathbf{x}_0)$ *satisfying* $\mathrm{vol}(\mathrm{R}') = (1-\sigma)^{\mathrm{n}}\mathrm{vol}(\mathrm{R})$.

Proof The first thing we do is "normalise" the problem to one that is more easily managed. First let us define

$$\tilde{U} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} + \boldsymbol{x}_0 \in U\}, \quad \tilde{R} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{x} + \boldsymbol{x}_0 \in R\}.$$

Thus we shift U and R so that $\mathbf{0}$ is the centre of $\tilde{R} \subseteq \tilde{U}$. Now let us define $\tilde{\boldsymbol{\phi}}\colon \tilde{U} \to \mathbb{R}^n$ by $\tilde{\boldsymbol{\phi}}(\boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{x} - \boldsymbol{x}_0) - \boldsymbol{\phi}(\boldsymbol{x}_0)$ so that $\tilde{\boldsymbol{\phi}}(\mathbf{0}) = \mathbf{0}$, $\boldsymbol{D}\tilde{\boldsymbol{\phi}}(\mathbf{0}) = \boldsymbol{I}_n$, and

$$\|\boldsymbol{D}\tilde{\boldsymbol{\phi}}(\boldsymbol{x}) - \boldsymbol{D}\tilde{\boldsymbol{\phi}}(\boldsymbol{x}')\|_{\infty,\infty} \le \sigma, \qquad \boldsymbol{x}, \boldsymbol{x}' \in \tilde{R}.$$

Thus we shift the image of the map $\boldsymbol{\phi}$ so that $\mathbf{0}$ is mapped to $\mathbf{0}$. The lemma will follow if we can show that $\tilde{\boldsymbol{\phi}}(\tilde{R})$ contains a rectangle with volume $(1-\sigma)^n \mathrm{vol}(R)$. That is to say, without loss of generality we can additionally assume in the hypotheses of the lemma that $\boldsymbol{x}_0 = \mathbf{0}$ and that $\boldsymbol{\phi}(\mathbf{0}) = \mathbf{0}$. This is stage one of our normalisation.

Now we assume that $\boldsymbol{\phi}$ is as determined in the preceding paragraph and normalise it further. Let us suppose that

$$R = [x_{0,1} - a_1, x_{0,1} + a_1] \times \cdots \times [x_{0,n} - a_n, x_{0,n} + a_n],$$

where $\boldsymbol{x}_0 = (x_{0,1}, \dots, x_{0,n})$ so that $\mathrm{vol}(R) = 2^n a_1 \cdots a_n$. Now define a linear map $\chi\colon \mathbb{R}^n \to \mathbb{R}^n$ by

$$\chi(x_1, \dots, x_n) = (a_1 x_1, \dots, a_n x_n).$$

We then define $U_1 = \chi^{-1}(U)$ and $R_1 = \chi^{-1}(R)$. Note that

$$R_1 = [-1, 1] \times \cdots \times [-1, 1].$$

so that $\mathrm{vol}(R_1) = 2^n$. Now define and $\boldsymbol{\phi}_1\colon U_1 \to \mathbb{R}^n$ by $\boldsymbol{\phi}_1 = \chi^{-1} \circ \boldsymbol{\phi} \circ \chi$. Note that $\boldsymbol{\phi}_1(\mathbf{0}) = 0$ and that $\boldsymbol{D}\boldsymbol{\phi}_1(\mathbf{0}) = \boldsymbol{I}_n$, the latter following from the Chain Rule. We claim that if $\boldsymbol{\phi}_1(R_1)$ contains a cube R_1' with $\mathrm{vol}(R_1') = (1-\sigma)^n \mathrm{vol}(R_1)$ then $\boldsymbol{\phi}(R)$ contains a rectangle R' with $\mathrm{vol}(R') = (1-\sigma)^n \mathrm{vol}(R)$. Indeed,

$$R_1' \subseteq \boldsymbol{\phi}_1(R_1) = \chi^{-1} \circ \boldsymbol{\phi} \circ \chi(R_1) \quad \Longrightarrow \quad \chi(R_1') \subseteq \boldsymbol{\phi}(\chi(R_1)) = \boldsymbol{\phi}(R).$$

We take $R' = \chi(R_1')$ and note that

$$\begin{aligned} \mathrm{vol}(R') = \mathrm{vol}(\chi(R_1')) &= \det \chi \mathrm{vol}(R_1') \ge (1-\sigma)^n \det \chi \mathrm{vol}(R_1) \\ &= (1-\sigma)^n 2^n a_1 \cdots a_n = (1-\sigma)^n \mathrm{vol}(R), \end{aligned}$$

as desired.

In summary, the above arguments show that we can add the following hypotheses to those of the lemma:

1. $\boldsymbol{x}_0 = \boldsymbol{0}$;
2. $\boldsymbol{\phi}(\boldsymbol{0}) = \boldsymbol{0}$;
3. $R = [-1, 1] \times \cdots \times [-1, 1]$.

Moreover, we have also shown that if we add these hypotheses, the conclusions of the lemma will follow if we can show that $\boldsymbol{\phi}(R)$ contains a cube R' for which $\mathrm{vol}(R') = (1-\sigma)^n \mathrm{vol}(R)$. This is what we now do.

Let $\boldsymbol{y}$ have the property that $\|\boldsymbol{y}\|_\infty \le (1-\sigma)$. Define $\psi_{\boldsymbol{y}} \colon \mathbb{R}^n \to \mathbb{R}^n$ be defined by

$$\psi_{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\phi}(\boldsymbol{x}) + \boldsymbol{y}.$$

If $\boldsymbol{x} \in R_1$ then

$$\begin{aligned}
\|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{x}\|_\infty &= \|\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{0}) - \boldsymbol{D\phi}(\boldsymbol{0}) \cdot \boldsymbol{x}\|_\infty \\
&= \|\boldsymbol{D\phi}(\boldsymbol{x}') \cdot \boldsymbol{x} - \boldsymbol{D\phi}(\boldsymbol{0}) \cdot \boldsymbol{x}\|_\infty \\
&\le \|\boldsymbol{x}\|_\infty \sup\left\{\|\boldsymbol{D\phi}(\boldsymbol{x}') - \boldsymbol{D\phi}(\boldsymbol{0})\|_{\infty,\infty} \;\middle|\; \boldsymbol{x}' \in R_1\right\} \le \sigma,
\end{aligned}$$

where $\boldsymbol{x}' \in R_1$ is such that

$$\boldsymbol{\phi}(\boldsymbol{x}) - \boldsymbol{\phi}(\boldsymbol{0}) = \boldsymbol{D\phi}(\boldsymbol{x}') \cdot (\boldsymbol{x} - \boldsymbol{0}),$$

using the Mean Value Theorem. Therefore, if $\boldsymbol{x} \in R_1$ and if $\|\boldsymbol{y}\|_\infty \le (1-\sigma)$,

$$\|\psi_{\boldsymbol{y}}(\boldsymbol{x})\|_\infty \le \sigma + (1-\sigma) = 1.$$

Thus $\psi_{\boldsymbol{y}}$ maps R_1 into R_1 provided that $\|\boldsymbol{y}\|_\infty \le (1-\sigma)$. Therefore, for $\boldsymbol{x}_1, \boldsymbol{x}_2 \in R_1$ and if $\|\boldsymbol{y}\|_\infty \le (1-\sigma)$,

$$\begin{aligned}
\|\psi_{\boldsymbol{y}}(\boldsymbol{x}_1) - \psi_{\boldsymbol{y}}(\boldsymbol{x}_2)\|_\infty &= \|\boldsymbol{x}_1 - \boldsymbol{x}_2 - (\boldsymbol{\phi}(\boldsymbol{x}_1) - \boldsymbol{\phi}(\boldsymbol{x}_2))\|_\infty \\
&= \|(\boldsymbol{x}_1 - \boldsymbol{x}_2) - \boldsymbol{D\phi}(\boldsymbol{x}') \cdot (\boldsymbol{x}_1 - \boldsymbol{x}_2)\|_\infty \\
&\le \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_\infty \|\boldsymbol{D\phi}(\boldsymbol{x}') - \boldsymbol{D\phi}(\boldsymbol{0})\|_{\infty,\infty} \\
&\le \sigma \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|_\infty,
\end{aligned}$$

where $\boldsymbol{x}' \in R_1$ is such that

$$\boldsymbol{\phi}(\boldsymbol{x}_1) - \boldsymbol{\phi}(\boldsymbol{x}_2) = \boldsymbol{D\phi}(\boldsymbol{x}') \cdot (\boldsymbol{x}_1 - \boldsymbol{x}_2),$$

using the Mean Value Theorem. Since $\sigma \in (0, 1)$ we can conclude from Theorem 1.11.3 that $\psi_{\boldsymbol{y}}$ possesses a unique fixed point $\boldsymbol{x} \in R_1$. Thus

$$\psi_{\boldsymbol{y}}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{\phi}(\boldsymbol{x}) + \boldsymbol{y} = \boldsymbol{x} \quad \Longrightarrow \quad \boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{y}.$$

Thus we have shown that $\boldsymbol{\phi}(R_1)$ contains a rectangle R' with sides of length at least $(1-\sigma)$. Note that

$$\mathrm{vol}(R') \ge 2^n (1-\sigma)^n = (1-\sigma)^n \mathrm{vol}(R),$$

as desired. ▼

We shall on several occasions require a condition guaranteeing that the image of a set of zero volume has zero volume.

3 Lemma *Let* $U \subseteq \mathbb{R}^n$, $A \subseteq U$, *and* $\phi\colon U \to \mathbb{R}^n$ *have the following properties:*

(i) U *is open;*

(ii) $\mathrm{vol}(A) = 0$ *and* $\mathrm{cl}(A) \subseteq U$*;*

(iii) ϕ *is continuously differentiable.*

Then $\mathrm{vol}(\phi(A)) = 0$*.*

Proof Note that A is bounded since, by definition of zero volume, it can be contained in the union of a finite number of compact rectangles. Thus $\mathrm{cl}(A)$ is closed and bounded and so compact by the Heine–Borel Theorem.

First we claim that if R is a compact rectangle intersecting A then it is possible to cover $A \cap R$ with compact rectangles $R_1, \dots, R_m$ having the properties that $R_j \subseteq R$ and $R_j \subseteq U$ for $j \in \{1, \dots, m\}$ and that $\sum_{j=1}^m \mathrm{vol}(R_j) \le \mathrm{vol}(R)$. To see this, note that $R \cap \mathrm{cl}(A)$ is compact since it is a closed subset of the compact set $\mathrm{cl}(A)$. Let $\boldsymbol{x} \in \mathrm{cl}(A) \cap R$. Since $\boldsymbol{x} \in U$ and since U is open there exists a fat compact rectangle $R_{\boldsymbol{x}}$ containing $\boldsymbol{x}$ in its interior and having the property that $R_{\boldsymbol{x}} \subseteq U$. Since $\mathrm{cl}(A) \cap R$ is compact there exists a finite subcover of the open cover $(\mathrm{int}(R_{\boldsymbol{x}}))_{\boldsymbol{x} \in \mathrm{cl}(A) \cap R}$. Let the corresponding rectangles in this subcover be denoted by $R_{\boldsymbol{x}_1}, \dots, R_{\boldsymbol{x}_m}$ and define $R_j = R_{\boldsymbol{x}_j} \cap R$ for $j \in \{1, \dots, m\}$. The compact rectangles $R_1, \dots, R_m$ are then contained in R, contained in U, and they cover $\mathrm{cl}(A) \cap R$. However, because of overlapping, it is possible that the total volume of the rectangles might exceed that of R. However, it is possible to write the rectangles $R_1, \dots, R_m$ as a finite disjoint union of rectangles (this is sort of obvious, but see Proposition III-2.1.2). These finitely many disjoint rectangles will then have a total volume not exceeding that of R.

Denote

$$M = \sup\left\{\|\boldsymbol{D}\phi(\boldsymbol{x})\|_{\infty,\infty} \;\middle|\; \boldsymbol{x} \in \mathrm{cl}(A)\right\}.$$

Since $\mathrm{cl}(A)$ is compact M is finite because $\boldsymbol{D}\phi$ is continuous. Now let $\epsilon \in \mathbb{R}_{>0}$ and let $R_1, \dots, R_k$ be rectangles covering A and such that $\sum_{j=1}^k \mathrm{vol}(R_j) < \frac{\epsilon}{M}$; this is possible since $\mathrm{vol}(A) = 0$. From the preceding paragraph it is possible to choose the rectangles $R_1, \dots, R_k$ so that $R_j \subseteq U$ for each $j \in \{1, \dots, k\}$. That $\phi(A)$ has zero volume now follows from Lemma 1. ▼

Let us next prove a lemma ensuring that certain kinds of sets we shall encounter are Jordan measurable.

4 Lemma *Let* $U \subseteq \mathbb{R}^n$, $A \subseteq U$, *and* $\phi\colon U \to \mathbb{R}^n$ *have the following properties:*

(i) U *is open;*

(ii) A *is bounded and Jordan measurable with* $\mathrm{cl}(A) \subseteq U$*, and*

(iii) ϕ *has the following properties:*

  (a) ϕ *is continuously differentiable;*

  (b) $\phi|\,\mathrm{int}(A)$ *is injective;*

  (c) $(\phi|\,\mathrm{int}(A))^{-1}|\phi(\mathrm{int}(A))$ *is continuously differentiable.*

Then $\phi(A)$ *is bounded and Jordan measurable.*

Proof First we show that $\mathrm{bd}(\phi(A)) \subseteq \phi(\mathrm{bd}(A))$. Let $\boldsymbol{y} \in \mathrm{bd}(\phi(A))$ so that there exists a sequence $(\boldsymbol{y}_j)_{j \in \mathbb{Z}_{>0}}$ in $\phi(A)$ converging to $\boldsymbol{y}$. Let $(\boldsymbol{x}_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in A such that $\phi(\boldsymbol{x}_j) = \boldsymbol{y}_j$ for each $j \in \mathbb{Z}_{>0}$. There are two possibilities.

1. *For every $k \in \mathbb{Z}_{>0}$ there exists $j_k \geq k$ such that $\boldsymbol{x}_{j_k} \notin \mathrm{bd}(A)$:* In this case the subsequence $(\boldsymbol{y}_{j_k})_{k\in\mathbb{Z}_{>0}}$ converges to $\boldsymbol{y}$ and each element $\boldsymbol{y}_{j_k}$ has the property that $\boldsymbol{y}_{j_k} = \boldsymbol{\phi}(\boldsymbol{x}_{j_k})$ for $\boldsymbol{x}_{j_k} \in \mathrm{int}(A)$. Let us rename the subsequence $(\boldsymbol{y}_{j_k})_{k\in\mathbb{Z}_{>0}}$ to $(\boldsymbol{y}_j)_{j\in\mathbb{Z}_{>0}}$ so that $\boldsymbol{y}_j = \boldsymbol{\phi}(\boldsymbol{x}_j)$ for $\boldsymbol{x}_j \in \mathrm{int}(A)$. Since $\mathrm{cl}(A)$ is compact, the sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ has a convergent subsequence $(\boldsymbol{x}_{j_k})_{k\in\mathbb{Z}_{>0}}$ converging to $\boldsymbol{x} \in \mathrm{cl}(A)$. We claim that $\boldsymbol{x} \in \mathrm{bd}(A)$. Indeed, if $\boldsymbol{x} \in \mathrm{int}(A)$ then
$$\boldsymbol{y} \in \boldsymbol{\phi}(\mathrm{int}(A)) = \mathrm{int}(\boldsymbol{\phi}(\mathrm{int}(A))) \subseteq \mathrm{int}(\boldsymbol{\phi}(A)),$$
using the fact that $\boldsymbol{\phi}(\mathrm{int}(A))$ is open. This contradicts the fact that $\boldsymbol{y} \in \mathrm{bd}(\boldsymbol{\phi}(A))$, and so we indeed must have $\boldsymbol{x} \in \mathrm{bd}(A)$. However, by Theorem 1.3.2, $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{y}$ since the sequence $(\boldsymbol{y}_{j_k})_{k\in\mathbb{Z}_{>0}}$ converges to $\boldsymbol{y}$. Thus $\boldsymbol{y} \in \boldsymbol{\phi}(\mathrm{bd}(A))$, as desired.
2. *There exists $N \in \mathbb{Z}_{>0}$ such that $\boldsymbol{x}_j \in \mathrm{bd}(A)$ for every $j \geq N$:* In this case we can without loss of generality (why?) suppose that $\boldsymbol{x}_j \in \mathrm{bd}(A)$ for every $j \in \mathbb{Z}_{>0}$. Because $\mathrm{bd}(A)$ is compact there exists a subsequence $(\boldsymbol{x}_{j_k})_{k\in\mathbb{Z}_{>0}}$ converging to some $\boldsymbol{x} \in \mathrm{bd}(A)$. But, just as in the previous case, $\boldsymbol{\phi}(\boldsymbol{x}) = \boldsymbol{y}$ and so $\boldsymbol{y} \in \boldsymbol{\phi}(\mathrm{bd}(A))$.

The lemma now follows from Lemma 3. ▼

Now we prove two lemmas that will allow us to conclude that, under the hypotheses of this case,
$$\mathrm{vol}(\boldsymbol{\phi}(R)) = \int_R |\det \boldsymbol{D\phi}(\boldsymbol{x})| \, \mathrm{d}\boldsymbol{x}.$$
We do this by establishing both inequalities.

5 Lemma *Let $\mathrm{U} \subseteq \mathbb{R}^n$, $\mathrm{R} \subseteq \mathrm{U}$, and $\boldsymbol{\phi}\colon \mathrm{U} \to \mathbb{R}^n$ have the following properties:*

(i) U is open;

(ii) R is a compact rectangle;

(iii) $\boldsymbol{\phi}$ has the following properties:

(a) $\boldsymbol{\phi}$ is continuously differentiable;

(b) $\boldsymbol{\phi}|\mathrm{U}$ is injective;

(c) $\boldsymbol{\phi}^{-1}\colon \boldsymbol{\phi}(\mathrm{U}) \to \mathrm{U}$ is continuously differentiable.

Then
$$\mathrm{vol}(\boldsymbol{\phi}(\mathrm{R})) \leq \int_{\mathrm{R}} |\det \mathbf{D}\boldsymbol{\phi}(\mathbf{x})| \, \mathrm{d}\mathbf{x}.$$

Proof Let $\boldsymbol{L} \in \mathrm{End}_{\mathbb{R}}(\mathbb{R}^n)$ be invertible and note that, by Theorem 1.6.38,
$$|\det \boldsymbol{L}^{-1}| \, \mathrm{vol}(\boldsymbol{\phi}(R)) = \mathrm{vol}(\boldsymbol{L}^{-1} \circ \boldsymbol{\phi}(R)).$$
Note that $\boldsymbol{x} \mapsto \boldsymbol{L}^{-1} \circ \boldsymbol{\phi}(\boldsymbol{x})$ is continuously differentiable and satisfies
$$\boldsymbol{D}(\boldsymbol{L}^{-1} \circ \boldsymbol{\phi})(\boldsymbol{x}) = \boldsymbol{L}^{-1} \circ \boldsymbol{D\phi}(\boldsymbol{x}).$$
Thus $\det \boldsymbol{D}(\boldsymbol{L}^{-1} \circ \boldsymbol{\phi})(\boldsymbol{x}) \neq 0$ for all $\boldsymbol{x} \in U$. By Lemmata 1 and 4 it then follows that $\boldsymbol{L} \circ \boldsymbol{\phi}(R)$ is bounded and Jordan measurable and satisfies
$$\mathrm{vol}(\boldsymbol{L}^{-1} \circ \boldsymbol{\phi}(R)) \leq M^n \, \mathrm{vol}(R)$$

where

$$M = \sup\left\{\|\boldsymbol{L} \circ \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})\|_{\infty,\infty} \;\middle|\; \boldsymbol{x} \in U\right\}.$$

Thus

$$\mathrm{vol}(\boldsymbol{\phi}(R)) \leq |\det \boldsymbol{L}|\, M^n \,\mathrm{vol}(R), \tag{1.37}$$

this equation holding for any invertible matrix $\boldsymbol{L}$.

Now let $\epsilon \in \mathbb{R}_{>0}$. Note that since $\boldsymbol{x} \mapsto |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|$ is continuous and bounded its restriction to R is Riemann integrable. Choose $\epsilon' \in \mathbb{R}_{>0}$ such that

$$\epsilon'\Big(1 + \int_R |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}\Big) + (\epsilon')^2 < \epsilon. \tag{1.38}$$

Choose $\delta_1 \in \mathbb{R}_{>0}$ such that, if $\|\boldsymbol{x} - \boldsymbol{x}'\|_\infty < \delta_1$, then

$$\Big|1 - \|\boldsymbol{D}\boldsymbol{\phi}^{-1}(\boldsymbol{x}) \circ \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x}')\|^n_{\infty,\infty}\Big| < \epsilon'. \tag{1.39}$$

This is possible since

$$\|\boldsymbol{D}\boldsymbol{\phi}^{-1}(\boldsymbol{x}) \circ \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})\|_{\infty,\infty} = 1, \qquad \boldsymbol{x} \in U,$$

and since $\|\cdot\|_{\infty,\infty}$, composition of matrices, and $\boldsymbol{D}\boldsymbol{\phi}$ are all uniformly continuous.

Now let $\delta_2 \in \mathbb{R}_{>0}$ be such that, if $\boldsymbol{P} = (P_1, \dots, P_n)$, $P_j = (I_{j1}, \dots, I_{jk_j})$, $j \in \{1, \dots, n\}$, is a partition of R satisfying $|\boldsymbol{P}| < \delta_2$ and if

$$\{\boldsymbol{\xi}_{l_1,\dots,l_n} \mid l_j \in \{1, \dots, k_j\},\ j \in \{1, \dots, n\}\}$$

is a selection from $\boldsymbol{P}$, then

$$\Big|\sum_{l_1=1}^{k_1} \cdots \sum_{l_n=1}^{k_n} |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{\xi}_{l_1,\dots,l_n})|\mathrm{vol}(R_{l_1,\dots,l_n}) - \int_R |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}\Big| < \epsilon'.$$

This implies that

$$\int_R |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} + \epsilon' < \sum_{l_1=1}^{k_1} \cdots \sum_{l_n=1}^{k_n} |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{\xi}_{l_1,\dots,l_n})|\mathrm{vol}(R_{l_1,\dots,l_n}). \tag{1.40}$$

Take $\delta = \min\{\delta_1, \delta_2\}$. Let $\boldsymbol{P} = (P_1, \dots, P_n)$ be a partition satisfying $|\boldsymbol{P}| < \delta$ and denote $P_j = (I_{j1}, \dots, I_{jk_j})$, $j \in \{1, \dots, n\}$. Let $\boldsymbol{x}_{l_1,\dots,l_n}$ be the centre of the subrectangle $R_{l_1,\dots,l_n}$, $l_j \in \{1, \dots, k_j\}$, $j \in \{1, \dots, n\}$. For brevity denote $L = \{1, \dots, k_1\} \times \cdots \times \{1, \dots, k_n\}$. For $(l_1, \dots, l_n) \in L$ denote

$$M_{l_1,\dots,l_n} = \sup\left\{\|\boldsymbol{D}\boldsymbol{\phi}^{-1}(\boldsymbol{x}_{l_1,\dots,l_n}) \circ \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})\|_{\infty,\infty} \;\middle|\; \boldsymbol{x} \in R_{l_1,\dots,l_n}\right\},$$

and note that our choice of δ, along with (1.39), ensures that

$$1 + \epsilon' \leq M^n_{l_1,\dots,l_n}. \tag{1.41}$$

Then

$$\mathrm{vol}(\boldsymbol{\phi}(R_{l_1,\dots,l_n})) \leq |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x}_{l_1,\dots,l_n})| M^n_{l_1,\dots,l_n} \mathrm{vol}(R_{l_1,\dots,l_n})$$

by (1.37).

Then, using (1.38), (1.40), and (1.41),

$$\begin{aligned}\operatorname{vol}(\boldsymbol{\phi}(R)) &= \sum_{(l_1,\dots,l_n)\in L} \operatorname{vol}(\boldsymbol{\phi}(R_{l_1,\dots,l_n}))\\ &\le \sum_{(l_1,\dots,l_n)\in L} |\det \boldsymbol{D\phi}(\boldsymbol{x}_{l_1,\dots,l_n})| M^n_{l_1,\dots,l_n} \operatorname{vol}(R_{l_1,\dots,l_n})\\ &\le (1+\epsilon') \sum_{(l_1,\dots,l_n)\in L} |\det \boldsymbol{D\phi}(\boldsymbol{x}_{l_1,\dots,l_n})| \operatorname{vol}(R_{l_1,\dots,l_n})\\ &\le (1+\epsilon')\Big(\int_R |\det \boldsymbol{D\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} + \epsilon'\Big)\\ &< \int_R |\det \boldsymbol{D\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} + \epsilon\end{aligned}$$

Since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary, this gives the lemma. ▼

6 Lemma *Let* $\mathrm{U} \subseteq \mathbb{R}^n$, $\mathrm{R} \subseteq \mathrm{U}$, *and* $\boldsymbol{\phi}\colon \mathrm{U} \to \mathbb{R}^n$ *have the following properties:*

(i) U *is open;*

(ii) R *is a compact rectangle;*

(iii) $\boldsymbol{\phi}$ *has the following properties:*

  (a) $\boldsymbol{\phi}$ *is continuously differentiable;*

  (b) $\boldsymbol{\phi}|\mathrm{U}$ *is injective;*

  (c) $\boldsymbol{\phi}^{-1}\colon \boldsymbol{\phi}(\mathrm{U}) \to \mathrm{U}$ *is continuously differentiable.*

Then

$$\int_{\mathrm{R}} |\det \mathbf{D}\boldsymbol{\phi}(\mathbf{x})|\,\mathrm{d}\mathbf{x} \le \operatorname{vol}(\boldsymbol{\phi}(\mathrm{R})).$$

Proof For $\epsilon \in \mathbb{R}_{>0}$ let $\epsilon' \in \mathbb{R}_{>0}$ be sufficiently small that

$$(1-\epsilon')^{-n}\operatorname{vol}(\boldsymbol{\phi}(R)) + \epsilon' = \operatorname{vol}(\boldsymbol{\phi}(R)) + \epsilon, \tag{1.42}$$

This being possible since the expression on the left is a continuous function of ϵ' in a neighbourhood of 0 and has the value $\operatorname{vol}(\boldsymbol{\phi}(R))$ when $\epsilon' = 0$.

Now let $\delta_1 \in \mathbb{R}_{>0}$ be sufficiently small that, if $\boldsymbol{P}$ is a partition of R with $|\boldsymbol{P}| < \delta_1$, then

$$\|\boldsymbol{D\phi}(\boldsymbol{x}) - \boldsymbol{D\phi}(\boldsymbol{x}')\|_{\infty,\infty} < \epsilon'$$

whenever $\boldsymbol{x}$ and $\boldsymbol{x}'$ lie in the same subrectangle of $\boldsymbol{P}$; this is possible since $\boldsymbol{D\phi}|R$ is uniformly continuous, being a continuous function on a compact set. Let $\boldsymbol{P}$ be a partition such that $|\boldsymbol{P}| < \delta_1$, write $\boldsymbol{P} = (P_1, \dots, P_n)$ with $P_j = (I_{j1}, \dots, I_{jk_j})$, and let

$$(\boldsymbol{\xi}_{l_1,\dots,l_n} \mid \ l_j \in \{1,\dots,k_j\},\ j \in \{1,\dots,n\})$$

be the selection from $\boldsymbol{P}$ having the property that $\boldsymbol{\xi}_{l_1,\dots,l_n}$ is the centre of $R_{l_1,\dots,l_n}$. Note that the map

$$\boldsymbol{x} \mapsto \boldsymbol{D\phi}^{-1}(\boldsymbol{\xi}_{l_1,\dots,l_n}) \circ \boldsymbol{\phi}(\boldsymbol{x})$$

has the property that its derivative at $\xi_{l_1,\dots,l_n}$ is $\boldsymbol{I}_n$. By Lemma 2 we thus have

$$|\det \boldsymbol{D}\boldsymbol{\phi}^{-1}(\xi_{l_1,\dots,l_n})|\mathrm{vol}(\boldsymbol{\phi}(R_{l_1,\dots,l_n})) \geq (1-\epsilon')^n \mathrm{vol}(R_{l_1,\dots,l_n}) \tag{1.43}$$

for every $l_j \in \{1,\dots,k_j\}$, $j \in \{1,\dots,n\}$.

Now let $\delta_2 \in \mathbb{R}_{>0}$ have the property that if $\boldsymbol{P}$ is a partition with $|\boldsymbol{P}| < \delta_2$, if $\boldsymbol{P} = (P_1,\dots,P_n)$ with $P_j = (I_{j1},\dots,I_{jk_j})$, and if

$$(\xi_{l_1,\dots,l_n} \mid \ l_j \in \{1,\dots,k_j\},\ j \in \{1,\dots,n\})$$

is a selection from $\boldsymbol{P}$, then

$$\Big|\sum_{l_1=1}^{k_1}\cdots\sum_{l_n=1}^{k_n}|\det \boldsymbol{D}\boldsymbol{\phi}(\xi_{l_1,\dots,l_n})|\mathrm{vol}(R_{l_1,\dots,l_n}) - \int_R |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}\Big| < \epsilon'.$$

This implies that

$$\int_R |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} < \sum_{l_1=1}^{k_1}\cdots\sum_{l_n=1}^{k_n}|\det \boldsymbol{D}\boldsymbol{\phi}(\xi_{l_1,\dots,l_n})|\mathrm{vol}(R_{l_1,\dots,l_n}) + \epsilon'. \tag{1.44}$$

Now let $\delta = \min\{\delta_1,\delta_2\}$ and let $\boldsymbol{P}$ be a partition such that $|\boldsymbol{P}| < \delta$, write $\boldsymbol{P} = (P_1,\dots,P_n)$ with $P_j = (I_{j1},\dots,I_{jk_j})$, and let

$$(\xi_{l_1,\dots,l_n} \mid \ l_j \in \{1,\dots,k_j\},\ j \in \{1,\dots,n\})$$

be the selection from $\boldsymbol{P}$ having the property that $\xi_{l_1,\dots,l_n}$ is the centre of $R_{l_1,\dots,l_n}$. Now, using (1.42), (1.43), and (1.44) we compute

$$\begin{aligned}\int_R |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} &< \sum_{l_1=1}^{k_1}\cdots\sum_{l_n=1}^{k_n}|\det \boldsymbol{D}\boldsymbol{\phi}(\xi_{l_1,\dots,l_n})|\mathrm{vol}(R_{l_1,\dots,l_n}) + \epsilon' \\ &\leq (1-\epsilon')^{-n}\sum_{l_1=1}^{k_1}\cdots\sum_{l_n=1}^{k_n}\mathrm{vol}(\boldsymbol{\phi}(R_{l_1,\dots,l_n})) + \epsilon' \\ &\leq (1-\epsilon')^{-n}\mathrm{vol}(\boldsymbol{\phi}(R)) + \epsilon' = \mathrm{vol}(\boldsymbol{\phi}(R)) + \epsilon,\end{aligned}$$

and the result then follows since $\epsilon \in \mathbb{R}_{>0}$ is arbitrary. ▼

From the previous two lemmata we have

$$\int_{\boldsymbol{\phi}(R)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \int_R |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}$$

if f is the constant function having the value 1. By linearity of the integral, this case follows.

Case 3

Here we weaken the hypotheses from the previous case only by allowing f to be Riemann integrable:

1. U is open;
2. A is a compact rectangle (now denoted by R for the remainder of this case) with $R \subseteq U$;
3. ϕ has the following properties:
   (a) ϕ is continuously differentiable;
   (b) ϕ is injective;
   (c) $\phi^{-1}: \phi(U) \to U$ is continuously differentiable;
4. f is bounded and Riemann integrable.

   First let us determine that $f \circ \phi$ is Riemann integrable on R.

7 Lemma *Let* $U \subseteq \mathbb{R}^n$, $A \subseteq U$, $\phi: U \to \mathbb{R}^n$, *and* $f: \phi(A) \to \mathbb{R}$ *have the following properties:*

(i) U *is open;*
(ii) A *is Jordan measurable with* $\mathrm{cl}(A) \subseteq U$*;*
(iii) ϕ *has the following properties:*
  (a) ϕ *is continuously differentiable;*
  (b) ϕ *is injective;*
  (c) $\phi^{-1}: \phi(U) \to U$ *is continuously differentiable;*
(iv) f *is bounded and Riemann integrable.*

Then $(f \circ \phi)|A$ *is Riemann integrable.*

Proof From Lemma 4 we know that $\phi(A)$ is Jordan measurable. By asking that $f(y) = 0$ for $y \in \mathrm{bd}(\phi(A))$ (actually, we can have f be arbitrarily defined on $\mathrm{bd}(\phi(A))$ by Proposition 1.6.21) we can without loss of generality suppose that A is closed. Let $D \subseteq \phi(A)$ be the set of points of discontinuity of f. Let

$$M = \sup\{\|D\phi^{-1}(x)\|_{\infty,\infty} \mid x \in A\},$$

noting that M is finite since A is compact and $D\phi^{-1}$ is continuous.

Let $\epsilon \in \mathbb{R}_{>0}$. Since f is bounded and Riemann integrable it follows that D has zero volume by Theorems 1.6.11 and 1.6.16. By Lemma 3 this means that $\mathrm{vol}(\phi^{-1}(D)) = 0$.
what Moreover, $f \circ \phi(x)$ is continuous if $x \notin \phi^{-1}(D)$ by . Thus $(f \circ \phi)|A$ is discontinuous at a set of points whose volume is zero, and so is Riemann integrable by Theorems 1.6.11 and 1.6.16. ▼

Since R is bounded and Jordan measurable and since $|\det D\phi|$ is continuous on R, it follows that $|\det D\phi||R$ is Riemann integrable. Thus $(f \circ \phi)|\det D\phi|$ is Riemann integrable on R, being a product of Riemann integrable functions.

Let $\epsilon \in \mathbb{R}_{>0}$. For P_ϵ a partition of R such that

$$\int_R f \circ \phi(x)\,\mathrm{d}x - A_-(f \circ \phi, P_\epsilon) < \frac{\epsilon}{2M}, \quad A_+(f \circ \phi, P_\epsilon) - \int_R f \circ \phi(x)\,\mathrm{d}x < \frac{\epsilon}{2M},$$

denote by $\mathscr{R}(\boldsymbol{P}_\epsilon)$ the subrectangles of $\boldsymbol{P}_\epsilon$. For $R' \in \mathscr{R}(\boldsymbol{P}_\epsilon)$ denote

$$\begin{aligned} m_{R'}(f) &= \inf\{f \circ \boldsymbol{\phi}(\boldsymbol{x}) \mid \boldsymbol{x} \in \operatorname{cl}(R')\} = \inf\{f(\boldsymbol{y}) \mid \boldsymbol{y} \in \boldsymbol{\phi}(\operatorname{cl}(R'))\}, \\ M_{R'}(f) &= \sup\{f \circ \boldsymbol{\phi}(\boldsymbol{x}) \mid \boldsymbol{x} \in \operatorname{cl}(R')\} = \sup\{f(\boldsymbol{y}) \mid \boldsymbol{y} \in \boldsymbol{\phi}(\operatorname{cl}(R'))\}. \end{aligned}$$

Note that

$$\sum_{R' \in \mathscr{R}(\boldsymbol{P}_\epsilon)} m_{R'}(f)\operatorname{vol}(\boldsymbol{\phi}(R')) \le \int_{\boldsymbol{\phi}(R)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \le \sum_{R' \in \mathscr{R}(\boldsymbol{P}_\epsilon)} M_{R'}(f)\operatorname{vol}(\boldsymbol{\phi}(R')).$$

Since the theorem holds for Case 2 this is equivalent to

$$\begin{aligned} \sum_{R' \in \mathscr{R}(\boldsymbol{P}_\epsilon)} \int_{R'} m_{R'}(f)|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} &\le \int_{\boldsymbol{\phi}(R)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \\ &\le \sum_{R' \in \mathscr{R}(\boldsymbol{P}_\epsilon)} \int_{R'} M_{R'}(f)|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}. \end{aligned}$$

We also clearly have

$$\begin{aligned} \sum_{R' \in \mathscr{R}(\boldsymbol{P}_\epsilon)} \int_{R'} m_{R'}(f)|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} &\le \int_{R} f \circ \boldsymbol{\phi}(\boldsymbol{x})|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} \\ &\le \sum_{R' \in \mathscr{R}(\boldsymbol{P}_\epsilon)} \int_{R'} M_{R'}(f)|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}. \end{aligned}$$

The definitions of the upper and lower step functions then give

$$\begin{aligned} \Big| \sum_{R' \in \mathscr{R}(\boldsymbol{P}_\epsilon)} \int_{R'} M_{R'}(f)|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} - \sum_{R' \in \mathscr{R}(\boldsymbol{P}_\epsilon)} \int_{R'} m_{R'}(f)|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} \Big| \\ \le M(A_+(f \circ \boldsymbol{\phi}, \boldsymbol{P}_\epsilon) - A_-(f \circ \boldsymbol{\phi}, \boldsymbol{P}_\epsilon)) < \epsilon. \end{aligned}$$

It then follows that

$$\Big| \int_R f \circ \boldsymbol{\phi}(\boldsymbol{x})|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} - \int_{\boldsymbol{\phi}(R)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} \Big| < \epsilon,$$

which is the theorem in this case.

Case 4

Next we weaken the hypotheses by allowing A to be an arbitrary bounded, Jordan measurable subset of U:

1. U is open;
2. A is bounded and Jordan measurable with $\operatorname{cl}(A) \subseteq U$;
3. $\boldsymbol{\phi}$ has the following properties:
   (a) $\boldsymbol{\phi}$ is continuously differentiable;

(b) $\phi|\operatorname{int}(\operatorname{cl}(A))$ is injective;

(c) $(\phi|\operatorname{int}(\operatorname{cl}(A)))^{-1}|\phi(\operatorname{int}(\operatorname{cl}(A)))$ is continuously differentiable;

4. f is bounded and Riemann integrable.

Since A is bounded and Jordan measurable it follows that vol(bd(A)) we can assume without loss of generality, by Proposition 1.6.21, that A is closed. Let us denote

$$\begin{aligned} M_1 &= \sup\{\|\boldsymbol{D}\phi(\boldsymbol{x})\|_{\infty,\infty} \mid \boldsymbol{x} \in A\}, \\ M_2 &= \sup\{|f(\boldsymbol{y})| \mid \boldsymbol{y} \in \phi(A)\}, \\ M_3 &= \sup\{|f \circ \phi(\boldsymbol{x})||\det \boldsymbol{D}\phi(\boldsymbol{x})| \mid \boldsymbol{x} \in A\}. \end{aligned}$$

Let us first show that $(f \circ \phi)|A$ is Riemann integrable. We observe that this does not follow from Lemma 7 since we do not ask that ϕ be invertible on U, only on $\operatorname{int}(A)$.

8 Lemma *Let* $U \subseteq \mathbb{R}^n$, $A \subseteq U$, $\phi\colon U \to \mathbb{R}^n$, *and* $f\colon \phi(A) \to \mathbb{R}$ *have the following properties:*

(i) U *is open;*

(ii) A *is bounded and Jordan measurable with* $\operatorname{cl}(A) \subseteq U$*;*

(iii) ϕ *has the following properties:*

(a) ϕ *is continuously differentiable;*

(b) $\phi|\operatorname{int}(\operatorname{cl}(A))$ *is injective;*

(c) $(\phi|\operatorname{int}(\operatorname{cl}(A)))^{-1}|\phi(\operatorname{int}(\operatorname{cl}(A)))$ *is continuously differentiable;*

(iv) f *is bounded and Riemann integrable.*

Then $(f \circ \phi)|A$ *is Riemann integrable.*

Proof Let $\epsilon \in \mathbb{R}_{>0}$. Since bd($A$) has volume zero by Theorems 1.6.11, 1.6.13, and 1.6.16 let $R_1, \dots, R_k$ be rectangles that cover bd(A) and whose volumes sum to less than $\frac{\epsilon}{2}$. We can assume, without loss of generality (by "fattening the rectangles a little if necessary) that $\operatorname{bd}(A) \subseteq \operatorname{int}(\cup_{j=1}^k R_j)$. Let $B = A \cap (\cup_{j=1}^k R_k)$ and let $C = A \setminus B$. Let $D \subseteq \phi(A)$ be the subset of points at which f is discontinuous, noting that $\operatorname{vol}(D) = 0$ by Theorems 1.6.11 and 1.6.16. Let us write $\phi_B = \phi|B$ and $\phi_C = \phi|C$ so that $\phi^{-1}(D) = \phi_B^{-1}(D) \cup \phi_C^{-1}(D)$ is the set of points of discontinuity of $f \circ \phi$. Since $\phi_B^{-1}(D) = \phi^{-1}(D) \cap B$ and since $B \subseteq \cup_{j=1}^k R_k$, it follows that $\phi_B^{-1}(D)$ is covered by the rectangles $R_1, \dots, R_k$ whose volumes sum to less than $\frac{\epsilon}{2}$. Now note that

$$M = \sup\{\|\boldsymbol{D}\phi^{-1}(\boldsymbol{y})\|_{\infty,\infty} \mid \boldsymbol{y} \in \phi(C)\}$$

is bounded since $\operatorname{bd}(A) \subseteq \operatorname{int}(\cup_{j=1}^k R_j)$. Since $\operatorname{vol}(D) = 0$ it follows from Lemma 3 that $\operatorname{vol}(\phi_B^{-1}(D)) = 0$ and so is covered by finitely many rectangles whose volumes sum to less than $\frac{\epsilon}{2}$. Thus $\phi^{-1}(D)$ is covered by rectangles whose volumes sum to less than ϵ. Thus $(f \circ \phi)|A$ is discontinuous on a set with zero volume, and so it Riemann integrable by Theorems 1.6.11 and 1.6.16. ▼

Let $\epsilon \in \mathbb{R}_{>0}$. Since $\operatorname{vol}(\operatorname{bd}(A)) = 0$ it follows that there exists rectangles $R_1, \dots, R_k$ covering bd(A) such that

$$\sum_{j=1}^k \operatorname{vol}(R_j) < \min\Big\{\frac{\epsilon}{2M_1^n M_2}, \frac{\epsilon}{2M_3}\Big\}.$$

As we argued in the proof of Lemma 3 we can suppose that $\mathrm{cl}(R_j) \subseteq U$, $j \in \{1, \dots, k\}$. Let $B_\epsilon = A \cap (\cup_{j=1}^k R_j)$ and $C_\epsilon = A \setminus B_\epsilon$ so that $\mathrm{vol}(B_\epsilon) < \min\{\frac{\epsilon}{2M_1^n M_2}, \frac{\epsilon}{2M_3}\}$. (We observe that B_ϵ and C_ϵ are Jordan measurable since they are formed by finite unions, finite intersections, and set theoretic differences of Jordan measurable sets, cf. Section III-2.1.1.) By Lemma 1 we have $\mathrm{vol}(\boldsymbol{\phi}(B_\epsilon)) \le M_1^n \mathrm{vol}(B_\epsilon)$. Therefore,

$$\left|\int_{\boldsymbol{\phi}(A)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} - \int_{\boldsymbol{\phi}(C_\epsilon)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right| = \left|\int_{\boldsymbol{\phi}(B_\epsilon)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right| \le M_2 \mathrm{vol}(\boldsymbol{\phi}(B_\epsilon)) < \frac{\epsilon}{2}.$$

We also have

$$\begin{aligned}\left|\int_A f\circ\boldsymbol{\phi}(\boldsymbol{x})|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} - \int_{C_\epsilon} f\circ\boldsymbol{\phi}(\boldsymbol{x})|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}\right| \\ = \left|\int_{B_\epsilon} f\circ\boldsymbol{\phi}(\boldsymbol{x})|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}\right| \le M_3 \mathrm{vol}(B_\epsilon) < \tfrac{\epsilon}{2}.\end{aligned}$$

Using the fact that the theorem holds for Case 3 we then compute

$$\begin{aligned}&\left|\int_A f\circ\boldsymbol{\phi}(\boldsymbol{x})|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} - \int_{\boldsymbol{\phi}(A)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right| \\ &\quad\le \left|\int_A f\circ\boldsymbol{\phi}(\boldsymbol{x})|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} - \int_{C_\epsilon} f\circ\boldsymbol{\phi}(\boldsymbol{x})|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x}\right| \\ &\quad\quad + \left|\int_{C_\epsilon} f\circ\boldsymbol{\phi}(\boldsymbol{x})|\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} - \int_{\boldsymbol{\phi}(C_\epsilon)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right| \\ &\quad\quad + \left|\int_{\boldsymbol{\phi}(A)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} - \int_{\boldsymbol{\phi}(C_\epsilon)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}\right| \\ &\quad< \tfrac{\epsilon}{2} + 0 + \tfrac{\epsilon}{2} = \epsilon,\end{aligned}$$

which gives the theorem in this case.

Case 5

Next we allow A to be possibly unbounded:

1. U is open;
2. A is Jordan measurable with $\mathrm{cl}(A) \subseteq U$;
3. $\boldsymbol{\phi}$ is has the following properties:
   (a) $\boldsymbol{\phi}$ is continuously differentiable;
   (b) $\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(A))$ is injective;
   (c) $(\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(A)))^{-1}|\boldsymbol{\phi}(\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(A)))$ is continuously differentiable;
4. f is locally bounded, Riemann integrable, and takes values in $\mathbb{R}_{\ge 0}$. defined?

As in the preceding case, we can without loss of generality assume that A is closed. This is not perfectly obvious from the beginning, but it can easily be seen as we go through the proof that this assumption can indeed be made without loss of generality.

Let $r \in \mathbb{R}_{>0}$ and let

$$R_r = [-r, r] \times \cdots \times [-r, r].$$

Denote $A_r = A \cap R_r$ and define

$$\alpha_r = \sup\{\|\boldsymbol{D\phi}(\boldsymbol{x})\|_{\infty,\infty} \mid \boldsymbol{x} \in A_r\}$$

and take $M(r) = \max\{r, \alpha_r\}$. By Lemma 1 (more precisely, from the proof of that lemma) it follows that $\boldsymbol{\phi}(A_r) \subseteq R_{M(r)}$. Since A is Jordan measurable, A_r is bounded and Jordan measurable. Therefore, by Lemma 4 it follows that $\boldsymbol{\phi}(A_r)$ is bounded and Jordan measurable. Define $f_r\colon R_{M(r)} \to \mathbb{R}_{\geq 0}$ by

$$f_r(\boldsymbol{y}) = \begin{cases} f(\boldsymbol{y}), & \boldsymbol{y} \in \boldsymbol{\phi}(A_r), \\ 0, & \text{otherwise.} \end{cases}$$

Since f is locally bounded and Riemann integrable and since $\boldsymbol{\phi}(A_r)$ is bounded and Jordan measurable, it follows that f_r is Riemann integrable. Moreover, since $\lim_{r\to\infty} M(r) = \infty$ it follows that the limit

$$\lim_{r\to\infty} \int_{R_{M(r)}} f_r(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}$$

exists by Riemann integrability of f.

Now, the hypotheses of Case 4 apply to U, A_r, $\boldsymbol{\phi}$, and f_r. Therefore,

$$\int_{A_r} f_r \circ \boldsymbol{\phi}(\boldsymbol{x}) |\det \boldsymbol{D\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{\phi}(A_r)} f_r(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}.$$

Therefore,

$$\lim_{r\to\infty} \int_{A_r} f_r \circ \boldsymbol{\phi}(\boldsymbol{x}) |\det \boldsymbol{D\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} = \lim_{r\to\infty} \int_{\boldsymbol{\phi}(A_r)} f_r(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = \lim_{r\to\infty} \int_{R_{M(r)}} f_r(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}$$

and the first limit is thus equal to the last limit. But this means exactly that

$$\int_A f \circ \boldsymbol{\phi}(\boldsymbol{x}) |\det \boldsymbol{D\phi}(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{\phi}(A)} f(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y},$$

by the definition of the integral (i.e., Definition 1.6.22) in each case. This gives the theorem in this case.

Case 6

Next we remove the restriction that f be locally bounded:

1. U is open;
2. A is Jordan measurable with $\mathrm{cl}(A) \subseteq U$;
3. $\boldsymbol{\phi}$ is has the following properties:
   (a) $\boldsymbol{\phi}$ is continuously differentiable;
   (b) $\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(A))$ is injective;
   (c) $(\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(A)))^{-1}|\boldsymbol{\phi}(\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(A)))$ is continuously differentiable;

4. f is Riemann integrable and takes values in $\mathbb{R}_{\geq 0}$.

For $M \in \mathbb{R}_{>0}$ define $f_M\colon A \to \mathbb{R}_{\geq 0}$ and $(f \circ \boldsymbol{\phi})_M\colon A \to \mathbb{R}_{\geq 0}$ just as in Definition 1.6.22. Note that $(f \circ \boldsymbol{\phi})_M = f_M \circ \boldsymbol{\phi}$. Since f_M is bounded and Riemann integrable it follows from Case 5 that

$$\int_A f_M \circ \boldsymbol{\phi}(\boldsymbol{x}) |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})| \,\mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{\phi}(A)} f_M(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y}.$$

Thus we have

$$\lim_{M\to\infty} \int_A (f \circ \boldsymbol{\phi})_M(\boldsymbol{x}) |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})| \,\mathrm{d}\boldsymbol{x} = \lim_{M\to\infty} \int_{\boldsymbol{\phi}(A)} f_M(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y}$$

which gives

$$\int_A f \circ \boldsymbol{\phi}(\boldsymbol{x}) |\det \boldsymbol{D}\boldsymbol{\phi}(\boldsymbol{x})| \,\mathrm{d}\boldsymbol{x} = \int_{\boldsymbol{\phi}(A)} f(\boldsymbol{y}) \,\mathrm{d}\boldsymbol{y}$$

by the definition (i.e., Definition Definition 1.6.22) of the integral on each side of this equality.

Case 7

Here we eliminate the requirement that f be nonnegative:

1. U is open;
2. A is Jordan measurable with $\mathrm{cl}(A) \subseteq U$;
3. $\boldsymbol{\phi}$ is has the following properties:
   (a) $\boldsymbol{\phi}$ is continuously differentiable;
   (b) $\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(A))$ is injective;
   (c) $(\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(A)))^{-1}|\boldsymbol{\phi}(\boldsymbol{\phi}|\,\mathrm{int}(\mathrm{cl}(A)))$ is continuously differentiable;
4. f is Riemann integrable.

This follows since the positive/negative decomposition has the property that $(f \circ \boldsymbol{\phi})_+(\boldsymbol{x}) = f_+ \circ \boldsymbol{\phi}(\boldsymbol{x})$ and $(f \circ \boldsymbol{\phi})_-(\boldsymbol{x}) = f_- \circ \boldsymbol{\phi}(\boldsymbol{x})$. Thus the preceding case can be applied separately to the positive and negative parts of f. ■

Let us give some examples to illustrate the theorem.

1.6.40 Examples (Change of variables formula)

1. Let us take $U = \mathbb{R}$, $A = \mathbb{R}$, and define $\phi\colon U \to \mathbb{R}$ and $f\colon (-\frac{\pi}{2}, \frac{\pi}{2}) \to \mathbb{R}$ by $\phi(x) = \tan^{-1}(x)$ and $f(y) = 1$, respectively. We claim that $\boldsymbol{D}\phi(x) = \frac{1}{1+x^2}$. To see this write $\tan(x) = \frac{\sin(x)}{\cos(x)}$, compute

$$\tan'(x) = \frac{\sin'(x)\cos(x) - \sin(x)\cos'(x)}{\cos(x)^2} = \frac{1}{\cos(x)^2},$$

using the quotient rule. By the Inverse Function Theorem,

$$\phi'(x) = \frac{1}{(\phi^{-1})'(\phi(x))} = \frac{1}{\tan'(\tan^{-1}(x))} = (\cos(\tan^{-1}(x)))^2.$$

If $\theta = \tan^{-1}(x)$ then

$$\frac{\sin(\theta)}{\cos(\theta)} = x.$$

Since $\sin(\theta) > 0$ for $\theta \in (0, \pi)$, we have $\sin(\theta) = \sqrt{1 - \cos(\theta)}$ by (I-3.26). Therefore,

$$\frac{1-\cos(\theta)^2}{\cos(\theta)^2} = x^2 \quad \implies \quad \cos(\theta)^2 = \frac{1}{1+x^2}$$

as desired. Therefore,

$$\int_A f \circ \phi(x) |\det \boldsymbol{D}\phi(x)| \, \mathrm{d}x = \int_{\mathbb{R}} \frac{1}{1+x^2} \, \mathrm{d}x, \quad \int_{\phi(A)} f(y) \, \mathrm{d}y = \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \mathrm{d}y.$$

The change of variables formula then gives

$$\int_{\mathbb{R}} \frac{1}{1+x^2} \, \mathrm{d}x = \pi.$$

2. Let us consider $U = \mathbb{R}$ and for $k \in \mathbb{Z}$ define

$$A_k = [-(k - \tfrac{1}{2})\pi, (k - \tfrac{1}{2})\pi].$$

Then take $\phi \colon U \to \mathbb{R}$ and $f \colon [-1, 1] \to \mathbb{R}$ to be defined by $\phi(x) = \sin(x)$ and $f(y) = 1$, respectively. Let us compare the change of variables formula of Proposition I-3.4.27 with that of Theorem 1.6.39.

Note that the hypotheses of Proposition I-3.4.27 are satisfied for every $k \in \mathbb{Z}_{>0}$. We have

$$\int_a^b f \circ \phi(x) \phi'(x) \, \mathrm{d}x = \int_{-(k-\frac{1}{2})\pi}^{(k-\frac{1}{2})\pi} \cos(x) \, \mathrm{d}x,$$

$$\int_{\phi(a)}^{\phi(b)} \mathrm{d}y = \int_{(-1)^k}^{(-1)^{k+1}} \mathrm{d}y.$$

Thus the conclusions of Proposition I-3.4.27 give

$$\int_{-(k-\frac{1}{2})\pi}^{(k-\frac{1}{2})\pi} \sin(x)\cos(x) \, \mathrm{d}x = 2(-1)^{k+1}.$$

The hypotheses of Theorem 1.6.39 hold only when $k = 1$. Indeed, when $k > 1$ we do not have injectivity of ϕ on $\mathrm{int}(A_k)$. Moreover, we have

$$\int_{A_k} f \circ \phi(x) |\det \boldsymbol{D}\phi(x)| \, \mathrm{d}x = \int_{-(k-\frac{1}{2})\pi}^{(k-\frac{1}{2})\pi} |\cos(x)| \, \mathrm{d}x = 4k - 2,$$

$$\int_{\phi(A_k)} f(y) \, \mathrm{d}y = \int_{-1}^{1} \mathrm{d}y = 2.$$

Thus we see that the conclusions of Theorem 1.6.39 hold only when $k = 1$, just in line with when the hypotheses hold.

The point is that the two change of variables formulae, Proposition I-3.4.27 and Theorem 1.6.39, are not the same, and one needs to exercise some care in using them. Moreover, the example shows that the hypotheses that the injectivity of $\phi|\,\mathrm{int}(A)$ is necessary; this fact is overlooked in some treatments of the change of variable formula. The final resolution of all of this will not be achieved until we consider the change of variable formula for the Lebesgue integral in . what?

3. We next consider a well-known situation, that of changing from Cartesian coordinates (x, y) for $\mathbb{R}^2$ to polar coordinates (r, θ). To set this up we take $U = \mathbb{R}^2$,

$$A = [0, \infty) \times [-\pi, \pi]$$

and define $\boldsymbol{\phi}\colon U \to \mathbb{R}^2$ by $\phi(r, \theta) = (r\cos(\theta), r\sin(\theta))$. We compute

$$\boldsymbol{D\phi}(r, \theta) = \begin{bmatrix} \cos(\theta) & -r\sin(\theta) \\ \sin(\theta) & r\cos(\theta) \end{bmatrix},$$

and so $\det \boldsymbol{D\phi}(r, \theta) = r$. Note that

$$\boldsymbol{\phi}(\mathrm{int}(A)) = \mathbb{R}^2 \setminus \{(x, y) \mid \ y = 0,\ x \le 0\}.$$

One can check that $\boldsymbol{\phi}|\,\mathrm{int}(A)$ is injective and that

$$(\boldsymbol{\phi}|\,\mathrm{int}(A))^{-1}(x, y) = (\sqrt{x^2 + y^2}, \mathrm{atan}(x, y)), \qquad (x, y) \in \boldsymbol{\phi}(\mathrm{int}(A)).$$

This function is continuously differentiable on $\boldsymbol{\phi}(\mathrm{int}(A))$. Thus, for any Riemann integrable function $f\colon \boldsymbol{\phi}(\mathrm{int}(A)) \to \mathbb{R}$ we have

$$\int_A f(r\cos(\theta), r\sin(\theta)) r\,\mathrm{d}r\mathrm{d}\theta = \int_{\boldsymbol{\phi}(A)} f(x, y)\,\mathrm{d}x\mathrm{d}y,$$

which is the usual change of variable formula for this case.

4. Let us give a concrete illustration of the preceding polar coordinate formula. We wish to compute

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp(-x^2 - y^2)\,\mathrm{d}x\mathrm{d}y.$$

Using the change of variables formula for polar coordinate derived above we have

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \exp(-x^2 - y^2)\,\mathrm{d}x\mathrm{d}y = \int_{\mathbb{R}_{\ge 0}} \int_{[-\pi, \pi]} r\exp(-r^2)\,\mathrm{d}\theta\mathrm{d}r.$$

We obviously have

$$\int_{[-\pi, \pi]} r\exp(-r^2)\,\mathrm{d}\theta = 2\pi r\exp(-r^2),$$

and so by Fubini's Theorem we have

$$\int_{\mathbb{R}}\int_{\mathbb{R}} \exp(-x^2 - y^2)\,\mathrm{d}x\mathrm{d}y = 2\pi \int_{\mathbb{R}_{\geq 0}} r\exp(-r^2)\,\mathrm{d}r.$$

To evaluate the integral on the right we use the change of variables formula again with $U = \mathbb{R}$, $A = \mathbb{R}_{\geq 0}$, $\phi(r) = r^2$, and $f(y) = \exp(-y)$. The change of variables formula then gives

$$2\int_{\mathbb{R}_{\geq 0}} r\exp(-r^2)\,\mathrm{d}r = \int_{\mathbb{R}_{\geq 0}} \exp(-y)\,\mathrm{d}y = 1$$

or

$$\int_{\mathbb{R}_{\geq 0}} r\exp(-r^2)\,\mathrm{d}r = \frac{1}{2}.$$

This gives

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(-x^2 - y^2)\,\mathrm{d}x\mathrm{d}y = \pi.$$ •

1.6.9 Notes

Exercises

1.6.1 Let $A \subseteq \mathbb{R}^n$ be bounded, let $f\colon A \to \mathbb{R}$ be bounded, and suppose that, for each $\epsilon \in \mathbb{R}_{>0}$, there exists Riemann integrable functions $\overline{f}_\epsilon, \underline{f}_\epsilon\colon A \to \mathbb{R}$ such that

$$\underline{f}_\epsilon(\boldsymbol{x}) \leq f(\boldsymbol{x}) \leq \overline{f}_\epsilon(\boldsymbol{x}), \qquad \boldsymbol{x} \in A$$

and

$$\int_A \overline{f}_\epsilon(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_A \underline{f}_\epsilon(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} < \epsilon.$$

Show that f is Riemann integrable and that

$$I(f) = \lim_{\epsilon\to 0}\int_A \overline{f}_\epsilon(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \lim_{\epsilon\to 0}\int_A \underline{f}_\epsilon(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.$$

1.6.2 Let $R \subseteq \mathbb{R}^n$ be a fat compact rectangle, let $f\colon R \to \mathbb{R}$ be Riemann integrable (in the sense of Definition 1.6.7), and let $A \subseteq R$. Denote by $f_A\colon R \to \mathbb{R}$ the function

$$f_A(\boldsymbol{x}) = \begin{cases} f(\boldsymbol{x}), & \boldsymbol{x} \in A, \\ 0, & \boldsymbol{x} \notin A. \end{cases}$$

Answer the following questions.

(a) Show that f_A is Riemann integrable if A is Jordan measurable.

(b) Is it true that, if f_A is Riemann integrable, then A is Jordan measurable?

1.6.3 Let $A_1, \dots, A_k \subseteq \mathbb{R}^n$ have volume zero. Show that $A_1 \cup \dots \cup A_k$ has volume zero.

1.6.4 Let $\mathsf{V} \subseteq \mathbb{R}^n$ be a subspace with dimension less than n. Show that $\mathsf{V} \cap K$ has volume zero for any compact subset K of $\mathbb{R}^n$.

1.6.5 Let $A, B, A_1, \dots, A_k \subseteq \mathbb{R}^n$ be bounded and Jordan measurable sets.

(a) Show that $\cup_{j=1}^k A_j$ is bounded and Jordan measurable.

(b) Show that $\cap_{j=1}^k A_j$ is bounded and Jordan measurable.

(c) Show that $A - B$ is bounded and Jordan measurable.

1.6.6 Use Fubini's Theorem to directly show that the disk of radius r in $\mathbb{R}^2$ has area πr^2.

1.6.7 Show that $\lim_{n\to\infty} V_n(r) = 0$ for each $r \in \mathbb{R}_{>0}$, where $V_n(r)$ is the volume of the ball of radius r in $\mathbb{R}^n$.

Interesting fact: Note that since $V_2(r) > V_1(r)$ and since $\lim_{n\to\infty} V_n(r) = 0$ it follows that there exists some n_0 such that $V_{n_0}(r) = \max\{V_n(r) \mid n \in \mathbb{Z}_{>0}\}$. It turns out that this "largest" ball is the ball in $\mathbb{R}^5$.

1.6.8 Let $\boldsymbol{L} \in \operatorname{End}_{\mathbb{R}}(\mathbb{R}^2)$ be a 2×2 elementary matrix and let $R \subseteq \mathbb{R}^2$ be a rectangle. For each of the three possible types of elementary matrices, draw the picture of R and $\boldsymbol{L}(R)$, and verify for yourself that Theorem 1.6.38 holds in each case.

Section 1.7

Sequences and series of functions

In this section we generalise the results of Section I-3.6 to functions defined on subsets of $\mathbb{R}^n$. Much of the discussion will take a similar form to our discussion of functions whose domain is $\mathbb{R}$. However, because of the more general context, we will give some results that are of a more advanced nature.

Do I need to read this section? If a reader is acquainted with the results in Section I-3.6 then this section can be bypassed on a first reading. However, when we come to use the greater generality of functions of multiple variables, the reader will want to refer back to this section to be sure that all of the extensions from the single variable case work as expected. •

1.7.1 Uniform convergence

1.7.1 Theorem (Weierstrass M-test)

1.7.2 The Weierstrass Approximation Theorem

We now give the multivariable version of the Weierstrass Approximation Theorem presented in Section I-3.6.6 for the single-variable case. As we shall see, there are no substantial difficulties with adapting our single-variable proof to the multivariable case. Thus we limit the discussion, and get right to the point.

1.7.2 Definition (Polynomial functions) A function $P\colon \mathbb{R}^n \to \mathbb{R}$ is a polynomial function if

$$P(x_1, \dots, x_n) = \sum_{(k_1,\dots,k_n)\in\mathbb{Z}^n_{\geq 0}} a_{k_1\cdots k_n} x_1^{k_1} \cdots x_n^{k_n},$$

where the set of numbers $a_{k_1\cdots k_n} \in \mathbb{R}$, $(k_1, \dots, k_n) \in \mathbb{Z}^n_{\geq 0}$ have the property that the set

$$\left\{(k_1, \dots, k_n) \in \mathbb{Z}^n_{\geq 0} \;\middle|\; a_{k_1\cdots k_n} \neq 0\right\}$$

is finite. •

In Section I-4.4.7 we discuss multivariable polynomials in a little detail, so the reader may be interested in reading about this material there. However, we shall be interested in only the most pedestrian aspects of such objects. Indeed our interest is in the following polynomials, recalling from Definition I-3.6.19 the notation P_k^m for the single-variable Bernstein polynomials.

1.7.3 Definition (Multivariate Bernstein polynomial, multivariate Bernstein approximation) For $m_1, \dots, m_n \in \mathbb{Z}_{\geq 0}$ and for $k_j \in \{0, 1, \dots, m_j\}$, $j \in \{1, \dots, n\}$, the polynomial function

$$P^{m_1 \cdots m_n}_{k_1 \cdots k_n}(x_1, \dots, x_n) = P^{m_1}_{k_1}(x_1) \cdots P^{m_n}_{k_n}(x_n) = \binom{m_1}{k_1} \cdots \binom{m_n}{k_n} x_1^{k_1}(1 - x_1)^{m_1 - k_1} \cdots x_n^{k_n}(1 - x_n)^{m_n - k_n}$$

is a ***Bernstein polynomial*** in n-variables. For a continuous function $f: R \to \mathbb{R}$ defined on a fat compact rectangle

$$R = [a_1, b_1] \times \cdots \times [a_n, b_n],$$

the **$(\mathbf{m_1}, \dots, \mathbf{m_n})$*th Bernstein approximation*** of f is the function $B^R_{m_1 \cdots m_n} f: R \to \mathbb{R}$ defined by

$$B^R_{m_1 \cdots m_n} f(x_1, \dots, x_n) = \sum_{k_1=0}^{m_1} \cdots \sum_{k_n=0}^{m_n} f\left(\tfrac{k_1}{m_1}, \dots, \tfrac{k_n}{m_n}\right) P^{m_1 \cdots m_n}_{k_1 \cdots k_n}(x_1, \dots, x_n). \qquad \bullet$$

We may now state the multivariable Weierstrass Approximation Theorem.

1.7.4 Theorem (Multivariable Weierstrass Approximation Theorem) *Let* $K \subseteq \mathbb{R}^n$ *be a compact set and let* $f: K \to \mathbb{R}$ *be continuous. Then there exists a sequence* $(P_m)_{m \in \mathbb{Z}_{>0}}$ *of polynomial functions on* $\mathbb{R}^n$ *such that the sequence* $(P_m|K)_{m \in \mathbb{Z}_{>0}}$ *converges uniformly to* f.

Proof First let us consider the case when $K = R$ is a fat compact rectangle. We can without loss of generality take the case when $R = [0, 1]^n$, and for brevity denote $B_{m_1 \cdots m_n} f = B^R_{m_1 \cdots m_n} f$. We will show that the sequence $(B_{m_1 \cdots m_n} f)_{(m_1, \dots, m_n) \in \mathbb{Z}_{\geq 0}}$ converges uniformly to f on R. That is to say, given $\epsilon \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{Z}_{>0}$ such that, whenever $m_1, \dots, m_n \geq N$,

$$\left| B_{m_1 \cdots m_n} f(\boldsymbol{x}) - f(\boldsymbol{x}) \right| < \epsilon, \qquad \boldsymbol{x} \in R.$$

Let $\epsilon \in \mathbb{R}_{>0}$. Since a continuous function on the compact set R is uniformly continuous (Theorem 1.3.33) it follows that there exists $\delta \in \mathbb{R}_{>0}$ such that

$$\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} \leq \delta \quad \Longrightarrow \quad |f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq \tfrac{\epsilon}{2}.$$

Also define

$$M = \sup\{|f(\boldsymbol{x})| \mid \boldsymbol{x} \in R\},$$

noting that this is finite by Theorem 1.3.31. Now, it $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} \leq \delta$ then

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leq \tfrac{\epsilon}{2} \leq \tfrac{\epsilon}{2} + \tfrac{2M}{n\delta^2}(x_j - y_j)^2$$

for every $j \in \{1, \dots, n\}$. If $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} > \delta$ then

$$(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2 > \delta^2.$$

This means that, for some $j_0 \in \{1, \dots, n\}$, $(x_{j_0} - y_{j_0})^2 > \frac{\delta^2}{n}$. Therefore,

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \le 2M \le 2M\Big(\frac{x_{j_0} - y_{j_0}}{\sqrt{n}\delta}\Big)^2 \le \frac{\epsilon}{2} + \frac{2M}{n\delta^2}(x_{j_0} - y_{j_0})^2.$$

Thus, for every $\boldsymbol{x}, \boldsymbol{y} \in R$ we have

$$|f(\boldsymbol{x}) - f(\boldsymbol{y})| \le \frac{\epsilon}{2} + \frac{2M}{n\delta^2}(x_{j_0} - y_{j_0})^2$$

for some $j_0 \in \{1, \dots, n\}$.

Define $f_0 \colon R \to \mathbb{R}$ by $f_0(\boldsymbol{x}) = 1$ and, for $j_0 \in \{1, \dots, n\}$, define $f_{1,j}, f_{2,j} \colon R \to \mathbb{R}$ by

$$f_{1,j}(\boldsymbol{x}) = x_j, \quad f_{2,j}(\boldsymbol{x}) = x_j^2.$$

Using the lemma from the proof of Theorem I-3.6.21 and the Binomial Theorem, one can easily verify the following identities:

$$\begin{aligned} B_{m_1\cdots m_n} f_0(\boldsymbol{x}) &= 1; \\ B_{m_1\cdots m_n} f_{1,j}(\boldsymbol{x}) &= x_j; \\ B_{m_1\cdots m_n} f_{2,j}(\boldsymbol{x}) &= x_j^2 + \tfrac{1}{m_j}(x_j - x_j^2). \end{aligned}$$

In like manner one can also use the lemma of Theorem I-3.6.21 to verify that

$$|B_{m_1\cdots m_n} f(\boldsymbol{x})| \le B_{m_1\cdots m_n} g(\boldsymbol{x}), \qquad \boldsymbol{x} \in R$$

if $|f(\boldsymbol{x})| \le g(\boldsymbol{x})$ for every $\boldsymbol{x} \in R$.

Now fix $\boldsymbol{x}_0 = (x_{0,1}, \dots, x_{0,n}) \in R$. For $\boldsymbol{x} \in R$ let $j(\boldsymbol{x}) \in \{1, \dots, n\}$ be such that

$$|f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| \le \frac{\epsilon}{2} + \frac{2M}{n\delta^2}(x_{j(\boldsymbol{x})} - x_{0,j(\boldsymbol{x})})^2$$

For every $m_1, \dots, m_n \in \mathbb{Z}_{\ge 0}$ we have

$$\begin{aligned} |B_{m_1\cdots m_n} f(\boldsymbol{x}) - f(\boldsymbol{x}_0)| &= |B_{m_1\cdots m_n}(f - f(\boldsymbol{x}_0) f_0)(\boldsymbol{x})| \\ &\le B_{m_1\cdots m_n}\Big(\tfrac{\epsilon}{2} f_0 + \tfrac{2M}{n\delta^2}(f_{1,j(\boldsymbol{x})} - x_{0,j(\boldsymbol{x})} f_0)^2\Big)(\boldsymbol{x}) \\ &= \tfrac{\epsilon}{2} + \tfrac{2M}{n\delta^2}(x_{j(\boldsymbol{x})}^2 + \tfrac{1}{m_{j(\boldsymbol{x})}}(x_{j(\boldsymbol{x})} - x_{j(\boldsymbol{x})}^2) - 2x_{0,j(\boldsymbol{x})} x_{j(\boldsymbol{x})} + x_{0,j(\boldsymbol{x})}^2) \\ &= \tfrac{\epsilon}{2} + \tfrac{2M}{n\delta^2}(x_{j(\boldsymbol{x})} - x_{0,j(\boldsymbol{x})})^2 + \tfrac{2M}{n m_{j(\boldsymbol{x})}\delta^2}(x_{j(\boldsymbol{x})} - x_{j(\boldsymbol{x})}^2). \end{aligned}$$

Now take $\boldsymbol{x} = \boldsymbol{x}_0$, note that $j(\boldsymbol{x}_0)$ can be arbitrary, and then get, for any $j \in \{1, \dots, n\}$,

$$|B_{m_1\cdots m_n} f(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)| \le \frac{\epsilon}{2} + \frac{2M}{n m_j \delta^2}(x_{0,j} - x_{0,j}^2) \le \frac{\epsilon}{2} + \frac{M}{2n m_j \delta^2},$$

using the fact that $x - x^2 \le \frac{1}{4}$ for $x \in [0, 1]$. Therefore, if $N \in \mathbb{Z}_{>0}$ is sufficiently large that $\frac{M}{2nm\delta^2} < \frac{\epsilon}{2}$ for $m \ge N$ we have

$$|B_{m_1\cdots m_n} f(\boldsymbol{x}_0) - f(\boldsymbol{x}_0)| < \epsilon,$$

and this holds for every $\boldsymbol{x}_0 \in R$, giving us the desired uniform convergence in the case where K is a rectangle.

Now consider the case where K is a general compact set and let R be a fat compact rectangle such that $K \subseteq R$. By the Tietze Extension Theorem extend f to a continuous function $\hat{f}\colon R \to \mathbb{R}$ such that $\hat{f}|K = f$. Our computations above ensure that, for $\epsilon \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{Z}_{>0}$ such that, whenever $m_1, \dots, m_n \geq N$,

$$|B_{m_1 \cdots m_n}\hat{f}(\boldsymbol{x}) - \hat{f}(\boldsymbol{x})| < \epsilon, \qquad \boldsymbol{x} \in R.$$

If we $P_m = B_{m \cdots m}\hat{f}$, $m \in \mathbb{Z}_{>0}$, this gives the sequence of polynomial functions converging uniformly to f on K. ■

This can then easily be extended to maps taking values in Euclidean space by applying the preceding theorem to each component. Let us say that a map $\boldsymbol{P}\colon \mathbb{R}^n \to \mathbb{R}^m$ is a ***polynomial map*** if

$$\boldsymbol{P}(x_1, \dots, x_n) = (P_1(x_1, \dots, x_n), \dots, P_m(x_1, \dots, x_n))$$

for polynomial functions $P_j\colon \mathbb{R}^n \to \mathbb{R}$.

1.7.5 Corollary (Weierstrass Approximation Theorem for vector-valued maps) *Let $\mathrm{K} \subseteq \mathbb{R}^\mathrm{n}$ be a compact set and let $\mathbf{f}\colon \mathrm{K} \to \mathbb{R}^\mathrm{m}$ be continuous. Then there exists a sequence $(\mathbf{P}_\mathrm{m})_{\mathrm{m} \in \mathbb{Z}_{>0}}$ of polynomial maps on $\mathbb{R}^\mathrm{n}$, taking values in $\mathbb{R}^\mathrm{m}$, such that the sequence $(\mathbf{P}_\mathrm{m}|\mathrm{K})_{\mathrm{m} \in \mathbb{Z}_{>0}}$ converges uniformly to $\mathbf{f}$.*

1.7.3 Swapping limits with other operations

In this section we prove some of the same results as in Section I-3.6.7 concerning the swapping of limits and other operations, like integration and differentiation. One significant extension we give in this section concerns limit theorems for Riemann integration. In Section I-3.6.7 we showed that for uniformly convergent sequences one can swap limit and integral. However, this is true, even for the Riemann integral in a more general setting. Here we state these results. These results are really best suited to the domain of Lebesgue integration which we discuss in Chapter III-2. However, since some version of these results are valid for the more easily understood Riemann integral, it is interesting to record them. Moreover, by comparing what is true for the Riemann integral with what is true for the more general Lebesgue integral, one can get a better appreciation of the value of the Lebesgue integral.

First we record the commutativity of the Riemann integral with increasing sequences of functions.

1.7.6 Theorem (The Monotone Convergence Theorem for the Riemann integral) *Let $\mathrm{R} \subseteq \mathbb{R}^\mathrm{n}$ be a fat compact rectangle and let $(\mathrm{f_j})_{\mathrm{j} \in \mathbb{Z}_{>0}}$ be a sequence of $\mathbb{R}$-valued functions on R satisfying the following conditions:*

(i) $\mathrm{f_j}(\mathbf{x}) \geq 0$ *for each* $\mathbf{x} \in \mathrm{R}$ *and* $\mathrm{j} \in \mathbb{Z}_{>0}$*;*

(ii) $\mathrm{f_{j+1}}(\mathbf{x}) \geq \mathrm{f_j}(\mathbf{x})$ *for each* $\mathbf{x} \in \mathrm{R}$ *and* $\mathrm{j} \in \mathbb{Z}_{>0}$*;*

(iii) $\mathrm{f_j}$ *is Riemann integrable (in the sense of Definition 1.6.22) for each* $\mathrm{j} \in \mathbb{Z}_{>0}$*;*

(iv) *the map* $f\colon R \to \mathbb{R}_{\geq 0}$ *defined by* $f(x) = \lim_{j\to\infty} f_j(x)$ *exists and is Riemann integrable (in the sense of Definition 1.6.22).*

Then

$$\lim_{j\to\infty} \int_R f_j(x)\,dx = \int_R f(x)\,dx.$$

Proof We first prove a couple of lemmata.

1 Lemma *Let* $R \subseteq \mathbb{R}^n$ *be a fat compact rectangle, let* $f\colon R \to \mathbb{R}$ *be bounded and Riemann integrable with*

$$M = \sup\{|f(x)| \mid x \in R\},$$

and suppose that $\int_R f(x)\,dx \geq m\,\mathrm{vol}(R)$ *for some* $m \in \mathbb{R}_{>0}$. *Then the set*

$$\left\{x \in R \;\middle|\; f(x) \geq \tfrac{m}{2}\mathrm{vol}(R)\right\}$$

contains a finite union of rectangles whose total volume is bounded below by $\frac{m}{4M}\mathrm{vol}(R)$.

Proof Let $\boldsymbol{P}$ be a partition of R for which

$$0 \leq \int_R f(x)\,dx - A_-(f, \boldsymbol{P}) \leq \frac{m}{4}\mathrm{vol}(R).$$

Therefore $A_-(f, \boldsymbol{P}) \geq \frac{3m}{4}\mathrm{vol}(R)$. Let us write $\boldsymbol{P} = (P_1, \dots, P_n)$ with $P_j = (I_{j1}, \dots, I_{jk_j})$, $j \in \{1, \dots, n\}$. Let

$$E = \left\{x \in R \;\middle|\; f(x) \geq \tfrac{m}{2}\right\}$$

and denote

$$L_1 = \{(l_1, \dots, l_n) \in \{1, \dots, k_1\} \times \cdots \times \{1, \dots, k_n\} \mid R_{l_1,\dots,l_m} \subseteq E\}$$

and

$$L_2 = (\{1, \dots, k_1\} \times \cdots \times \{1, \dots, k_n\}) \setminus L_1.$$

We then have

$$\begin{aligned}
\frac{3m}{4}\mathrm{vol}(R) \leq A_-(f, \boldsymbol{P}) &= \sum_{(l_1,\dots,l_n)\in L_1} \inf\{f(x) \mid x \in \mathrm{cl}(R_{l_1,\dots,l_n})\}\mathrm{vol}(R_{l_1,\dots,l_n}) \\
&\quad + \sum_{(l_1,\dots,l_n)\in L_2} \inf\{f(x) \mid x \in \mathrm{cl}(R_{l_1,\dots,l_n})\}\mathrm{vol}(R_{l_1,\dots,l_n}) \\
&\leq \sum_{(l_1,\dots,l_n)\in L_1} M\mathrm{vol}(R_{l_1,\dots,l_n}) + \sum_{(l_1,\dots,l_n)\in L_1} \frac{m}{2}\mathrm{vol}(R_{l_1,\dots,l_n}) \\
&\leq \sum_{(l_1,\dots,l_n)\in L_1} M\mathrm{vol}(R_{l_1,\dots,l_n}) + \frac{m}{2}\mathrm{vol}(R).
\end{aligned}$$

Therefore,

$$\sum_{(l_1,\dots,l_n)\in L_1} \mathrm{vol}(R_{l_1,\dots,l_n}) \geq \frac{m}{4M}\mathrm{vol}(R),$$

giving the lemma. ▼

Using the preceding lemma we prove the following result.

2 Lemma *Let* $R \subseteq \mathbb{R}^n$ *be a fat compact rectangle, let* $(g_j)_{j \in \mathbb{Z}_{>0}}$ *be a sequence of* $\mathbb{R}$*-valued functions on* R *satisfying the following conditions:*

(i) $g_j(\boldsymbol{x}) \geq 0$ *for each* $\boldsymbol{x} \in R$ *and* $j \in \mathbb{Z}_{>0}$*;*

(ii) $g_{j+1}(\boldsymbol{x}) \leq g_j(\boldsymbol{x})$ *for each* $\boldsymbol{x} \in R$ *and* $j \in \mathbb{Z}_{>0}$*;*

(iii) g_j *is Riemann integrable (in the sense of Definition 1.6.22) for each* $j \in \mathbb{Z}_{>0}$*;*

(iv) $\lim_{j\to\infty} g_j(\boldsymbol{x}) = 0$ *for all* $\boldsymbol{x} \in R$.

Then

$$\lim_{j\to\infty} \int_R g_j(\boldsymbol{x})\, d\boldsymbol{x} = 0.$$

Proof The hypotheses ensure that the sequence whose jth term is $\int_R g_j(\boldsymbol{x})\, d\boldsymbol{x}$ is monotonically decreasing and positive. Therefore, it converges by Theorem I-2.3.8. Let us denote the limit by $\tilde{L} \geq 0$ and suppose, in fact, that $\tilde{L} > 0$. Let us denote $L = \frac{\tilde{L}}{\operatorname{vol}(R)}$. For $j \in \mathbb{Z}_{>0}$, let $g_{j,M}\colon R \to \mathbb{R}_{\geq 0}$ be defined by $g_{j,M}(\boldsymbol{x}) = \min\{g_j(\boldsymbol{x}), M\}$. Since g_1 is Riemann integrable in the sense of Definition 1.6.22, let $M_0 \in \mathbb{R}_{>0}$ be such that $M_0 > \frac{2L}{5}\operatorname{vol}(R)$ and such that

$$\int_R g_1(\boldsymbol{x})\, d\boldsymbol{x} - \int_R g_{1,M_0}(\boldsymbol{x})\, d\boldsymbol{x} \leq \frac{L}{5}\operatorname{vol}(R).$$

For each $j \in \mathbb{Z}_{>0}$ we have

$$\{\boldsymbol{x} \in R \mid g_j(\boldsymbol{x}) \geq M_0\} \subseteq \{\boldsymbol{x} \in R \mid g_1(\boldsymbol{x}) \geq M_0\}$$

since $g_j(\boldsymbol{x}) \leq g_1(\boldsymbol{x})$ for all $\boldsymbol{x} \in R$. This gives

$$0 \leq \int_R (g_j(\boldsymbol{x}) - g_{j,M_0}(\boldsymbol{x}))\, d\boldsymbol{x} \leq \int_R (g_1(\boldsymbol{x}) - g_{1,M_0}(\boldsymbol{x}))\, d\boldsymbol{x} \leq \frac{L}{5}\operatorname{vol}(R).$$

Since $\int_R g_j(\boldsymbol{x})\, d\boldsymbol{x} \geq L\operatorname{vol}(R)$ (by definition of L) it follows that $\int_R g_{j,M_0}(\boldsymbol{x})\, d\boldsymbol{x} \geq \frac{4L}{5}\operatorname{vol}(R)$. Now, for $j \in \mathbb{Z}_{>0}$, define

$$E_j = \left\{\boldsymbol{x} \in R \;\middle|\; g_j(\boldsymbol{x}) \geq \tfrac{2L}{5}\operatorname{vol}(R)\right\}.$$

Since $M_0 \geq \frac{2L}{5}\operatorname{vol}(R)$ we also have

$$E_j = \left\{\boldsymbol{x} \in R \;\middle|\; g_{j,M_0}(\boldsymbol{x}) \geq \tfrac{2L}{5}\operatorname{vol}(R)\right\}.$$

By Lemma 1 the set E_j contains a finite number of rectangles whose total volume is bounded below by $\frac{L}{5M_0}\operatorname{vol}(R)$. By Theorem 1.6.11 and Exercise 1.2.12 it follows that the set

$$D = \cup_{j\in\mathbb{Z}_{>0}}\{\boldsymbol{x} \in R \mid g_j(\boldsymbol{x}) \text{ is discontinuous at } \boldsymbol{x}\}$$

has measure zero. Therefore, there is a countable collection of open rectangles covering D and having total volume bounded above by $\frac{L}{10M_0}\operatorname{vol}(R)$. Denote by U the union of these rectangles. We claim that $E_j \not\subseteq U$ for each $j \in \mathbb{Z}_{>0}$. Indeed, if $E_j \subseteq U$ then $\operatorname{vol}(E_j) \leq \operatorname{vol}(U)$, but this cannot be since $\operatorname{vol}(E_j) \geq \frac{L}{5M_0}\operatorname{vol}(R)$ and $\operatorname{vol}(U) \leq \frac{L}{10M_0}\operatorname{vol}(R)$. Let $\boldsymbol{x} \in \operatorname{cl}(E_j) \setminus E_j$. Thus $g_j(\boldsymbol{x}) < \frac{2L}{5}\operatorname{vol}(R)$ by definition of E_j. There then exists a sequence $(\boldsymbol{x}_k)_{k\in\mathbb{Z}_{>0}}$ in E_j converging to $\boldsymbol{x}_0$. Since $g_j(\boldsymbol{x}_k) \geq \frac{2L}{5}\operatorname{vol}(R)$ for each $k \in \mathbb{Z}_{>0}$ by definition of E_j it follows that $\lim_{k\to\infty} g_j(\boldsymbol{x}_k) \neq g_j(\boldsymbol{x})$, and so g_j is discontinuous at $\boldsymbol{x}$. Thus $\boldsymbol{x} \in D \subseteq U$. This shows that $\mathrm{cl}(E_j) \subseteq E_j \cup U$. Now, for $j \in \mathbb{Z}_{>0}$, define $F_j = \mathrm{cl}(E_j) - U$ so that $F_j \subseteq E_j$. Thus F_j is bounded since E_j is bounded. We claim that it is also closed. To see this, let $(\boldsymbol{x}_k)_{k\in\mathbb{Z}_{>0}}$ be a sequence in F_j converging to $\boldsymbol{x}$. Since $F_j \subseteq \mathrm{cl}(E_j)$ it follows that $\boldsymbol{x} \in \mathrm{cl}(E_j)$. We also claim that $\boldsymbol{x} \notin U$. Indeed, since U is open, if $\boldsymbol{x} \in U$ it must follow that $\boldsymbol{x}_k \in U$ for sufficiently large k, contradicting the fact that $(\boldsymbol{x}_k)_{k\in\mathbb{Z}_{>0}}$ is a sequence in F_j. Thus $\boldsymbol{x} \in \mathrm{cl}(E_j) - U = F_j$. Thus F_j is closed and so compact by the Heine–Borel Theorem. Since $E_{j+1} \subseteq E_j$ it follows that $F_{j+1} \subseteq F_j$. By Proposition 1.2.39 it follows that $\cap_{j\in\mathbb{Z}_{>0}} F_j$ is nonempty. Thus, $\cap_{j\in\mathbb{Z}_{>0}} E_j$ is nonempty. Thus there exists $\boldsymbol{x} \in R$ such that $g_j(\boldsymbol{x}) \geq \frac{2L}{5}\mathrm{vol}(R)$, contradicting the fact that the sequence $(g_j)_{j\in\mathbb{Z}_{>0}}$ converges pointwise to zero. ▼

Now we proceed with the proof of the theorem. With $(f_j)_{j\in\mathbb{Z}_{>0}}$ and f as in the statement of the theorem, let $g_j = f - f_j$ for $j \in \mathbb{Z}_{>0}$. One can easily verify that the sequence $(g_j)_{j\in\mathbb{Z}_{>0}}$ satisfies the hypotheses of Lemma 2. Thus, by the lemma,

$$0 = \lim_{j\to\infty} \int_R g_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \lim_{j\to\infty} \int_R (f(\boldsymbol{x}) - f_j(\boldsymbol{x}))\,\mathrm{d}\boldsymbol{x} = \int_R f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \lim_{j\to\infty} \int_R f_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},$$

where we have used Proposition 1.6.28. This gives the result. ■

Let us give some examples which show the value and limitations of the Monotone Convergence Theorem for the Riemann integral.

1.7.7 Examples (The Monotone Convergence Theorem for the Riemann integral)

1. Let $(f_j)_{j\in\mathbb{Z}_{>0}}$ be an enumeration of the rational numbers in the interval $[0,1]$; such an enumeration is possible by Exercise I-2.1.3. Define a sequence of functions $(g_j)_{j\in\mathbb{Z}_{>0}}$ from $[0,1]$ to $\mathbb{R}_{\geq 0}$ by

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

Then define $f_k = \sum_{j=1}^{k} g_j$. One easily verifies that the sequence of functions $(f_k)_{k\in\mathbb{Z}_{>0}}$ satisfies the first three hypotheses of the Monotone Convergence Theorem. Moreover, since the Riemann integral of each of the functions g_j, $j \in \mathbb{Z}_{>0}$, is zero (why?) it follows by Proposition 1.6.28 that each of functions f_k, $k \in \mathbb{Z}_{>0}$, has Riemann integral zero. Thus

$$\lim_{k\to\infty} f_k(x)\,\mathrm{d}x = 0.$$

However, the pointwise limit of the sequence $(f_k)_{k\in\mathbb{Z}_{>0}}$ is the function $f\colon [0,1] \to \mathbb{R}_{\geq 0}$ defined by

$$f(x) = \begin{cases} 1, & x \in \mathbb{Q}, \\ 0, & \text{otherwise,} \end{cases}$$

i.e., the characteristic function of $\mathbb{Q} \cap [0,1]$. However, we have already seen in Example I-3.4.10 that this function is not Riemann integrable. Thus the Monotone Convergence Theorem for the Riemann integral does not hold in this case.
Punchline: The condition that the pointwise limit function f is Riemann integrable appears in the *hypotheses* of the Monotone Convergence Theorem, not in its *conclusions*. This is a significant defect of the Riemann integral. As we shall see with the various versions of the Monotone Convergence Theorem in Chapter III-2, for more general notions of the integral the integrability of the pointwise limit function follows as a conclusion.

2. On $[0,1]$ consider the sequence of functions $(f_j)_{j\in\mathbb{Z}_{>0}}$ given by

$$f_j(x) = \begin{cases} \frac{1}{(jx)^{1/2}}, & x \in (0,1], \\ 0, & x = 0. \end{cases}$$

One can readily verify (cf. Example 1.6.26) that each of the functions f_j is Riemann integrable. Moreover, for each $x \in [0,1]$ it follows that $\lim_{j\to\infty} f_j(x) = 0$. Thus the pointwise limit of the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ is the zero function. Therefore, the limit function is Riemann integrable. Note that this sequence does not quite satisfy the hypotheses of the Monotone Convergence Theorem since the sequence $(f_j(x))_{j\in\mathbb{Z}_{>0}}$ is monotonically decreasing, not increasing, for each $x \in [0,1]$. However, the Monotone Convergence Theorem more or less obviously applies to this case as well (also see Lemma 2 in the proof of the Monotone Convergence Theorem). Indeed, the Monotone Convergence Theorem gives

$$\lim_{j\to\infty} \int_0^1 \frac{1}{(jx)^{1/2}}\,\mathrm{d}x = 0.$$

This can also be checked directly.
Punchline: The Monotone Convergence Theorem applies to sequences of possibly unbounded functions. •

The following result gives conditions, in the absence of positivity of the functions in the sequence, under which we can swap limit and integral.

1.7.8 Theorem (Dominated Convergence Theorem for the Riemann integral) *Let* $R \subseteq \mathbb{R}^n$ *be a fat compact rectangle and let* $(f_j)_{j\in\mathbb{Z}_{>0}}$ *be a sequence of* $\mathbb{R}$*-valued functions on* R *satisfying the following conditions:*

(i) there exists $M \in \mathbb{R}_{>0}$ *such that* $f_j(\mathbf{x}) \le M$ *for each* $\mathbf{x} \in R$ *and* $j \in \mathbb{Z}_{>0}$*;*

(ii) f_j *is Riemann integrable for each* $j \in \mathbb{Z}_{>0}$*;*

(iii) the map $f\colon R \to \mathbb{R}$ *defined by* $f(\mathbf{x}) = \lim_{j\to\infty} f_j(\mathbf{x})$ *exists and is Riemann integrable.*

Then

$$\lim_{j\to\infty} \int_R f_j(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_R f(\mathbf{x})\,\mathrm{d}\mathbf{x}.$$

Proof We first prove a lemma.

1 Lemma *Let* $(\mathrm{A_j})_{\mathrm{j}\in\mathbb{Z}_{>0}}$ *be a sequence of subsets of* $\mathbb{R}^{\mathrm{n}}$ *having the properties*

(i) $\mathrm{A_{j+1}} \subseteq \mathrm{A_j}$, $\mathrm{j} \in \mathbb{Z}_{>0}$, *and*

(ii) $\cap_{\mathrm{j}\in\mathbb{Z}_{>0}}\mathrm{A_j} = \varnothing$.

For $\mathrm{j} \in \mathbb{Z}_{>0}$ *define*

$$\nu_{\mathrm{j}} = \inf\{\mathrm{vol(B)} \mid \mathrm{B} \subseteq \mathrm{A_j} \text{ is a finite union of rectangles}\}.$$

Then $\lim_{\mathrm{j}\to\infty} \nu_{\mathrm{j}} = 0$.

Proof If there exists $N \in \mathbb{Z}_{>0}$ such that A_N contains no set which is a finite union of fat rectangles then it follows that the sets A_j, $j \geq N$, contain no set which is a finite union of fat rectangles. In this case, the lemma holds vacuously. Thus we can suppose, without loss of generality, that each set A_j contains a set which is a finite union of fat rectangles. Since the sequence of subsets $(A_j)_{j\in\mathbb{Z}_{>0}}$ is decreasing with respect to the partial order of inclusion it follows that the sequence $(\nu_j)_{j\in\mathbb{Z}_{>0}}$ is a decreasing sequence of strictly positive numbers. This sequence converges by Theorem I-2.3.8. Suppose that it converges to $L \in \mathbb{R}_{>0}$. For each $j \in \mathbb{Z}_{>0}$ let $B_j \subseteq A_j$ be a finite union of closed fat rectangles having the property that

$$\mathrm{vol}(B_j) = \nu_j - \tfrac{L}{2^j}. \tag{1.45}$$

For $m \in \mathbb{Z}_{>0}$ let us define $K_m = \cap_{j=1}^m B_j$. Since K_m is an intersection of closed sets it is closed by Exercise 1.2.3. Since the sets K_m, $m \in \mathbb{Z}_{>0}$, are obviously bounded it follows from the Heine–Borel Theorem that they are compact. We next claim that K_m is nonempty for each $m \in \mathbb{Z}_{>0}$. Let $j \in \mathbb{Z}_{>0}$. If $B \subseteq A_j \setminus B_j$ is a finite union of rectangles then we have

$$\mathrm{vol}(B) + \mathrm{vol}(B_j) = \mathrm{vol}(B \cup B_j) \leq \nu_j$$

since B and B_j are disjoint. By (1.45) it then follows that

$$\mathrm{vol}(B) \leq \frac{L}{2^j}. \tag{1.46}$$

Now, for $m \in \mathbb{Z}_{>0}$, let $B \subseteq A_m \setminus K_m$. By Proposition I-1.1.5 we have

$$B = (B \setminus B_1) \cup \cdots \cup (B \setminus B_m). \tag{1.47}$$

Since B and B_j are each finite unions of rectangles, $B \setminus B_1$ is a finite union of rectangles for each $j \in \{1, \dots, m\}$ (why?). Therefore, for each $j \in \{1, \dots, m\}$, $B \setminus B_j$ is a subset of $A_j \setminus E_j$ that is a finite union of rectangles. By (1.46) this means that $\mathrm{vol}(B \setminus B_j) < \frac{L}{2^n}$, $j \in \{1, \dots, m\}$. By (1.47) it follows that

$$\mathrm{vol}(B) \leq \sum_{j=1}^m \mathrm{vol}(B \setminus B_j) \leq L \sum_{j=1}^m \frac{1}{2^j} < L.$$

Now, since A_m must contain a set which is a finite union of rectangles with the union having volume at least L, and since any subset of $A_m \setminus K_m$ that is a finite union of rectangles has volume at most L, it follows that $K_m \neq \varnothing$. Now, by Proposition 1.2.39 it follows that $\cap_{m=1}^\infty K_m \neq \varnothing$. Since $K_j \subseteq A_j$ for each $j \in \mathbb{Z}_{>0}$, it then follows that $\cap_{j=1}^\infty A_j \neq \varnothing$, so violating the hypotheses of the lemma. Thus the assumption that the sequence $(\nu_j)_{j\in\mathbb{Z}_{>0}}$ converges to a positive number is invalid. ▼

Next we prove the theorem for the case when the functions f_j, $j \in \mathbb{Z}_{>0}$, take values in $\mathbb{R}_{\geq 0}$ and when the limit function f is the zero function. In this case, let $\epsilon \in \mathbb{R}_{>0}$ and for $j \in \mathbb{Z}_{>0}$ define

$$A_j = \left\{ \boldsymbol{x} \in R \;\middle|\; f_k(\boldsymbol{x}) \geq \tfrac{\epsilon}{4\mathrm{vol}(R)} \text{ for some } k \geq j \right\}.$$

Clearly $A_{j+1} \subseteq A_j$ for all $j \in \mathbb{Z}_{>0}$. Moreover, since the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ converges pointwise to zero, $\cap_{j\in\mathbb{Z}_{>0}} A_j = \emptyset$. By the lemma above let $N \in \mathbb{Z}_{>0}$ be sufficiently large that, for $j \geq N$, if $B \subseteq A_j$ is a finite union of rectangles then $\mathrm{vol}(B) < \frac{\epsilon}{4M}$. Let $j \geq N$ and let $\boldsymbol{P}$ be a partition such that

$$\int_R f_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_R A_-(f_j, \boldsymbol{P})(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} < \frac{\epsilon}{2}.$$

Define

$$B = \left\{ \boldsymbol{x} \in R \;\middle|\; A_-(f_j, \boldsymbol{P})(\boldsymbol{x}) \geq \tfrac{\epsilon}{4\mathrm{vol}(R)} \right\}$$

and $B' = R \setminus B$. Since $A_-(f, \boldsymbol{P})$ is a step function, B, and therefore B', is a finite union of rectangles. We then have

$$\begin{aligned}
\int_R f_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \int_R f_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_R A_-(f, \boldsymbol{P})(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int_R A_-(f, \boldsymbol{P})(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&\leq \frac{\epsilon}{2} + \int_B A_-(f, \boldsymbol{P})(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} + \int_{B'} A_-(f, \boldsymbol{P})(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \\
&\leq \frac{\epsilon}{2} + M\mathrm{vol}(B) + \frac{\epsilon}{4\mathrm{vol}(R)}\mathrm{vol}(B') \leq \frac{\epsilon}{2} + \frac{\epsilon}{4} + \frac{\epsilon}{4} = \epsilon.
\end{aligned}$$

Thus $\lim_{j\to\infty} \int_R f_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 0$ giving the theorem in this case.

Finally to prove the theorem, given the sequence $(f_j)_{j\in\mathbb{Z}_{>0}}$ and f as in the statement of the theorem, define $g_j(\boldsymbol{x}) = |f(\boldsymbol{x}) - f_j(\boldsymbol{x})|$, $\boldsymbol{x} \in R$, $j \in \mathbb{Z}_{>0}$. By Propositions 1.6.28 and 1.6.31 it follows that the functions g_j, $j \in \mathbb{Z}_{>0}$, are Riemann integrable. Moreover, they take values in $\mathbb{R}_{\geq 0}$ and converge pointwise to zero. Therefore, by the special case we considered above we have

$$\lim_{j\to\infty} \left| \int_R f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_R f_j(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} \right| \leq \lim_{j\to\infty} \int_R |f(\boldsymbol{x}) - f_j(\boldsymbol{x})|\,\mathrm{d}\boldsymbol{x} = 0,$$

using Proposition 1.6.31. Thus the theorem follows. ■

1.7.4 Notes

The Dominated Convergence Theorem for the Riemann integral is due to Arzelà [1885] and Arzelà [1900], and the proof we give is an adaptation of the proof of Lewin [1986]. See also [Gordon 2000].

Section 1.8

Multivariable $\mathbb{R}$-power series

In this section we study the multivariable generalisation of $\mathbb{R}$-power series given in the single-variable case in Section I-3.7. The presentation here bears strong resemblance to the single-variable case, with the main difference being that the notation is more ponderous.

Do I need to read this section? This section should be read when needed, or when the reader becomes interested in reading it. •

1.8.1 Multivariable $\mathbb{R}$-formal power series

1.8.2 Multivariable $\mathbb{R}$-convergent power series

1.8.3 Multivariable $\mathbb{R}$-convergent power series and operations on functions

1.8.4 Multivariable Taylor series

1.8.5 Multivariable $\mathbb{R}$-power series

1.8.6 Notes

Section 1.9

Elementary convexity

Convexity is an important tool in many applications. Restricting domains for a problem to be convex is often a natural thing to do, and upon doing so one has at one's disposal a significant body of knowledge. In this section we work with convex subsets of $\mathbb{R}^n$. Many of the important applications of convexity arise in an infinite-dimensional setting, and we shall see a *huge* portion of this in Chapter III-6. However, it is useful to have at hand the more elementary finite-dimensional theory to get some substantial insight into the more difficult applications of convexity.

Do I need to read this section? The material in this section can be skipped until needed. •

1.9.1 Definitions

There are a few basic concepts one needs to get started, and these are them.

1.9.1 Definition (Convex set, cone, convex cone, affine subspace)

(i) A subset $C \subseteq \mathbb{R}^n$ is ***convex*** if, for each $\boldsymbol{x}_1, \boldsymbol{x}_2 \in C$, we have

$$\{s\boldsymbol{x}_1 + (1-s)\boldsymbol{x}_2 \mid s \in [0,1]\} \subseteq C.$$

(ii) A subset $K \subseteq \mathbb{R}^n$ is a ***cone*** if there exists $\boldsymbol{x}_0 \in \mathbb{R}^n$ such that, for each $\boldsymbol{x} \in K$, we have

$$\{\boldsymbol{x}_0 + \lambda(\boldsymbol{x} - \boldsymbol{x}_0) \mid \lambda \in \mathbb{R}_{\geq 0}\} \subseteq K.$$

The point $\boldsymbol{x}_0$ is the ***vertex*** of the cone.

(iii) A subset $K \subseteq \mathbb{R}^n$ is a ***convex cone*** if it is both convex and a cone.

(iv) A subset $\mathsf{A} \subseteq \mathbb{R}^n$ is an ***affine subspace*** if, for each $\boldsymbol{x}_1, \boldsymbol{x}_2 \in A$, we have

$$\{s\boldsymbol{x}_1 + (1-s)\boldsymbol{x}_2 \mid s \in \mathbb{R}\} \subseteq \mathsf{A}.$$ •

Note that the set $\{s\boldsymbol{x}_1 + (1-s)\boldsymbol{x}_2 \mid s \in [0,1]\}$ is the line segment in $\mathbb{R}^n$ between $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Thus a set is convex when the line segment connecting any two points in the set remains in the set. In a similar manner, $\{\boldsymbol{x}_0 + \lambda(\boldsymbol{x} - \boldsymbol{x}_0) \mid \lambda \in \mathbb{R}_{\geq 0}\}$ is the ray emanating from $\boldsymbol{x}_0 \in \mathbb{R}^n$ through the point $\boldsymbol{x}$. A set is thus a cone when the rays emanating from $\boldsymbol{x}_0$ through all points remain in the set. An affine subspace is a set where the (bi-infinite) line through any two points in the set remains in the set. We illustrate some of the intuition concerning these various sorts of sets in Figure 1.13.

1.9.2 Combinations and hulls

We shall be interested in generating convex sets, cones, and affine subspaces containing given sets.

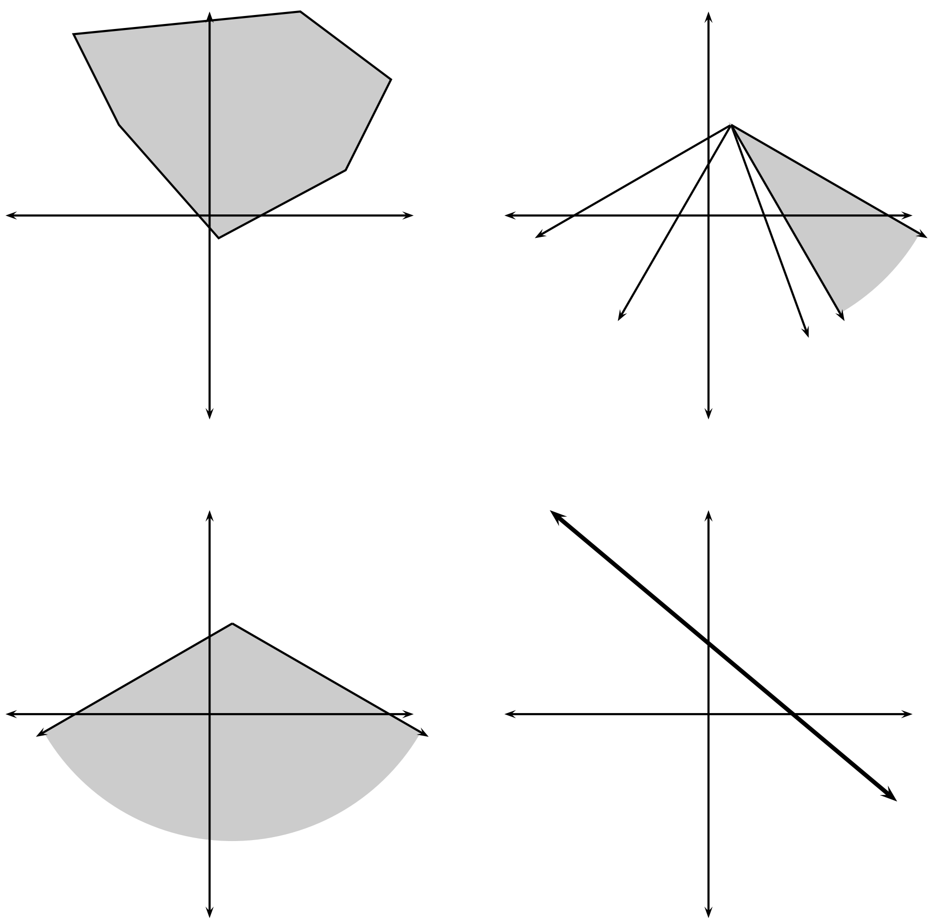

Figure 1.13 An illustration of a convex set (top left), a cone (top right), a convex cone (bottom left), and an affine subspace (bottom right)

1.9.2 Definition (Convex hull, coned hull, coned convex hull, affine hull) Let $S \subseteq \mathbb{R}^n$ be nonempty.

(i) A ***convex combination*** from S is a linear combination in $\mathbb{R}^n$ of the form

$$\sum_{j=1}^{k} \lambda_j \boldsymbol{v}_j, \qquad k \in \mathbb{Z}_{>0},\ \lambda_1, \dots, \lambda_k \in \mathbb{R}_{\geq 0},\ \sum_{j=1}^{k} \lambda_j = 1,\ \boldsymbol{v}_1, \dots, \boldsymbol{v}_k \in S.$$

(ii) The ***convex hull*** of S, denoted by $\mathrm{conv}(S)$, is the smallest convex subset of $\mathbb{R}^n$ containing S.

(iii) The ***coned hull*** of S with vertex $\boldsymbol{x}_0$, denoted by $\mathrm{cone}(S, \boldsymbol{x}_0)$, is the smallest cone in $\mathbb{R}^n$ with vertex at $\boldsymbol{x}_0$ containing S.

(iv) A ***coned convex combination*** from S with vertex at $\boldsymbol{x}_0$ is a linear combination in $\mathbb{R}^n$ of the form

$$\boldsymbol{x}_0 + \sum_{j=1}^{k} \lambda_j(\boldsymbol{v}_j - \boldsymbol{x}_0), \qquad k \in \mathbb{Z}_{>0},\ \lambda_1, \dots, \lambda_k \in \mathbb{R}_{\geq 0},\ \boldsymbol{v}_1, \dots, \boldsymbol{v}_k \in S.$$

(v) The ***coned convex hull*** of S with vertex $\boldsymbol{x}_0$, denoted by $\operatorname{conv}\operatorname{cone}(S, \boldsymbol{x}_0)$, is the smallest convex cone in $\mathbb{R}^n$ with vertex $\boldsymbol{x}_0$ containing S.

(vi) An ***affine combination*** from S is a linear combination in $\mathbb{R}^n$ of the form

$$\sum_{j=1}^{k} \lambda_j \boldsymbol{v}_j, \qquad k \in \mathbb{Z}_{>0},\ \lambda_1, \dots, \lambda_k \in \mathbb{R},\ \sum_{j=1}^{k} \lambda_j = 1,\ \boldsymbol{v}_1, \dots, \boldsymbol{v}_k \in S.$$

(vii) The ***affine hull*** of S, denoted by $\operatorname{aff}(S)$, is the smallest affine subspace of $\mathbb{R}^n$ containing S. •

1.9.3 Remark (Sensibility of hull definitions) The definitions of $\operatorname{conv}(S)$, $\operatorname{cone}(S, \boldsymbol{x}_0)$, $\operatorname{conv}\operatorname{cone}(S, \boldsymbol{x}_0)$, and $\operatorname{aff}(S)$ make sense because intersections of convex sets are convex, intersections of cones with vertex $\boldsymbol{x}_0$ are cones with vertex $\boldsymbol{x}_0$, and intersections of affine subspaces are affine subspaces. See Exercise 1.9.3. •

The terms "coned hull" and "coned convex hull" are not standard. In the literature these will often be called the "cone generated by S" and the "convex cone generated by S," respectively.

Convex combinations have the following useful property which also describes the convex hull.

1.9.4 Proposition (The convex hull is the set of convex combinations) *Let* $\mathrm{S} \subseteq \mathbb{R}^{\mathrm{n}}$ *be nonempty and denote by* $\mathrm{C(S)}$ *the set of convex combinations from* S*. Then* $\mathrm{C(S)} = \mathrm{conv(S)}$*.*

Proof First we show that $C(S)$ is convex. Consider two elements of $C(S)$ given by

$$\boldsymbol{x} = \sum_{j=1}^{k} \lambda_j \boldsymbol{u}_j, \quad \boldsymbol{y} = \sum_{l=1}^{m} \mu_l \boldsymbol{v}_l.$$

Then, for $s \in [0, 1]$ we have

$$s\boldsymbol{x} + (1 - s)\boldsymbol{y} = \sum_{j=1}^{k} s\lambda_j \boldsymbol{u}_j + \sum_{l=1}^{m} (1 - s)\mu_j \boldsymbol{v}_j.$$

For $r \in \{1, \dots, k + m\}$ define

$$\boldsymbol{w}_r = \begin{cases} \boldsymbol{u}_r, & r \in \{1, \dots, k\}, \\ \boldsymbol{v}_{r-k}, & r \in \{k + 1, \dots, k + m\} \end{cases}$$

and

$$\rho_r = \begin{cases} s\lambda_r, & r \in \{1, \dots, k\}, \\ (1-s)\mu_{r-k}, & r \in \{k+1, \dots, k+m\}. \end{cases}$$

Clearly $\boldsymbol{w}_r \in S$ and $\rho_r \in \mathbb{R}_{\geq 0}$ for $r \in \{1, \dots, k+m\}$. Also,

$$\sum_{r=1}^{k+m} \rho_r = \sum_{j=1}^{k} s\lambda_j + \sum_{l=1}^{m} (1-s)\mu_l = s + (1-s) = 1.$$

Thus $s\boldsymbol{x} + (1-s)\boldsymbol{y} \in C(S)$, and so $C(S)$ is convex.

This necessarily implies that $\mathrm{conv}(S) \subseteq C(S)$ since $\mathrm{conv}(S)$ is the smallest convex set containing S. To show that $C(S) \subseteq \mathrm{conv}(S)$ we will show by induction on the number of elements in the linear combination that all convex combinations are contained in the convex hull. This is obvious for the convex combination of one vector. So suppose that every convex combination of the form

$$\sum_{j=1}^{k} \lambda_j \boldsymbol{u}_j, \qquad k \in \{1, \dots, m\},$$

is in $\mathrm{conv}(S)$, and consider a convex combination from S of the form

$$\boldsymbol{y} = \sum_{l=1}^{m+1} \mu_l \boldsymbol{v}_l = \sum_{l=1}^{m} \mu_l \boldsymbol{v}_l + \mu_{m+1} \boldsymbol{v}_{m+1}.$$

If $\sum_{l=1}^{m} \mu_l = 0$ then $\mu_l = 0$ for each $l \in \{1, \dots, m\}$. Thus $\boldsymbol{y} \in \mathrm{conv}(S)$ by the induction hypothesis. So we may suppose that $\sum_{l=1}^{m} \mu_l \neq 0$ which means that $\mu_{m+1} \neq 1$. Let us define $\mu_l' = \mu_l (1 - \mu_{m+1})^{-1}$ for $l \in \{1, \dots, m\}$. Since

$$1 - \mu_{m+1} = \sum_{l=1}^{m} \mu_l$$

it follows that

$$\sum_{l=1}^{m} \mu_l' = 1.$$

Therefore,

$$\sum_{l=1}^{m} \mu_l' \boldsymbol{v}_l \in \mathrm{conv}(S)$$

by the induction hypothesis. But we also have

$$\boldsymbol{y} = (1 - \mu_{m+1}) \sum_{l=1}^{m} \mu_l' \boldsymbol{v}_l + \mu_{m+1} \boldsymbol{v}_{m+1}$$

by direct computation. Therefore, $\boldsymbol{y}$ is a convex combination of two elements of $\mathrm{conv}(S)$. Since $\mathrm{conv}(S)$ is convex, this means that $\boldsymbol{y} \in \mathrm{conv}(S)$, giving the result. ■

For cones one has a similar result.

1.9.5 Proposition (The set of positive multiples is the coned hull) *Let* $\mathrm{S} \subseteq \mathbb{R}^n$ *be nonempty and denote*

$$\mathrm{K}(\mathrm{S}, \mathbf{x}_0) = \{\lambda(\mathbf{x} - \mathbf{x}_0) \mid \mathbf{x} \in \mathrm{S},\ \lambda \in \mathbb{R}_{\geq 0}\}.$$

Then $\mathrm{K}(\mathrm{S}, \mathbf{x}_0) = \mathrm{cone}(\mathrm{S}, \mathbf{x}_0)$.

Proof Note that $K(S\boldsymbol{x}_0)$ is clearly a cone which contains S. Thus $\mathrm{cone}(S, \boldsymbol{x}_0) \subseteq K(S, \boldsymbol{x}_0)$. Now suppose that $\boldsymbol{y} \in K(S, \boldsymbol{x}_0)$. Thus $\boldsymbol{y} = \lambda(\boldsymbol{x} - \boldsymbol{x}_0)$ for $\boldsymbol{x} \in S$ and $\lambda \in \mathbb{R}_{\geq 0}$. Since $\mathrm{cone}(S)$ is a cone with vertex $\boldsymbol{x}_0$ containing $\boldsymbol{x}$, we must have $\boldsymbol{y} \in \mathrm{cone}(S)$, giving $K(S, \boldsymbol{x}_0) \subseteq \mathrm{cone}(S, \boldsymbol{x}_0)$. ■

Finally, one has an interpretation along these lines for convex cones.

1.9.6 Proposition (The coned convex hull is the set of coned convex combinations) *Let* $\mathrm{S} \subseteq \mathbb{R}^n$ *be nonempty and denote by* $\mathrm{K}'(\mathrm{S}, \mathbf{x}_0)$ *the set of coned convex combinations with vertex* $\mathbf{x}_0$ *from* S. *Then* $\mathrm{K}'(\mathrm{S}, \mathbf{x}_0) = \mathrm{conv\,cone}(\mathrm{S}, \mathbf{x}_0)$.

Proof Let $\boldsymbol{x}, \boldsymbol{y} \in K'(S, \boldsymbol{x}_0)$ and write

$$\boldsymbol{x} = \boldsymbol{x}_0 + \sum_{j=1}^{k} \lambda_j(\boldsymbol{u}_j - \boldsymbol{x}_0), \quad \boldsymbol{y} = \boldsymbol{x}_0 + \sum_{l=1}^{m} \mu_l(\boldsymbol{v}_l - \boldsymbol{x}_0).$$

Then, for $\lambda \in \mathbb{R}_{\geq 0}$,

$$\boldsymbol{x}_0 + \lambda(\boldsymbol{x} - \boldsymbol{x}_0) = \boldsymbol{x}_0 + \sum_{j=1}^{k} \lambda\lambda_j(\boldsymbol{u}_j - \boldsymbol{x}_0),$$

immediately giving $\boldsymbol{x}_0 + \lambda(\boldsymbol{x} - \boldsymbol{x}_0) \in K'(S, \boldsymbol{x}_0)$. We also have

$$\begin{aligned}(1-s)\boldsymbol{x} + s\boldsymbol{y} &= \boldsymbol{x}_0 + (1-s)(\boldsymbol{x} - \boldsymbol{x}_0) + s(\boldsymbol{y} - \boldsymbol{x}_0) \\ &= \boldsymbol{x}_0 + \sum_{j=1}^{k}(1-s)\lambda_j(\boldsymbol{u}_j - \boldsymbol{x}_0) + \sum_{l=1}^{m} s\mu_l(\boldsymbol{v}_l - \boldsymbol{x}_0)\end{aligned}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in K'(S, \boldsymbol{x}_0)$ and $s \in [0, 1]$. Thus $K'(S, \boldsymbol{x}_0)$ is convex. It is evident that $K'(S, \boldsymbol{x}_0)$ is also a cone, and so we must have $\mathrm{conv\,cone}(S, \boldsymbol{x}_0) \subseteq K'(S, \boldsymbol{x}_0)$.

Now let

$$\boldsymbol{y} = \boldsymbol{x}_0 + \sum_{j=1}^{k} \lambda_j(\boldsymbol{v}_j - \boldsymbol{x}_0) \in K'(S, \boldsymbol{x}_0).$$

By the fact that $\mathrm{conv\,cone}(S, \boldsymbol{x}_0)$ is a cone containing S we must have $\boldsymbol{x}_0 + k\lambda_j(\boldsymbol{v}_j - \boldsymbol{x}_0) \in \mathrm{conv\,cone}(S, \boldsymbol{x}_0)$ for $j \in \{1, \dots, k\}$. Since $\mathrm{conv\,cone}(S, \boldsymbol{x}_0)$ is convex and contains $\boldsymbol{x}_0 + k\lambda_j(\boldsymbol{v}_j - \boldsymbol{x}_0)$ for $j \in \{1, \dots, k\}$ we must have

$$\sum_{j=1}^{k} \frac{1}{k}(\boldsymbol{x}_0 + k\lambda_j(\boldsymbol{v}_j - \boldsymbol{x}_0)) = \boldsymbol{y} \in \mathrm{conv\,cone}(S, \boldsymbol{x}_0),$$

giving the result. ■

Finally, we prove the expected result for affine subspaces, namely that the affine hull is the set of affine combinations. In order to do this we first give a useful characterisation of affine subspaces.

1.9.7 Proposition (Characterisation of an affine subspace) *A nonempty subset* $\mathsf{A} \subseteq \mathbb{R}^n$ *is an affine subspace if and only if there exists* $\mathbf{x}_0 \in \mathbb{R}^n$ *and a subspace* $\mathsf{U} \subseteq \mathbb{R}^n$ *such that*

$$\mathsf{A} = \{\mathbf{x}_0 + \mathbf{u} \mid \mathbf{u} \in \mathsf{U}\}.$$

Proof Let $x_0 \in \mathsf{A}$ and define $\mathsf{U} = \{x - x_0 \mid x \in \mathsf{A}\}$. The result will be proved if we prove that U is a subspace. Let $x - x_0 \in \mathsf{U}$ for some $x \in \mathsf{A}$ and $a \in \mathbb{R}$. Then

$$a(x - x_0) = ax + (1 - a)x_0 - x_0,$$

and so $a(x - x_0) \in \mathsf{U}$ since $ax + (1 - a)x_0 \in \mathsf{A}$. For $x_1 - x_0, x_2 - x_0 \in \mathsf{U}$ with $x_1, x_2 \in \mathsf{A}$ we have

$$(x_1 - x_0) + (x_2 - x_0) = (x_1 + x_2 - x_0) - x_0.$$

Thus we will have $(x_1 - x_0) + (x_2 - x_0) \in \mathsf{U}$ if we can show that $x_1 + x_2 - x_0 \in \mathsf{A}$. However, we have

$$\begin{aligned}
& x_1 - x_0, x_2 - x_0 \in \mathsf{U}, \\
\implies\quad & 2(x_1 - x_0), 2(x_2 - x_0) \in \mathsf{U}, \\
\implies\quad & 2(x_1 - x_0) + x_0, 2(x_2 - x_0) + x_0 \in \mathsf{A}, \\
\implies\quad & \tfrac{1}{2}(2(x_1 - x_0) + x_0) + \tfrac{1}{2}(2(x_2 - x_0) + x_0) \in \mathsf{A},
\end{aligned}$$

which gives the result after we notice that

$$\tfrac{1}{2}(2(x_1 - x_0) + x_0) + \tfrac{1}{2}(2(x_2 - x_0) + x_0) = x_1 + x_2 - x_0. \qquad \blacksquare$$

Now we can characterise the affine hull as the set of affine combinations.

1.9.8 Proposition (The affine hull is the set of affine combinations) *Let* $\mathsf{S} \subseteq \mathbb{R}^n$ *be nonempty and denote by* $\mathsf{A}(\mathsf{S})$ *the set of affine combinations from* S*. Then* $\mathsf{A}(\mathsf{S}) = \mathrm{aff}(\mathsf{S})$*.*

Proof We first show that the set of affine combinations is an affine subspace. Choose $x_0 \in S$ and define

$$\mathsf{U}(S) = \{v - x_0 \mid v \in \mathsf{A}(S)\}.$$

We first claim that $\mathsf{U}(S)$ is the set of linear combinations of the form

$$\sum_{j=1}^{k} \lambda_j v_j, \qquad k \in \mathbb{Z}_{>0},\ \lambda_1, \dots, \lambda_k \in \mathbb{R},\ \sum_{j=1}^{k} \lambda_j = 0,\ v_1, \dots, v_k \in S. \tag{1.48}$$

To see this, note that if

$$u = \sum_{j=1}^{k} \lambda_j u_j - x_0 \in \mathsf{U}(S)$$

then we can write

$$u = \sum_{j=1}^{k+1} \lambda_j u_j, \qquad \lambda_1, \dots, \lambda_{k+1} \in \mathbb{R},\ \sum_{j=1}^{k+1} \lambda_j = 0,\ u_1, \dots, u_{k+1} \in S,$$

by taking $\lambda_{k+1} = -1$ and $\boldsymbol{u}_{k+1} = \boldsymbol{x}_0$. Similarly, consider a linear combination of the form (1.48). We can without loss of generality suppose that $\boldsymbol{x}_0 \in \{\boldsymbol{v}_1, \dots, \boldsymbol{v}_k\}$, since if this is not true we can simply add $0\boldsymbol{x}_0$ to the sum. Thus we suppose, without loss of generality, that $\boldsymbol{v}_k = \boldsymbol{x}_0$. We then have

$$\boldsymbol{u} = \Big(\sum_{j=1}^{k-1} \lambda_j \boldsymbol{v}_j + (\lambda_k + 1)\boldsymbol{x}_0\Big) - \boldsymbol{x}_0.$$

Since the term in the parenthesis is clearly an element of $\mathsf{A}(S)$ it follows that $u \in \mathsf{U}(S)$.

With this characterisation of $\mathsf{U}(S)$ it is then easy to show that $\mathsf{U}(S)$ is a subspace of $\mathbb{R}^n$. Moreover, it is immediate from Proposition 1.9.8 that $\mathsf{A}(S)$ is then an affine subspace. Since $\operatorname{aff}(S)$ is the smallest affine subspace containing S it follows that $\operatorname{aff}(S) \subseteq \mathsf{A}(S)$. To show that $\mathsf{A}(S) \subseteq \operatorname{aff}(S)$ we use induction on the number of elements in an affine combination in $\mathsf{A}(S)$. For an affine combination with one term this is obvious. So suppose that every affine combination of the form

$$\sum_{j=1}^{k} \lambda_j \boldsymbol{v}_j, \qquad k \in \{1, \dots, m\},$$

is in $\operatorname{aff}(S)$ and consider an affine combination of the form

$$\boldsymbol{x} = \sum_{j=1}^{m+1} \lambda_j \boldsymbol{v}_j = \sum_{j=1}^{m} \lambda_j \boldsymbol{v}_j + \lambda_{m+1} \boldsymbol{v}_{m+1}.$$

It must be the case that at least one of the numbers $\lambda_1, \dots, \lambda_{m+1}$ is not equal to 1. So, without loss of generality suppose that $\lambda_{m+1} \neq 1$ and then define $\lambda_j' = (1 - \lambda_{m+1}^{-1})\lambda_j$, $j \in \{1, \dots, m\}$. We then have

$$\sum_{j=1}^{m} \lambda_j' = 1,$$

so that

$$\sum_{j=1}^{m} \lambda_j' \boldsymbol{v}_j \in \operatorname{aff}(S)$$

by the induction hypothesis. It then holds that

$$\boldsymbol{x} = (1 - \lambda_{m+1}) \sum_{j=1}^{m} \lambda_j' \boldsymbol{v}_j + \lambda_{m+1} \boldsymbol{v}_{m+1}.$$

This is then in $\operatorname{aff}(S)$. ■

1.9.3 Topology of convex sets and cones

Let us now say a few words about the topology of convex sets. Note that every convex set is a subset of its affine hull. We first define the notion of the interior of a subset of an affine subspace.

1.9.9 Definition (Relative interior of a subset of an affine subspace) If $\mathsf{A} \subseteq \mathbb{R}^n$ is an affine subspace and if $S \subseteq \mathsf{A}$, a point $\boldsymbol{x} \in S$ is in the ***interior of*** **S** ***relative to*** **A** if there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{A} \cap \mathsf{B}^n(\epsilon, \boldsymbol{x}) \subseteq S$. The set of points in the interior of S relative to A is denoted by $\mathrm{int}_{\mathsf{A}}(S)$. •

The idea is that an affine subspace A that is a strict subset of $\mathbb{R}^n$ always has empty interior (why?). However, since A "looks like" $\mathbb{R}^k$ for some $k < n$, one would like to still talk about subsets of A as having an interior. This all makes a lot more sense after one knows a little point set topology, particularly as covered in Section III-1.4.1.

With the notion of the interior relative to an affine subspace, we make the following definition.

1.9.10 Definition (Relative interior and relative boundary) If $C \subseteq \mathbb{R}^n$ is a convex set, the set

$$\mathrm{rel\,int}(C) = \{\boldsymbol{x} \in C \mid \boldsymbol{x} \in \mathrm{int}_{\mathrm{aff}(C)}(C)\}$$

is the ***relative interior*** of C and the set $\mathrm{rel\,bd}(C) = \mathrm{cl}(C) \setminus \mathrm{rel\,int}(C)$ is the ***relative boundary*** of C. •

The point is that, while a convex set may have an empty interior, its interior can still be defined in a weaker, but still useful, sense. The notion of relative interior leads to the following useful concept.

1.9.11 Definition (Dimension of a convex set) Let $C \subseteq \mathbb{R}^n$ be convex and let $\mathsf{U} \subseteq \mathbb{R}^n$ be the subspace for which $\mathrm{aff}(C) = \{\boldsymbol{x}_0 + \boldsymbol{u} \mid \boldsymbol{u} \in \mathsf{U}\}$ for some $\boldsymbol{x}_0 \in \mathbb{R}^n$. The ***dimension*** of C, denoted by $\dim(C)$, is the dimension of the subspace U. •

The following result will be used in our development.

1.9.12 Proposition (Closures and relative interiors of convex sets and cones are convex sets and cones) *Let* $\mathrm{C} \subseteq \mathbb{R}^{\mathrm{n}}$ *be convex and let* $\mathrm{K} \subseteq \mathbb{R}^{\mathrm{n}}$ *be a convex cone with vertex* $\mathbf{x}_0$*. Then*

(i) $\mathrm{cl}(\mathrm{C})$ *is convex and* $\mathrm{cl}(\mathrm{K})$ *is a convex cone and*

(ii) $\mathrm{rel\,int}(\mathrm{C})$ *is convex and* $\mathrm{rel\,int}(\mathrm{K})$ *is a convex cone.*

Moreover, $\mathrm{aff}(\mathrm{C}) = \mathrm{aff}(\mathrm{cl}(\mathrm{C}))$ *and* $\mathrm{aff}(\mathrm{K}) = \mathrm{aff}(\mathrm{cl}(\mathrm{K}))$.

Proof (i) Let $\boldsymbol{x}, \boldsymbol{y} \in \mathrm{cl}(C)$ and let $s \in [0,1]$. Suppose that $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ and $(\boldsymbol{y}_j)_{j\in\mathbb{Z}_{>0}}$ are sequences in C converging to $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively. Note that $s\boldsymbol{x}_j + (1-s)\boldsymbol{y}_j \in C$ for each $j \in \mathbb{Z}_{>0}$. Moreover, if $\epsilon \in \mathbb{R}_{>0}$ then

$$\|s\boldsymbol{x} + (1-s)\boldsymbol{y} - s\boldsymbol{x}_j - (1-s)\boldsymbol{y}_j\|_{\mathbb{R}^n} \le s\|\boldsymbol{x} - \boldsymbol{x}_j\|_{\mathbb{R}^n} + (1-s)\|\boldsymbol{y} - \boldsymbol{y}_j\|_{\mathbb{R}^n} < \epsilon,$$

provided that j is sufficiently large that $s\|\boldsymbol{x} - \boldsymbol{x}_j\|_{\mathbb{R}^n} < \frac{\epsilon}{2}$ and $(1-s)\|\boldsymbol{y} - \boldsymbol{y}_j\|_{\mathbb{R}^n} < \frac{\epsilon}{2}$. Thus the sequence $(s\boldsymbol{x}_j + (1-s)\boldsymbol{y}_j)_{j\in\mathbb{Z}_{>0}}$ converges to $s\boldsymbol{x} + (1-s)\boldsymbol{y}$ and so $s\boldsymbol{x} + (1-s)\boldsymbol{y} \in \mathrm{cl}(C)$. This shows that $\mathrm{cl}(C)$ is convex. Since $C \subseteq \mathrm{cl}(C)$ it follows that $\mathrm{aff}(C) \subseteq \mathrm{aff}(\mathrm{cl}(C))$. Moreover, since $C \subseteq \mathrm{aff}(C)$ and since $\mathrm{aff}(C)$ is closed we have

$$\mathrm{cl}(C) \subseteq \mathrm{cl}(\mathrm{aff}(C)) = \mathrm{aff}(C),$$

so giving $\mathrm{aff}(C) = \mathrm{aff}(\mathrm{cl}(C))$ as desired.

An entirely similar argument shows that $\mathrm{cl}(K)$ is convex and that $\mathrm{aff}(K) = \mathrm{aff}(\mathrm{cl}(K))$.

(ii) Let us first consider the convex set C. To simplify matters, since the relative interior is the interior relative to the affine subspace containing C, and since the topology of an affine subspace is "the same as" Euclidean space, we shall assume that $\dim(C) = n$ and show that $\mathrm{int}(C)$ is convex; cf. Example 1.3.38–2.

We first prove a lemma.

1 Lemma *If* C *is a convex set, if* $x \in \mathrm{rel\,int}(C)$*, and if* $y \in \mathrm{cl}(C)$ *then*

$$[x, y) \triangleq \{sx + (1-s)y \mid s \in [0, 1)\}$$

is contained in $\mathrm{rel\,int}(C)$*.*

Proof As in the proof of (ii), let us assume, without loss of generality, that $\dim(C) = n$. Since $x \in \mathrm{int}(C)$ there exists $r \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(r, x) \subseteq C$. Since $y \in \mathrm{cl}(C)$, for every $\epsilon \in \mathbb{R}_{>0}$ there exists $y_\epsilon \in C \cap \mathsf{B}^n(\epsilon, y)$. Let $z = \alpha x + (1-\alpha)y \in [x, y)$ for $\alpha \in [0, 1)$, and define $\delta = \alpha r - (1-\alpha)\epsilon$. If ϵ is sufficiently small we can ensure that $\delta \in \mathbb{R}_{>0}$, and we assume that ϵ is so chosen. For $z' \in \mathsf{B}^n(\delta, z)$ we have

$$\begin{aligned} & \|z' - z\|_{\mathbb{R}^n} < \delta \\ \implies\quad & \|z' - (\alpha x + (1-\alpha)y_\epsilon + (1-\alpha)(y - y_\epsilon))\|_{\mathbb{R}^n} < \delta \\ \implies\quad & \|z' - (\alpha x + (1-\alpha)y_\epsilon)\|_{\mathbb{R}^n} \le \delta + (1-\alpha)\epsilon = \alpha r \\ \implies\quad & z' \in \{\alpha x' + (1-\alpha)y_\epsilon \mid x' \in \mathsf{B}^n(r, x)\}. \end{aligned}$$

Since $y_\epsilon \in C$ and $\mathsf{B}^n(r, x) \subseteq C$ it follows that $z' \in C$ and so $\mathsf{B}^n(\delta, z) \subseteq C$. This gives our claim that $[x, y) \subseteq \mathrm{int}(C)$. ▼

That $\mathrm{int}(C)$ is convex follows immediately since, if $x, y \in \mathrm{int}(C)$, Lemma 1 ensures that the line segment connecting x and y is contained in $\mathrm{int}(C)$.

Now consider the convex cone K with vertex x_0. We know now that $\mathrm{rel\,int}(K)$ is convex so we need only show that it is a cone. This, however, is obvious. Indeed, if $x \in \mathrm{rel\,int}(K)$ suppose that $x_0 + \lambda(x - x_0) \notin \mathrm{rel\,int}(K)$ for some $\lambda \in \mathbb{R}_{>0}$. Since $x_0 + \lambda(x - x_0) \in K$ we must then have $x_0 + \lambda(x - x_0) \in \mathrm{rel\,bd}(K)$. By Lemma 1 this means that $x_0 + (\lambda + \epsilon)(x - x_0) \notin K$ for all $\epsilon \in \mathbb{R}_{>0}$. This contradicts the fact that K is a cone. ■

The following result will also come up in our constructions.

1.9.13 Proposition (The closure of the relative interior) *If* $C \subseteq \mathbb{R}^n$ *is a convex set then* $\mathrm{cl}(\mathrm{rel\,int}(C)) = \mathrm{cl}(C)$*.*

Proof It is clear that $\mathrm{cl}(\mathrm{rel\,int}(C)) \subseteq \mathrm{cl}(C)$. Let $x \in \mathrm{cl}(C)$ and let $y \in \mathrm{rel\,int}(C)$. By Lemma 1 in the proof of Proposition 1.9.12 it follows that the half-open line segment $[y, x)$ is contained in $\mathrm{rel\,int}(C)$. Therefore, there exists a sequence $(x_j)_{j \in \mathbb{Z}_{>0}}$ in this line segment, and so in $\mathrm{rel\,int}(C)$, converging to x. Thus $x \in \mathrm{cl}(\mathrm{rel\,int}(C))$. ■

We close this section by giving a useful characterisation of compact convex sets. The proof of the theorem will involve some things that will be covered only in the sequel.

1.9.14 Theorem (Compact convex sets are homeomorphic to balls) *If* $\mathrm{C} \subseteq \mathbb{R}^{\mathrm{n}}$ *has dimension* k *then*

(i) $\mathrm{cl}(\mathrm{C})$ *is homeomorphic to* $\overline{\mathsf{B}}^{\mathrm{k}}(1, \mathbf{0})$*,*

(ii) $\mathrm{rel\,int}(\mathrm{C})$ *is homeomorphic to* $\mathsf{B}^{\mathrm{k}}(1, \mathbf{0})$*,*

(iii) $\mathrm{bd}(\mathrm{C})$ *is homeomorphic to*

$$\mathbb{S}^{\mathrm{k}-1} = \{\mathbf{x} \in \mathbb{R}^{\mathrm{k}} \mid \|\mathbf{x}\|_{\mathbb{R}^{\mathrm{k}}} = 1\},$$

the unit sphere in $\mathbb{R}^{\mathrm{k}}$*.*

Proof Suppose for simplicity and without loss of generality that $n = k$. Let $x_0 \in \mathrm{int}(C)$ and for $u \in \mathbb{S}^{n-1}$ let $\rho_{x_0,u} \colon \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ be defined by

$$\rho_{x_0,u}(s) = x_0 + su.$$

Since $x_0 \in \int(C)$, for s sufficiently small we have $\rho_{x_0,u}(s) \in \mathrm{int}(C)$. Define

$$s(x_0, u) = \sup\{s \in \mathbb{R}_{\geq 0} \mid \rho_{x_0,u}(s) \in \mathrm{rel\,int}(C)\},$$

this making sense by compactness of C and continuity of $\rho_{x,u}$. We claim that $\rho_{x_0,u}(s) \notin \mathrm{cl}(C)$ for $s > s(x_0, u)$. By Corollary 1.9.17 there exists a separating hyperplane P for $\{\rho_{x_0,u}(s(x_0, u)\}$ and C. We write

$$\mathsf{P} = \{x \in \mathbb{R}^n \mid \langle \lambda, x\rangle_{\mathbb{R}^n} = a\},$$

and assume that

$$C \subseteq \{x \in \mathbb{R}^n \mid \langle \lambda, x\rangle_{\mathbb{R}^n} \leq a\}.$$

Then

$$\rho_{x_0,u}(s) \in \mathrm{cl}(C) \quad \Longleftrightarrow \quad \langle \lambda, \rho_{x_0,u}(s)\rangle_{\mathbb{R}^n} \leq a \quad \Longleftrightarrow \quad s \leq s(x_0, u),$$

the latter implication following since the function $s \mapsto \langle \lambda, \rho_{x_0,u}(s)\rangle_{\mathbb{R}^n}$ is linear. Therefore, it follows that $\rho_{x_0,u}(s) \in \mathrm{bd}(C)$ if and only if $s = s(x_0, u)$. This, then defines a map $\psi_C \colon \mathbb{S}^{n-1} \to \mathrm{bd}(C)$ by $\psi_C(u) = \rho_{x_0,u}(s(x_0, u))$. Our arguments above directly imply that ψ_C is injective. To see that ψ_C is surjective, let $x \in \mathrm{bd}(C)$ and define $u_x = x_0 + \frac{x - x_0}{\|x - x_0\|_{\mathbb{R}^n}}$. Then $\rho_{x_0,u_x}(s) = x$ if we take $s = \|x - x_0\|_{\mathbb{R}^n}$. This not only establishes the surjectivity of ψ_C but gives an explicit formula for $\psi_C^{-1} \colon \mathrm{bd}(C) \to \mathbb{S}^{n-1}$:

$$\psi_C^{-1}(x) = \frac{x - x_0}{\|x - x_0\|_{\mathbb{R}^n}}.$$

Note that this is the restriction to $\mathrm{bd}(C)$ of the continuous map $x \mapsto \frac{x - x_0}{\|x - x_0\|_{\mathbb{R}^n}}$ from $\mathbb{R}^n \setminus \{x_0\}$ to $\mathbb{S}^{n-1}$. Thus ψ_C^{-1} is continuous by Proposition 1.3.24. It now follows from Theorem 1.3.43 that ψ_C^{-1} is, in fact, a homeomorphism and so too must ψ_C be a homeomorphism. This gives part (iii) of the theorem.

To prove the remaining two parts of the theorem we use a lemma.

1 Lemma *If* $\mathrm{C} \subseteq \mathbb{R}^{\mathrm{n}}$ *is an* n*-dimensional convex set and if* $\mathbf{x}_0 \in \mathrm{rel\,int}(\mathrm{C})$*, then*

$$\mathrm{cl}(\mathrm{C}) = \{(1-\mathrm{s})\mathbf{x}_0 + \mathrm{s}\mathbf{x}' \mid \mathrm{s} \in [0,1],\ \mathbf{x}' \in \mathrm{bd}(\mathrm{C})\}.$$

Moreover, if $\mathbf{x}'_1, \mathbf{x}'_2 \in \mathrm{bd}(\mathrm{C})$ *are distinct then*

$$\{(1-\mathrm{s})\mathbf{x}_0 + \mathrm{s}\mathbf{x}'_1 \mid \mathrm{s} \in [0,1]\} \cap \{(1-\mathrm{s})\mathbf{x}_0 + \mathrm{s}\mathbf{x}'_2 \mid \mathrm{s} \in [0,1]\} = \{\mathbf{x}_0\}.$$

Proof Convexity of $\mathrm{cl}(C)$ ensures that

$$\{(1-s)x_0 + sx' \mid s \in [0,1],\ x' \in \mathrm{bd}(C)\} \subseteq \mathrm{cl}(C).$$

For the converse inclusion, let $x \in \mathrm{cl}(C)$. If $x = x_0$ or it $x \in \mathrm{bd}(C)$ then we obviously have

$$x \in \{(1-s)x_0 + sx' \mid s \in [0,1],\ x' \in \mathrm{bd}(C)\}.$$

Thus we suppose that $x \in \mathrm{int}(C)$ and $x \neq x_0$. Similarly to the first part of the proof, let $v = x - x_0$ and define $\rho_{x_0,v}(t) = x_0 + tv$. For $t \in \mathbb{R}_{\geq 0}$ sufficiently small we have $\rho_{x_0,v}(t) \in \mathrm{int}(C)$ since $x_0 \in \mathrm{int}(C)$. Since C is compact we can define

$$t_0 = \sup\{t \in \mathbb{R}_{\geq 0} \mid \rho_{x_0,v}(t) \in \mathrm{int}(C)\}.$$

We then take $x' = x_0 + t_0(x - x_0)$ so that $x' \in \mathrm{bd}(C)$, just as we showed in the first part of the proof. Since $\rho_{x_0,v}(1) = x$ and since $x \in \mathrm{int}(C)$ we have $t_0 > 1$. Thus we write

$$x = (1 - t_0^{-1})y + t_0^{-1}x'$$

which gives

$$\mathrm{cl}(C) \subseteq \{(1-s)y + sx' \mid x' \in \mathrm{bd}(C),\ s \in [0,1]\}$$

since $t_0^{-1} < 1$.

Our arguments in the first part of the proof show that the ray

$$\{\rho_{x_0,v}(t) \mid t \geq 0\}$$

intersects $\mathrm{bd}(C)$ in exactly one point. From this it follows that two such rays passing through distinct boundary points of C can only intersect at x_0. ▼

Now let $x \in \mathrm{cl}(C)$. Suppose that $x \neq x_0$. Then, by the lemma, there exists a unique $x' \in \mathrm{bd}(C)$ and $s \in (0,1]$ such that $x = x_0 + s(x' - x_0)$. Moreover,

$$s = \frac{\|x - x_0\|_{\mathbb{R}^n}}{\|x' - x_0\|_{\mathbb{R}^n}}.$$

Thus we define $\phi_C\colon \mathrm{cl}(C) \to \overline{\mathsf{B}}^n(1, \mathbf{0})$ by

$$\phi_C(x) = \begin{cases} \frac{\|x-x_0\|_{\mathbb{R}^n}}{\|x'-x_0\|_{\mathbb{R}^n}} \psi_C^{-1}(x'), & x \neq x_0, \\ \mathbf{0}, & x = x_0. \end{cases}$$

Continuity of ϕ_C away from x_0 will follow if we can show that the boundary point x' is a continuous function of x. However, this follows since $x' = \psi_C(\frac{x-x_0}{\|x-x_0\|_{\mathbb{R}^n}})$, which is continuous away from x_0. To show continuity at x_0, define

$$m = \min\{\|x' - x_0\|_{\mathbb{R}^n} \mid x' \in \mathrm{bd}(C)\},$$

noting that $m \in \mathbb{R}_{>0}$ since $\mathrm{bd}(C)$ is compact (using Theorem 1.3.32). Now, if $\epsilon \in \mathbb{R}_{>0}$ let $\delta = m\epsilon$ and then note that if $x \in C \cap \mathsf{B}^n(\delta, x_0)$ we have

$$\|\boldsymbol{\phi}_C(\boldsymbol{x}) - \boldsymbol{\phi}_C(\boldsymbol{x}_0)\|_{\mathbb{R}^n} = \frac{\|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n}}{\|\boldsymbol{x}' - \boldsymbol{x}_0\|_{\mathbb{R}^n}} \|\boldsymbol{\psi}_C^{-1}(\boldsymbol{x}')\|_{\mathbb{R}^n} < \tfrac{\delta}{m} = \epsilon.$$

Thus $\boldsymbol{\phi}_C$ is continuous. Injectivity of $\boldsymbol{\phi}_C$ follows from the uniqueness part of the lemma above, and that $\boldsymbol{\phi}_C$ is a homeomorphism now follows from Theorem 1.3.43. ■

In Figure 1.14 we illustrate the idea behind the preceding theorem.

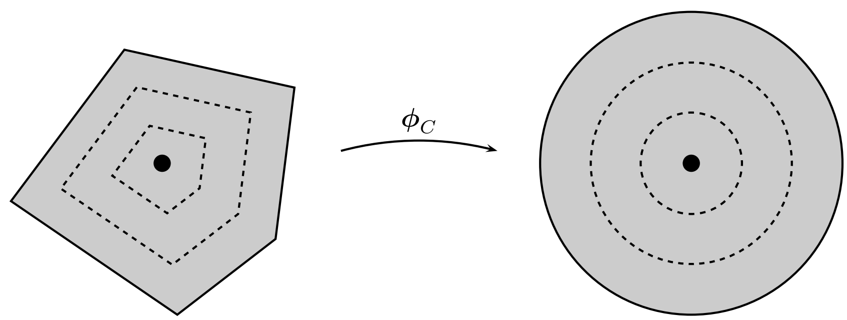

Figure 1.14 Compact convex sets are homeomorphic to balls

1.9.4 Separation theorems for convex sets

One of the most important properties of convex sets in convex analysis is the notion of certain types of convex sets being separated by hyperplanes. We shall give only an introduction to this very important topic, and refer the reader to the notes at the end of the chapter for additional references.

In order to make things clear, let us define all of our terminology precisely.

1.9.15 Definition (Hyperplane, half-space, support hyperplane)

(i) A ***hyperplane*** in $\mathbb{R}^n$ is a subset of the form

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{\lambda}, \boldsymbol{x} \rangle_{\mathbb{R}^n} = a\}$$

for some $\boldsymbol{\lambda} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and $a \in \mathbb{R}$. Such a hyperplane is denoted by $\mathsf{P}_{\lambda,a}$.

(ii) An ***half-space*** in $\mathbb{R}^n$ is a subset of the form

$$\{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{\lambda}, \boldsymbol{x} \rangle_{\mathbb{R}^n} > a\}$$

for some $\boldsymbol{\lambda} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ and $a \in \mathbb{R}$. We shall denote

$$H^-_{\lambda,a} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{\lambda}, \boldsymbol{x} \rangle_{\mathbb{R}^n} < a\}, \quad H^+_{\lambda,a} = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{\lambda}, \boldsymbol{x} \rangle_{\mathbb{R}^n} > a\}.$$

(iii) If $A \subseteq \mathbb{R}^n$, a ***support hyperplane*** for A is a hyperplane $\mathsf{P}_{\lambda,a}$ such that $A \subseteq H^+_{\lambda,a} \cup \mathsf{P}_{\lambda,a}$.

(iv) For subsets $A, B \subseteq \mathbb{R}^n$, a ***separating hyperplane*** is a hyperplane $\mathsf{P}_{\lambda,a}$ for which

$$A \subseteq H^+_{\lambda,a} \cup \mathsf{P}_{\lambda,a}, \; B \subseteq H^-_{\lambda,a} \cup \mathsf{P}_{\lambda,a}. \qquad \bullet$$

The following result is a basis for many separation theorems for convex sets.

1.9.16 Theorem (Convex sets possess supporting hyperplanes) *If $\mathsf{C} \subseteq \mathbb{R}^n$ is a convex set not equal to $\mathbb{R}^n$, then C possesses a supporting hyperplane.*

Proof Let $x_0 \notin \mathrm{cl}(C)$, let $z \in C$, and define $r = \|x_0 - z\|_{\mathbb{R}^n}$. Define $A = \mathrm{cl}(C) \cap \mathsf{B}^n(r, x_0)$ noting that A is a nonempty compact set. Define $f : A \to \mathbb{R}_{>0}$ by $f(y) = \|x_0 - y\|_{\mathbb{R}^n}$. The map f is continuous and so there exists $y_0 \in A \subseteq \mathrm{cl}(C)$ such that $f(y_0)$ is the minimum value of f. Let $\lambda = y_0 - x_0$ and $a = \langle y_0, y_0 - x_0\rangle_{\mathbb{R}^n}$. We will show that $\mathsf{P}_{\lambda,a}$ is a support hyperplane for C.

First let us show that $\mathsf{P}_{\lambda,a}$ separates $\{x_0\}$ and $\mathrm{cl}(C)$. A direct computation shows that

$$\langle \lambda, x_0\rangle_{\mathbb{R}^n} = -\|x_0 - y_0\|^2_{\mathbb{R}^n} + a < a.$$

To show that $\langle \lambda, x\rangle_{\mathbb{R}^n} \geq a$ for all $x \in \mathrm{cl}(C)$, suppose otherwise. Thus let $x \in C$ be such that $\langle \lambda, x\rangle_{\mathbb{R}^n} < a$. By Lemma 1 in the proof of Proposition 1.9.12 the line segment from y to y_0 is contained in $\mathrm{cl}(C)$. Define $g : [0,1] \to \mathbb{R}$ by $g(s) = \|(1-s)y_0 + sy - x_0\|^2_{\mathbb{R}^n}$. Thus g is the square of the distance from x_0 to points on the line segment from y to y_0. Note that $g(s) \geq g(0)$ for all $s \in (0,1]$ since y_0 is the closest point in $\mathrm{cl}(C)$ to x_0. A computation gives

$$g(s) = (1-s)^2\|y_0 - x_0\|^2_{\mathbb{R}^n} + 2s(1-s)\langle y - x_0, y_0 - x_0\rangle_{\mathbb{R}^n} + s^2\|y - x_0\|^2_{\mathbb{R}^n}$$

and another computation gives $g'(0) = 2(\langle \lambda, y\rangle_{\mathbb{R}^n} - a)$ which is strictly negative by our assumption about y. This means that g strictly decreases near zero, which contradicts the definition of y_0. Thus we must have $\langle \lambda, y\rangle_{\mathbb{R}^n} \geq a$ for all $y \in \mathrm{cl}(C)$. ■

During the course of the proof of the theorem we almost proved the following result.

1.9.17 Corollary (Separation of convex sets and points) *If $\mathsf{C} \subseteq \mathbb{R}^n$ is convex and if $\mathsf{x}_0 \notin \mathrm{int}(\mathsf{C})$ then there exists a separating hyperplane P for $\{\mathsf{x}_0\}$ and C. Moreover, if $\mathsf{x}_0 \notin \mathrm{cl}(\mathsf{C})$ then P may be chosen such that $\mathsf{x}_0 \notin \mathsf{P}$ and such that $\mathsf{P} \cap \mathsf{C} = \varnothing$.*

Proof If $x_0 \notin \mathrm{cl}(C)$ then, from the proof of Theorem 1.9.16, there exists $\lambda \in \mathbb{R}^n$ and $a \in \mathbb{R}$ such that $\langle \lambda, x_0\rangle_{\mathbb{R}^n} < a$ and such that $\langle \lambda, x\rangle_{\mathbb{R}^n} \geq a$ for all $x \in C$. By continuity of the inner product (as shown in ?) it follows that $\langle \lambda, x\rangle_{\mathbb{R}^n} \geq a$ for every $x \in \mathrm{cl}(C)$. Therefore, there exists $\epsilon \in \mathbb{R}_{>0}$ such that what

$$\langle \lambda, x_0\rangle_{\mathbb{R}^n} < a + \epsilon, \quad \langle \lambda, x\rangle_{\mathbb{R}^n} > a - \epsilon, \qquad x \in \mathrm{cl}(C).$$

Thus $x_0 \notin \mathsf{P}_{\lambda,a}$ and $\mathsf{P}_{\lambda,a} \cap \mathrm{cl}(C) = \varnothing$. Note that this, at the same time, gives the first statement in the corollary when $x_0 \in \mathrm{cl}(C)$ and the final statement in the corollary.

If $x_0 \in \mathrm{bd}(C)$ then let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}^n \setminus \mathrm{cl}(C)$ converging to x_0. For each $j \in \mathbb{Z}_{>0}$ let $\lambda_j \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and $a_j \in \mathbb{R}$ have the property that

$$\begin{aligned} \langle \lambda_j, x_j\rangle_{\mathbb{R}^n} &\leq a_j, \qquad j \in \mathbb{Z}_{>0}, \\ \langle \lambda_j, y\rangle_{\mathbb{R}^n} &> a_j, \qquad y \in C,\; j \in \mathbb{Z}_{>0}. \end{aligned}$$

(That it is possible to find such sequences is a consequence of the continuity of an inner product, as proved in .) Let us without loss of generality take $a_j = \langle \boldsymbol{\lambda}_j, \boldsymbol{x}_j \rangle_{\mathbb{R}^n}$; this corresponds to choosing the hyperplane separating C from $\boldsymbol{x}_j$ to pass through $\boldsymbol{x}_j$. Let $\boldsymbol{\alpha}_j = \frac{\boldsymbol{\lambda}_j}{\|\boldsymbol{\lambda}_j\|_{\mathbb{R}^n}}$, $j \in \mathbb{Z}_{>0}$. The sequence $(\boldsymbol{\alpha}_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence in the $(n-1)$-sphere which is compact. Thus we can choose a convergent subsequence which we also denote, by an abuse of notation, by $(\boldsymbol{\alpha}_j)_{j\in\mathbb{Z}_{>0}}$. Let $\boldsymbol{\alpha} \in \mathbb{R}^n$ denote the limit of this sequence. Defining $c_j = \langle \boldsymbol{\alpha}_j, \boldsymbol{x}_j \rangle_{\mathbb{R}^n}$ we then have

where?

$$\begin{aligned} \langle \boldsymbol{\alpha}_j, \boldsymbol{x}_j \rangle_{\mathbb{R}^n} &= c_j, \qquad j \in \mathbb{Z}_{>0}, \\ \langle \boldsymbol{\alpha}_j, \boldsymbol{y} \rangle_{\mathbb{R}^n} &> c_j, \qquad \boldsymbol{y} \in C,\ j \in \mathbb{Z}_{>0}. \end{aligned}$$

Let $c = \lim_{j\to\infty} c_j$. For $\boldsymbol{y} \in C$ this gives

$$\begin{aligned} \langle \boldsymbol{\alpha}, \boldsymbol{x}_0 \rangle_{\mathbb{R}^n} &= \lim_{j\to\infty} \langle \boldsymbol{\alpha}_j, \boldsymbol{x}_j \rangle_{\mathbb{R}^n} = c, \\ \langle \boldsymbol{\alpha}, \boldsymbol{y} \rangle_{\mathbb{R}^n} &= \lim_{j\to\infty} \langle \boldsymbol{\alpha}_j, \boldsymbol{y} \rangle_{\mathbb{R}^n} \geq c \end{aligned}$$

(again using continuity of the inner product), as desired. ■

The following consequence of Theorem 1.9.16 is also of independent interest.

1.9.18 Corollary (Disjoint convex sets are separated) *If* $\mathsf{C}_1, \mathsf{C}_2 \subseteq \mathbb{R}^n$ *are disjoint convex sets, then there exists a hyperplane separating* C_1 *and* C_2.

Proof Define

$$C_1 - C_2 = \{\boldsymbol{x}_1 - \boldsymbol{x}_2 \mid \boldsymbol{x}_1 \in C_1,\ \boldsymbol{x}_2 \in C_2\}.$$

One checks directly that $C_1 - C_2$ is convex. Since C_1 and C_2 are disjoint it follows that $\boldsymbol{0} \notin C_1 - C_2$. By Theorem 1.9.16 there exists a hyperplane P, passing through $\boldsymbol{0}$, separating $C_1 - C_2$ from $\boldsymbol{0}$. We claim that this implies that the same hyperplane P, appropriately translated, separates C_1 and C_2. To see this note that P gives rise to $\boldsymbol{\lambda} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that

$$\langle \boldsymbol{\lambda}, \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle_{\mathbb{R}^n} \geq 0, \qquad \boldsymbol{x}_1 \in C_1,\ \boldsymbol{x}_2 \in C_2.$$

Let

$$a_1 = \inf\{\langle \boldsymbol{\lambda}, \boldsymbol{x}_1 \rangle_{\mathbb{R}^n} \mid \boldsymbol{x}_1 \in C_1\}, \quad a_2 = \sup\{\langle \boldsymbol{\lambda}, \boldsymbol{x}_2 \rangle_{\mathbb{R}^n} \mid \boldsymbol{x}_2 \in C_2\}$$

so that $a_1 - a_2 \geq 0$. For any $a \in [a_2, a_1]$ we have

$$\begin{aligned} \langle \boldsymbol{\lambda}, \boldsymbol{x}_1 \rangle_{\mathbb{R}^n} &\geq a, \qquad \boldsymbol{x}_1 \in C_1, \\ \langle \boldsymbol{\lambda}, \boldsymbol{x}_2 \rangle_{\mathbb{R}^n} &\leq a, \qquad \boldsymbol{x}_2 \in C_2, \end{aligned}$$

giving the separation of C_1 and C_2, as desired. ■

The following result is now a fairly general separation theorem for convex sets.

1.9.19 Theorem (A general separation theorem) *If* $\mathrm{C}_1, \mathrm{C}_2 \subseteq \mathbb{R}^{\mathrm{n}}$ *are convex sets, then they possess a separating hyperplane if and only if either of the following two conditions holds:*

(i) there exists a hyperplane P *such that* $\mathrm{C}_1, \mathrm{C}_2 \subseteq \mathsf{P}$*;*

(ii) $\operatorname{rel\,int}(\mathrm{C}_1) \cap \operatorname{rel\,int}(\mathrm{C}_2) = \varnothing$.

Proof Suppose that C_1 and C_2 possess a separating hyperplane P. Therefore, there exists $\boldsymbol{\lambda} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and $a \in \mathbb{R}$ such that

$$\begin{aligned} \langle \boldsymbol{\lambda}, \boldsymbol{x}_1 \rangle_{\mathbb{R}^n} &\geq a, \qquad \boldsymbol{x}_1 \in C_1, \\ \langle \boldsymbol{\lambda}, \boldsymbol{x}_2 \rangle_{\mathbb{R}^n} &\leq a, \qquad \boldsymbol{x}_2 \in C_2. \end{aligned}$$

If $\langle \boldsymbol{\lambda}, \boldsymbol{x} \rangle_{\mathbb{R}^n} = a$ for all $\boldsymbol{x} \in C_1 \cup C_2$ then (i) holds. Now suppose that $\langle \boldsymbol{\lambda}, \boldsymbol{x}_1 \rangle_{\mathbb{R}^n} > a$ for some $\boldsymbol{x}_1 \in C_1$ (a similar argument will obviously apply if this holds for some $\boldsymbol{x}_2 \in C_2$) and let $\boldsymbol{x}_0 \in \operatorname{rel\,int}(C_1)$. Since P is a support hyperplane for C_1 and since $C_1 \not\subset \mathsf{P}$, it follows that the relative interior, and so $\boldsymbol{x}_0$, lies in the appropriate half-space defined by P. Since P separates C_1 and C_2 this precludes $\boldsymbol{x}_0$ from being in C_2. Thus (ii) holds.

Now suppose that (i) holds. It is then clear that P is a separating hyperplane for C_1 and C_2.

Finally, suppose that (ii) holds. From Proposition 1.9.12 and Corollary 1.9.18 it holds that $\operatorname{rel\,int}(C_1)$ and $\operatorname{rel\,int}(C_2)$ possess a separating hyperplane. Thus there exists $\boldsymbol{\lambda} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ and $a \in \mathbb{R}$ such that

$$\begin{aligned} \langle \boldsymbol{\lambda}, \boldsymbol{x}_1 \rangle_{\mathbb{R}^n} &\leq a, \qquad \boldsymbol{x}_1 \in \operatorname{rel\,int}(C_1), \\ \langle \boldsymbol{\lambda}, \boldsymbol{x}_2 \rangle_{\mathbb{R}^n} &\geq a, \qquad \boldsymbol{x}_2 \in \operatorname{rel\,int}(C_2). \end{aligned}$$

Therefore, by Proposition 1.9.13 we also have

$$\begin{aligned} \langle \boldsymbol{\lambda}, \boldsymbol{x}_1 \rangle_{\mathbb{R}^n} &\leq a, \qquad \boldsymbol{x}_1 \in \operatorname{cl}(C_1), \\ \langle \boldsymbol{\lambda}, \boldsymbol{x}_2 \rangle_{\mathbb{R}^n} &\geq a, \qquad \boldsymbol{x}_2 \in \operatorname{cl}(C_2), \end{aligned}$$

which implies this part of the theorem. ■

1.9.5 Simplices and simplex cones

We now concern ourselves with special examples of convex sets and convex cones, and show that these special objects can always be found as neighbourhoods in general convex sets and cones.

We begin with the definitions.

1.9.20 Definition (Affine independence, simplex, simplex cone) Let $n \in \mathbb{Z}_{>0}$.

(i) A set $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\} \subseteq \mathbb{R}^n$ is ***affinely independent*** if the set $\{\boldsymbol{x}_1 - \boldsymbol{x}_0, \dots, \boldsymbol{x}_k - \boldsymbol{x}_0\}$ is linearly independent.

(ii) A **k-*simplex*** is the convex hull of a set $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ of affinely independent points. We shall denote this simplex by $\Delta(\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$.

(iii) A **k-*simplex cone*** with vertex $\boldsymbol{x}_0$ is the coned convex hull with vertex $\boldsymbol{x}_0$ of a set $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ such that $\{\boldsymbol{x}_1 - \boldsymbol{x}_0, \dots, \boldsymbol{x}_k - \boldsymbol{x}_0\}$ is linearly independent. We shall denote this simplex cone by $K(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k; \boldsymbol{x}_0)$. •

The following result shows that choice of $\boldsymbol{x}_0$ from $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ is not important in the definition affine independence.

1.9.21 Proposition (Characterisation of affine independence) *For vectors* $\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k\} \subseteq \mathbb{R}^n$ *the following statements are equivalent:*

(i) the set is affinely independent;

(ii) for any $j \in \{1, \dots, n\}$*, the set* $\{\mathbf{x}_0 - \mathbf{x}_j, \mathbf{x}_1 - \mathbf{x}_j, \dots, \mathbf{x}_{j-1} - \mathbf{x}_j, \mathbf{x}_{j+1} - \mathbf{x}_j, \dots, \mathbf{x}_k - \mathbf{x}_j\}$ *is linearly independent.*

Proof Suppose that the set is affinely independent and let $j \in \{1, \dots, k\}$. Let

$$c_0(\boldsymbol{x}_0 - \boldsymbol{x}_j) + c_1(\boldsymbol{x}_1 - \boldsymbol{x}_j) + \dots + c_{j-1}(\boldsymbol{x}_{j-1} - \boldsymbol{x}_j) + c_{j+1}(\boldsymbol{x}_{j+1} - \boldsymbol{x}_j) + \dots + c_k(\boldsymbol{x}_k - \boldsymbol{x}_j) = \boldsymbol{0}$$

for $c_0, c_1, \dots, c_{j-1}, c_{j+1}, \dots, c_k \in \mathbb{R}$. Adding and subtracting $\boldsymbol{x}_0$ to each of the terms in the sum yields

$$c_1(\boldsymbol{x}_1 - \boldsymbol{x}_0 + \dots + c_{j-1}(\boldsymbol{x}_{j-1} - \boldsymbol{x}_0) + c_{j+1}(\boldsymbol{x}_{j+1} - \boldsymbol{x}_0 + \dots + c_k(\boldsymbol{x}_k - \boldsymbol{x}_0) - (c_0 + c_1 + \dots + c_{j-1} + c_{j+1} + \dots + c_k)(\boldsymbol{x}_j - \boldsymbol{x}_0) = \boldsymbol{0}.$$

Affine independence of $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ gives

$$\begin{aligned} c_1 &= \dots = c_{j-1} = c_{j+1} = \dots = c_k = 0, \\ c_0 &+ c_1 + \dots + c_{j-1} + c_{j+1} + \dots + c_k = 0, \end{aligned}$$

which gives $c_0, c_1, \dots, c_{j-1}, c_{j+1}, \dots, c_k$ all zero and so gives the second statement.

The converse is proved in exactly the same way, swapping the rôles of $\boldsymbol{x}_0$ and $\boldsymbol{x}_j$. ■

The idea of a simplex and a simplex cone is that they form the "simplest" possible examples of convex sets and convex cones. The notion of "simplicity" considered here is the following: A convex set C_1 (resp. convex cone K_1) is simpler than a convex set C_2 (resp. a convex cone K_2) if the minimum number of points needed to specify the convex hull (resp. coned convex hull) of C_1 (resp. K_1) is less than the minimum number of points needed to specify the convex hull (resp. coned convex hull) of C_2 (resp. K_2).

Let us give the standard examples of such objects.

1.9.22 Examples (Standard n-simplex, standard n-simplex cone)

1. The ***standard*** $\mathbf{n}$***-simplex*** is the subset of $\mathbb{R}^n$ given by

$$\Delta_n = \left\{\boldsymbol{x} \in \mathbb{R}^n \;\middle|\; x_1, \dots, x_n \in \mathbb{R}_{\geq 0}, \ \sum_{j=1}^{n} x_j \leq 1\right\}.$$

Thus Δ_n is the convex hull of the n standard basis vectors along with the origin.

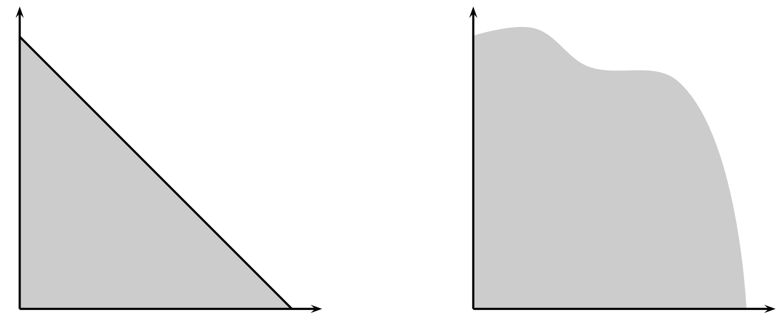

Figure 1.15 The standard 2-simplex (left) and the standard 2-simplex cone (right)

2. The ***standard* n*-simplex cone*** is the subset of $\mathbb{R}^n$ given by

$$K_n = \{\boldsymbol{x} \in \mathbb{R}^n \mid x_1, \dots, x_n \in \mathbb{R}_{\geq 0}\}.$$

Note that K_n is the coned convex hull with vertex at $\mathbf{0}$ of the n standard basis vectors.

In Figure 1.15 we depict the standard n-simplex and the standard n-simplex cone when $n = 2$. •

The following result about the dimension of simplices and simplex cones is intuitively clear.

1.9.23 Proposition (Dimension of simplices and simplex cones) *If* $\Delta, \mathrm{K} \subseteq \mathbb{R}^n$ *are a* k*-simplex and a* k*-simplex cone with vertex at* $\mathbf{x}_0$*, respectively, then* $\dim(\Delta) = \dim(\mathrm{K}) = \mathrm{k}$.

Proof Let us first consider the k-simplex $\Delta = \Delta(\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$. Clearly $\mathrm{aff}(\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}) \subseteq \mathrm{aff}(\Delta)$ since $\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\} \subseteq \Delta$. Let $\boldsymbol{x} \in \mathrm{aff}(\Delta)$ so that

$$\boldsymbol{x} = \sum_{l=1}^{m} \mu_l \boldsymbol{y}_l$$

for $\mu_1, \dots, \mu_m \in \mathbb{R}$ summing to 1 and for $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \in \Delta$. For each $l \in \{1, \dots, m\}$ we have

$$\boldsymbol{y}_l = \sum_{j=0}^{k} \lambda_{lj} \boldsymbol{x}_j$$

for $\lambda_{l0}, \lambda_{l1}, \dots, \lambda_{lk} \in \mathbb{R}_{\geq 0}$ summing to 1. Therefore,

$$\boldsymbol{x} = \sum_{l=1}^{m} \sum_{j=0}^{k} \mu_j \lambda_{lj} \boldsymbol{x}_j = \sum_{j=0}^{k} \Big(\sum_{l=1}^{m} \lambda_{lj} \mu_l \Big) \boldsymbol{x}_j,$$

and so $\boldsymbol{x} \in \mathrm{aff}(\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\})$ since

$$\sum_{j=0}^{k} \Big(\sum_{l=1}^{m} \lambda_{lj} \mu_l \Big) = 1.$$

Thus $\mathrm{aff}(\Delta) = \mathrm{aff}(\{x_0, x_1, \dots, x_k\})$. That $\dim(\Delta) = k$ follows since the subspace corresponding to the affine subspace $\mathrm{aff}(\{x_0, x_1, \dots, x_k\})$ is generated by $\{x_1 - x_0, \dots, x_k - x_0\}$, and this subspace has dimension k.

The proof for the k-simplex cone K follows in an entirely similar manner, merely with convex combinations being replaced by coned convex combinations. ■

One of the things we will need to be able to do is find neighbourhoods of points in convex sets and convex cones that are simplices and simplex cones, respectively. For convex sets we have the following result.

1.9.24 Proposition (Existence of simplicial neighbourhoods) *Let $C \subseteq \mathbb{R}^n$ be convex and of dimension k, let $x_0 \in \mathrm{rel\,int}(C)$, and let U be a neighbourhood of x_0 in $\mathbb{R}^n$. Then there exists a k-simplex $\Delta \subseteq C$ such that $\Delta \subseteq U$ and $x_0 \in \mathrm{rel\,int}(\Delta)$.*

Proof Let $r \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^n(r, x_0) \subseteq U$ and such that $\mathsf{B}^n(r, x_0) \cap \mathrm{aff}(C) \subseteq C$. The existence of such an r follows since $x_0 \in \mathrm{rel\,int}(C)$. Let $\{v_1, \dots, v_k\}$ be an orthogonal basis (see) for the subspace $\mathsf{U}(C)$ corresponding to $\mathrm{aff}(C)$ and suppose that $v_1, \dots, v_k \in \mathsf{B}^n(r, 0)$. Then $y_j \triangleq x_0 + v_j \in \mathsf{B}^n(r, x_0)$, $j \in \{1, \dots, k\}$.

what

what

We now use a linear algebra lemma; the reader can refer to for a discussion of inner product spaces.

1 Lemma *If $(\mathsf{V}, \langle\cdot,\cdot\rangle)$ is a finite-dimensional $\mathbb{R}$-inner product space and if $\{v_1, \dots, v_n\}$ is a basis for V, then there exists $v_0 \in \mathsf{V}$ such that $\langle v_0, v_j\rangle < 0$ for every $j \in \{1, \dots, n\}$.*

Proof Let $\mathsf{L}\colon \mathsf{V} \to \mathbb{R}^n$ be the unique linear map defined by asking that $\mathsf{L}(v_j)$ be equal to e_j, the jth standard basis vector for $\mathbb{R}^n$. Note that if we take $e_0 = (-1, \dots, -1) \in \mathbb{R}^n$ then, with respect to the standard inner product, $\langle e_0, e_j\rangle_{\mathbb{R}^n} = -1 < 0$, $j \in \{1, \dots, n\}$. Let $\alpha \in (\mathbb{R}^n)'$ correspond to e_0 under the identification of $\mathbb{R}^n$ with $(\mathbb{R}^n)'$ induced by the standard inner product (see) and take $\beta = \mathsf{L}'(\alpha)$. Then

what?

$$\beta(v_j) = \mathsf{L}'(\alpha) \cdot v_j = \alpha \cdot L(v_j) = \alpha \cdot e_j = -1$$

for $j \in \{1, \dots, n\}$. Then take v_0 to correspond to β under the identification of V' with V using the inner product on V (again, see). We then have $\langle v_0, v_j\rangle = -1$, $j \in \{1, \dots, n\}$. ▼

what

We now apply the lemma to the subspace $\mathsf{U}(C)$ (with the inner product induced from that on $\mathbb{R}^n$) to assert the existence of $v_0 \in \mathsf{U}(C)$ such that $\langle v_0, v_j\rangle_{\mathbb{R}^n} < 0$ for $j \in \{1, \dots, k\}$. We may assume that $\|v_0\|_{\mathbb{R}^n} < r$. We claim that the set $\{v_0, v_1, \dots, v_k\}$ is affinely independent. Indeed, suppose that

$$c_1(v_1 - v_0) + \dots + c_k(v_k - v_0) = 0.$$

Then $c_j(\langle v_j, v_j\rangle_{\mathbb{R}^n} - \langle v_j, v_0\rangle_{\mathbb{R}^n}) = 0$ for $j \in \{1, \dots, k\}$. Since $\langle v_j, v_j\rangle_{\mathbb{R}^n} - \langle v_j, v_0\rangle_{\mathbb{R}^n} \in \mathbb{R}_{>0}$ it follows that $c_j = 0$ for $j \in \{1, \dots, k\}$, so giving affine independence of $\{v_0, v_1, \dots, v_k\}$. Define $y_0 = x_0 + v_0 \in \mathsf{B}^n(r, x_0)$ and take $\Delta = \mathrm{conv}(\{y_0, y_1, \dots, y_k\})$.

We claim that $\Delta \in \mathsf{B}^n(r, x_0) \subseteq U$. Indeed, if $x \in \Delta$ then we can write x as a convex combination:

$$x = \sum_{j=0}^{k} \lambda_j y_j \quad \implies \quad x - x_0 = \sum_{j=0}^{k} \lambda_j (y_j - x_0) = \sum_{j=0}^{k} v_j.$$

Applying the triangle inequality a bunch of times gives

$$\|x - x_0\|_{\mathbb{R}^n} \leq \sum_{j=0}^{k} \lambda_j \|v_j\|_{\mathbb{R}^n} < r,$$

as desired.

Finally, we claim that $x_0 \in \operatorname{rel\,int}(\Delta)$. This will follow if we can show that $\mathbf{0} \in \operatorname{rel\,int}(\operatorname{conv}(\{v_0, v_1, \dots, v_k\}))$. By the lemma above and since we chose the basis $\{v_1, \dots, v_k\}$ to be orthogonal,

$$v_0 = \sum_{j=1}^{k} \frac{\langle v_j, v_0 \rangle_{\mathbb{R}^n}}{\|v_j\|_{\mathbb{R}^n}^2} v_j \quad \Longrightarrow \quad \|v_1\|_{\mathbb{R}^n}^2 \cdots \|v_k\|_{\mathbb{R}^n}^2 v_0 - \sum_{j=1}^{k} \langle v_j, v_0 \rangle_{\mathbb{R}^n} v_j = 0,$$

showing that $\mathbf{0}$ is a linear combination of the vectors $\{v_0, v_1, \dots, v_k\}$ with the coefficients being strictly positive. By scaling the coefficients this linear combination can be made convex with all coefficients positive. Therefore, $\mathbf{0} \in \operatorname{rel\,int}(\operatorname{conv}(\{v_0, v_1, \dots, v_k\}))$, as desired. ∎

For cones we have a similarly styled result.

1.9.25 Proposition (Existence of simplex cone neighbourhoods) *Let* $\mathsf{K} \subseteq \mathbb{R}^n$ *be a convex cone with vertex at* $\mathbf{y}_0$ *of dimension* k*, let* $\mathbf{x}_0 \in \operatorname{rel\,int}(\mathsf{K}) \setminus \{\mathbf{y}_0\}$*, and let* U *be a neighbourhood of* $\mathbf{x}_0 \in \mathbb{R}^n$*. Then there exists a* k*-simplex cone* $\mathsf{K}_0 \subseteq \mathsf{K}$ *with vertex* $\mathbf{y}_0$ *such that* $\mathsf{K}_0 \subseteq \operatorname{cone}(\mathsf{U}, \mathbf{y}_0)$ *and* $\mathbf{x}_0 \in \operatorname{rel\,int}(\mathsf{K}_0)$*.*

Proof For simplicity let us assume that $y_0 = \mathbf{0}$. This can be done without loss of generality by translating the vertex to $\mathbf{0}$, applying the result with $y_0 = \mathbf{0}$, and then translating back to y_0.

Denote by P_{x_0} the orthogonal complement to x_0 and let

$$U_{x_0} = \{v \in \mathsf{P}_{x_0} \mid x_0 + v \in U\}.$$

Note that U_{x_0} is a neighbourhood of $\mathbf{0}$ in P_{x_0}. By Proposition 1.9.24 let $\Delta \subseteq P_{x_0}$ be a $(k-1)$-simplex contained in U_{x_0} and having $\mathbf{0}$ in its relative interior. Then define K_0 to be the coned convex hull of $x_0 + \Delta \triangleq \{x_0 + v \mid v \in \Delta\}$, noting that K_0 is then the coned convex hull of the points $x_j \triangleq x_0 + v_j$, $j = \{1, \dots, k\}$, where the points $v_1, \dots, v_k$ are defined so that their convex hull is Δ.

We claim that $K_0 \subseteq \operatorname{cone}(U, \mathbf{0})$. This follows since

$$x_0 + \Delta \subseteq \{x_0 + v \mid v \in U_{x_0}\} \subseteq U,$$

and so $K_0 = \operatorname{cone}(x_0 + \Delta, \mathbf{0}) \subseteq \operatorname{cone}(U, \mathbf{0})$.

We also claim that $x_0 \in \operatorname{rel\,int}(K_0)$. Since $\mathbf{0} \in \operatorname{rel\,int}(\Delta)$ we can write

$$x_0 = x_0 + \mathbf{0} = x_0 + \sum_{j=1}^{k} \lambda_j v_j$$

for appropriate $\lambda_1, \dots, \lambda_k \in \mathbb{R}_{>0}$ summing to 1. Therefore

$$x_0 = \sum_{j=1}^{k} \lambda_j (v_j + x_0),$$

and so x_0 is a linear combination of the points $x_1, \dots, x_k$ with strictly positive coefficients. Thus $x_0 \in \operatorname{rel\,int}(K_0)$. ■

A useful property of simplices is that they define affine functions by knowing the values of the function on vertices.

1.9.26 Proposition (Affine functions on simplices) *Consider the* k*-simplex* $\Delta(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k)$ *and let* $a_0, a_1, \dots, a_k \in \mathbb{R}$. *Then there exists* $\boldsymbol{\alpha} \in (\mathbb{R}^n)'$ *and* $\beta \in \mathbb{R}$ *such that* $\boldsymbol{\alpha} \cdot \mathbf{x}_j + \beta = a_j$, $j \in \{0, 1, \dots, k\}$. *Moreover, if* $\boldsymbol{\alpha}' \in (\mathbb{R}^n)'$ *and* $\beta' \in \mathbb{R}$ *define an affine function*

$$\mathbf{x} \mapsto \boldsymbol{\alpha}' \cdot \mathbf{x} + \beta',$$

such that $\boldsymbol{\alpha} \cdot \mathbf{x}_j + \beta = a_j$, $j \in \{0, 1, \dots, k\}$, *then*

$$\boldsymbol{\alpha}' \cdot \mathbf{x} + \beta' = \boldsymbol{\alpha} \cdot \mathbf{x} + \beta$$

for all $\mathbf{x} \in \operatorname{aff}(\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k\})$.

Proof Define the linearly independent set of vectors $\{v_1, \dots, v_k\}$ by $v_j = x_j - x_0$, $j \in \{1, \dots, k\}$. If necessary, extend these vectors to a basis $\{v_1, \dots, v_k, v_{k+1}, \dots, v_n\}$. Let $\{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n\}$ be the dual basis. Then write

$$\boldsymbol{\alpha} = c_1 \boldsymbol{\alpha}_1 + \dots + c_n \boldsymbol{\alpha}_n$$

for $c_1, \dots, c_n \in \mathbb{R}$.

For the existence assertion of the result, take $c_j = a_j - a_0$, $j \in \{1, \dots, n\}$, and take $\beta = a_0 - \boldsymbol{\alpha} \cdot x_0$. Then compute

$$\boldsymbol{\alpha} \cdot x_j + \beta = \boldsymbol{\alpha} \cdot (x_j - x_0) + \beta + \boldsymbol{\alpha} \cdot x_0 = a_j, \qquad j \in \{1, \dots, k\}.$$

We also immediately have $\boldsymbol{\alpha} \cdot x_0 + \beta = a_0$. This gives the existence of $\boldsymbol{\alpha} \in (\mathbb{R}^n)'$ and $\beta \in \mathbb{R}$ as desired.

Now suppose that $\boldsymbol{\alpha}' \in (\mathbb{R}^n)'$ and $\beta' \in \mathbb{R}$ satisfy $\boldsymbol{\alpha}' \cdot x_j + \beta = a_j$ for $j \in \{0, 1, \dots, k\}$. Then $\beta' = a_0 - \boldsymbol{\alpha}' \cdot x_0$. Let $x \in \operatorname{aff}(\{x_0, x_1, \dots, x_k\})$ and write

$$x = x_0 + \xi_1(x_1 - x_0) + \dots + \xi_k(x_k - x_0)$$

for some suitable $\xi_1, \dots, \xi_k \in \mathbb{R}$ (why is this possible?). Then

$$\begin{aligned}
\boldsymbol{\alpha}' \cdot x + \beta' &= \boldsymbol{\alpha}' \cdot x_0 + \xi_1(a_1 - a_0) + \dots + \xi_k(a_k - a_0) + a_0 - \boldsymbol{\alpha}' \cdot x_0 \\
&= a_0 + \xi_1(a_1 - a_0) + \dots + \xi_k(a_k - a_0) \\
&= \boldsymbol{\alpha} \cdot x_0 + \beta + \xi_1 \boldsymbol{\alpha} \cdot (x_1 - x_0) + \dots + \xi_k \boldsymbol{\alpha} \cdot (x_k - x_0) \\
&= \boldsymbol{\alpha} \cdot x + \beta,
\end{aligned}$$

giving the uniqueness of the linear function on $\operatorname{aff}(\{x_0, x_1, \dots, x_k\})$. ■

1.9.6 Barycentric coordinates

In this section we give a useful parameterisation of simplices and simplex cones. The idea is that, using the parameterisation, one makes a simplex or a simplex cone homeomorphic to the standard simplex or the standard simplex cone, respectively.

To get started we make the following observation about simplices.

1.9.27 Proposition (Property of simplexes) *For every* $\mathbf{x} \in \Delta(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k)$ *there exists unique*

$$(\lambda_0, \lambda_1, \dots, \lambda_k) \in \left\{(\mu_0, \mu_1, \dots, \mu_k) \in [0, 1]^{k+1} \;\middle|\; \sum_{j=0}^{k} \mu_j = 1\right\}$$

such that

$$\mathbf{x} = \sum_{j=0}^{k} \lambda_j \mathbf{v}_j$$

Proof This is left to the reader as Exercise 1.9.8. ■

Note that the set of λ's appearing in the linear combination in the previous proposition have the property that the point

$$\sum_{j=0}^{k} \lambda_j \boldsymbol{e}_j$$

lies in the standard k-simplex if we take the convention that $\boldsymbol{e}_0 = \mathbf{0}$. Indeed, if $\Delta = \Delta(\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$ the map $\beta \colon \Delta_k \to \Delta$ defined by

$$\beta\Big(\sum_{j=0}^{k} \lambda_j \boldsymbol{e}_j\Big) = \sum_{j=0}^{k} \lambda_j \boldsymbol{x}_j$$

defines a homeomorphism of Δ_k with Δ. This homeomorphism has a name.

1.9.28 Definition (Barycentric coordinates for a simplex) The parameterisation of $\Delta = \Delta(\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$ by the set

$$\left\{(\lambda_0, \lambda_1, \dots, \lambda_k) \in [0, 1]^{k+1} \;\middle|\; \sum_{j=0}^{k} \lambda_j = 1\right\}$$

defined by

$$(\lambda_0, \lambda_1, \dots, \lambda_k) \mapsto \sum_{j=0}^{k} \lambda_j \boldsymbol{v}_j$$

gives ***barycentric coordinates*** for Δ. The ***barycentre*** of Δ is the image under this parameterisation of the point $(\frac{1}{k+1}, \dots, \frac{1}{k+1})$, and is denoted by $\mathrm{bc}(\Delta)$. •

The most insightful interpretation of the barycentre is given in Proposition 1.9.48.

A similar construction can be made for a k-simplex cone $K = K(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k; \boldsymbol{x}_0)$ with vertex $\boldsymbol{x}_0$. We fix some nonzero vector $\boldsymbol{v}_0 \in \operatorname{rel\,int}(K) \setminus \{\boldsymbol{x}_0\}$ and let $\mathsf{P}_{\boldsymbol{v}_0}$ be the orthogonal complement to $\boldsymbol{v}_0 - \boldsymbol{x}_0$. We may suppose, without loss of generality (by scaling if necessary), that

$$\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \{\boldsymbol{v}_0 + \boldsymbol{x} \mid \boldsymbol{x} \in \mathsf{P}_{\boldsymbol{v}_0}\},$$

i.e., that the points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ lie in a plane parallel to $\mathsf{P}_{\boldsymbol{v}_0}$ passing through $\boldsymbol{v}_0$. We then define a $(k-1)$-simplex $\Delta_{\boldsymbol{v}_0} \subseteq \mathsf{P}_{\boldsymbol{v}_0}$ by asking that

$$\Delta_{\boldsymbol{v}_0} = \{\boldsymbol{x} \in \mathsf{P}_{\boldsymbol{v}_0} \mid \boldsymbol{v}_0 + \boldsymbol{x} \in K\}$$

(we leave it to the reader to check that $\Delta_{\boldsymbol{v}_0}$ is indeed a $(k-1)$-simplex). We then let $(\lambda_1, \dots, \lambda_k) \in \Delta_{k-1}$ be barycentric coordinates for $\Delta_{\boldsymbol{v}_0}$. Noting that the vertices of $\Delta_{\boldsymbol{v}_0}$ are in fact $\boldsymbol{x}_1 - \boldsymbol{v}_0, \dots, \boldsymbol{x}_k - \boldsymbol{v}_0$, this means that a point in $\Delta_{\boldsymbol{v}_0}$ has the form

$$\lambda_1(\boldsymbol{x}_1 - \boldsymbol{v}_0) + \dots + \lambda_k(\boldsymbol{x}_k - \boldsymbol{v}_0). \tag{1.49}$$

A direct computation shows that if we define

$$l(\boldsymbol{x}) = \frac{\langle \boldsymbol{x} - \boldsymbol{x}_0, \boldsymbol{v}_0 - \boldsymbol{x}_0 \rangle_{\mathbb{R}^n}}{\|\boldsymbol{v}_0 - \boldsymbol{x}_0\|_{\mathbb{R}^n}^2}$$

then

$$l(\boldsymbol{x})^{-1}(\boldsymbol{x} - \boldsymbol{x}_0) - (\boldsymbol{v}_0 - \boldsymbol{x}_0) \in \mathsf{P}_{\boldsymbol{v}_0}.$$

Thus $l(\boldsymbol{x})^{-1}(\boldsymbol{x} - \boldsymbol{x}_0) + \boldsymbol{x}_0$ lies in the translation of $\mathsf{P}_{\boldsymbol{v}_0}$ through $\boldsymbol{v}_0$. Since K is a cone with vertex $\boldsymbol{x}_0$ this implies, in fact, that if $\boldsymbol{x} \in K$ then

$$l(\boldsymbol{x})^{-1}(\boldsymbol{x} - \boldsymbol{x}_0) - (\boldsymbol{v}_0 - \boldsymbol{x}_0) \in \Delta_{\boldsymbol{v}_0},$$

and so admits an expression of the form (1.49):

$$l(\boldsymbol{x})^{-1}(\boldsymbol{x} - \boldsymbol{x}_0) - (\boldsymbol{v}_0 - \boldsymbol{x}_0) = \lambda_1(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{v}_0) + \dots + \lambda_k(\boldsymbol{x})(\boldsymbol{x}_k - \boldsymbol{v}_0).$$

Therefore, a computation gives

$$\boldsymbol{x} = \boldsymbol{x}_0 + l(\boldsymbol{x})\Big(\boldsymbol{v}_0 - \boldsymbol{x}_0 + \Big(\lambda_1(\boldsymbol{x})(\boldsymbol{x}_1 - \boldsymbol{x}_0) + \dots + \lambda_k(\boldsymbol{x})(\boldsymbol{x}_k - \boldsymbol{x}_0)\Big)\Big).$$

Based on this development we make the following definition.

1.9.29 Definition (Barycentric coordinates for a simplex cone) Let $K = K(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k; \boldsymbol{x}_0)$. For $\boldsymbol{v}_0 \in \operatorname{rel\,int}(K) \setminus \{\boldsymbol{0}\}$ let $\mathsf{P}_{\boldsymbol{v}_0}$ be the orthogonal complement of $\boldsymbol{v}_0$ and define the $(k-1)$-simplex

$$\Delta_{\boldsymbol{v}_0} = \{\boldsymbol{x} \in \mathsf{P}_{\boldsymbol{v}_0} \mid \boldsymbol{x}_0 + \boldsymbol{v}_0 + \boldsymbol{x} \in K\}.$$

The parameterisation of K by the set

$$\mathbb{R}_{\geq 0} \times \Big\{(\lambda_1, \dots, \lambda_k) \in [0,1]^k \ \Big|\ \sum_{j=1}^{k} \lambda_j = 1\Big\}$$

defined by

$$(l, (\lambda_1, \dots, \lambda_k)) \mapsto \boldsymbol{x}_0 + l\Big(\boldsymbol{v}_0 - \boldsymbol{x}_0 + \Big(\lambda_1(\boldsymbol{x}_1 - \boldsymbol{x}_0) + \dots + \lambda_k)(\boldsymbol{x}_k - \boldsymbol{x}_0)\Big)\Big)$$

gives ***barycentric coordinates*** for K. •

For the reader for whom this definition of coordinates $(l, \boldsymbol{\lambda})$ for K is not immediately clear, we give an illustration of their meaning in Figure 1.16.

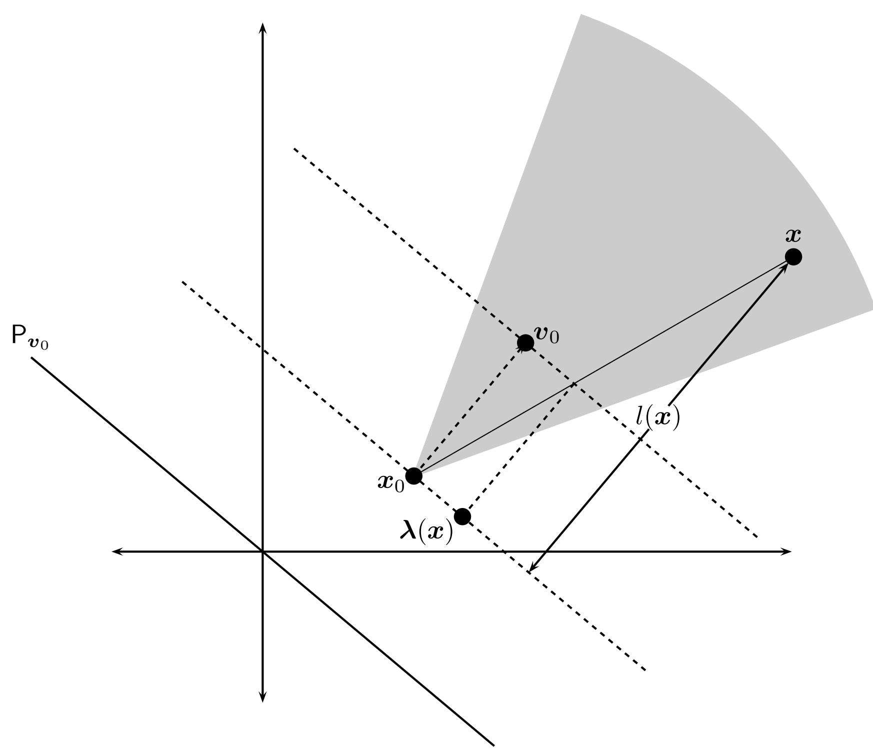

Figure 1.16 Barycentric coordinates for a simplex cone

1.9.7 Barycentric subdivision of simplices

Our next topic has to do with chopping up simplices into smaller simplices in a systematic manner. This, for reasons that are by no means obvious at this point, is an important process in studying the global topological (specifically, algebraic topological) properties of certain types of spaces. However, we shall only give the basic construction and its properties. Our only use of this construction will be in our proof of the Kakutani Fixed Point Theorem in Section 1.11.4. And since this result is presented mainly as entertainment, a reader can easily forgo the constructions we are now about to undergo.

While the idea of the barycentric subdivision is simple to understand, it is not so simple to define precisely. Indeed, many "definitions" have the character of, "Here's how to do it in a few low-dimensional cases, and the generalisation is obvious." However, if one wants to *do* anything with the notion, one must give a useful working definition. To get started we have the following notions.

1.9.30 Definition (Faces of a simplex) Consider a simplex $\Delta = \Delta(\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k)$. Let $m \in \{0, 1, \dots, k\}$ and let $J = \{j_0, j_1, \dots, j_m\}$ be a distinct subset of $\{0, 1, \dots, k\}$. The set

$$F_J = \Big\{ \sum_{l=0}^{m} \lambda_l \boldsymbol{x}_{j_l} \;\Big|\; \sum_{l=0}^{m} \lambda_l = 1,\ \lambda_0, \lambda_1, \dots, \lambda_m \in \mathbb{R}_{>0} \Big\}$$

is the **m-*dimensional face*** of Δ associated with the vertices J. The set of m-dimensional faces of Δ is denoted by $\mathscr{F}_m(\Delta)$. •

The following result records some more or less obvious statements about the faces of a simplex.

1.9.31 Proposition (Properties of faces of a simplex) *For* $\Delta = \Delta(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k)$ *the following statements hold:*

(i) $(\cup_{F \in \mathscr{F}_l(\Delta)} F) \cap (\cup_{F' \in \mathscr{F}_m(\Delta)} F') = \varnothing$ *if* $l \neq m$*;*

(ii) *if* $F \in \mathscr{F}_m(\Delta)$ *then* $\mathrm{cl}(F)$ *is an* m*-simplex;*

(iii) $\mathrm{rel\,int}(\Delta) = F$ *where* $\mathscr{F}_k(\Delta) = \{F\}$*;*

(iv) $\mathrm{rel\,bd}(\Delta) = \cup_{j=0}^{m-1} \mathscr{F}_j(\Delta)$.

Proof (i) This follows from the following obvious statement: a point in an m-dimensional face of Δ is an element

$$\sum_{j=0}^{k} \lambda_j \boldsymbol{x}_j$$

for which exactly m of the numbers $\lambda_0, \lambda_1, \dots, \lambda_k$ are nonzero.

(ii) Suppose that $F = F_J$ for $J = \{j_1, j_1, \dots, j_m\}$. We claim that

$$\mathrm{cl}(F) = \Big\{ \sum_{l=0}^{m} \lambda_l \boldsymbol{x}_{j_l} \;\Big|\; \sum_{l=0}^{m} \lambda_l = 1,\ \lambda_0, \lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0} \Big\}.$$

Indeed, if $\boldsymbol{x} \in \mathrm{cl}(F)$ then there exists a sequence

$$\Big(\boldsymbol{x}_r = \sum_{l=0}^{m} \lambda_{l,r}\boldsymbol{x}_{j_l}\Big)_{r\in\mathbb{Z}_{>0}}$$

in F such that $\lim_{r\to\infty} \boldsymbol{x}_r = \boldsymbol{x}$. Thus, taking $\lambda_l = \lim_{r\to\infty} \lambda_{l,r}$, we have $\boldsymbol{x} = \sum_{l=1} \lambda_l \boldsymbol{x}_{j_l}$. The result follows since

$$\sum_{l=0}^{m} \lambda_l = \sum_{l=0}^{m} \lim_{r\to\infty} \lambda_{l,r} = \lim_{r\to\infty} \sum_{l=0}^{m} \lambda_{l,r} = 1$$

and since $\lambda_l \in \mathbb{R}_{\geq 0}$ for $l \in \{0, 1, \dots, m\}$.

(iii) By the previous part of the result (particularly understanding its proof) we have $\mathrm{cl}(F) = \Delta$. By Proposition 1.9.13 it follows that $\mathrm{cl}(\mathrm{rel\,int}(\Delta)) = \Delta$. Thus $\mathrm{rel\,int}(\Delta) \subseteq F$. To show that $F \subseteq \mathrm{rel\,int}(\Delta)$ it suffices to show that F is relatively open since $\mathrm{rel\,int}(\Delta)$ is the largest open set whose closure is Δ. To see that F is relatively open, let $\boldsymbol{x}_0 \in F$ and write $\boldsymbol{x}_0 = \sum_{j=0}^{k} \lambda_{0,j}\boldsymbol{x}_j$ for $\lambda_{0,0}, \lambda_{0,1}, \dots, \lambda_{0,k} \in \mathbb{R}_{>0}$ summing to 1. Let us denote $\boldsymbol{\lambda}_0 = (\lambda_{0,0}, \lambda_{0,1}, \dots, \lambda_{0,k}) \in \mathbb{R}^{k+1}$. There then exists $\epsilon \in \mathbb{R}_{>0}$ such that

$$B_{\epsilon,\boldsymbol{\lambda}_0} \triangleq \mathsf{B}^n(\epsilon, \boldsymbol{\lambda}_0) \cap \mathrm{aff}(\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}) \subseteq \Big\{ \sum_{j=0}^{k} \lambda_j \boldsymbol{x}_j \;\Big|\; \sum_{j=0}^{k} \lambda_j = 1,\ \lambda_0, \lambda_1, \dots, \lambda_k \in \mathbb{R}_{>0} \Big\}.$$

Thus

$$T \triangleq \Big\{ \sum_{j=0}^{k} \lambda_j \boldsymbol{x}_j \;\Big|\; \sum_{j=0}^{k} \lambda_j = 1,\ \lambda_0, \lambda_1, \dots, \lambda_k \in \mathbb{R}_{>0} \Big\}$$

is relatively open in $\mathrm{aff}(\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\})$. Now consider the linear map L

$$\mathbb{R}^{k+1} \ni \boldsymbol{\lambda} \mapsto \sum_{j=0}^{k} \lambda_j \boldsymbol{x}_j \subseteq \mathbb{R}^n.$$

Let

$$A = \Big\{ \sum_{j=0}^{k} \lambda_j \boldsymbol{x}_j \;\Big|\; \sum_{j=0}^{k} \lambda_j = 1 \Big\}.$$

We have $\mathsf{L}|A$ as a continuous onto $\mathrm{aff}(\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\})$. As per Example 1.3.38–2 we have $\mathsf{L}|A$ as a homeomorphism onto $\mathrm{aff}(\{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\})$. Thus $\mathsf{L}|T$ is a homeomorphism onto F, and so F is relatively open.

(iv) Since $\Delta = \mathrm{rel\,bd}(\Delta) \mathbin{\mathring{\cup}} \mathrm{rel\,int}(\Delta)$, this part of the result follows from the previous part. ■

In order to introduce the barycentric subdivision we shall use the following notation. If Δ_1 and Δ_2 are simplices then we write $\Delta_1 \leq \Delta_2$ if $\mathrm{rel\,int}(\Delta_1) \in \mathscr{F}_m(\Delta_2)$ for some m. If $\Delta_1 \leq \Delta_2$ but $\Delta_1 \neq \Delta_2$ then we write $\Delta_1 < \Delta_2$.

1.9.32 Definition (Barycentric subdivision) Let Δ be a simplex. The ***barycentric subdivision*** of Δ is the collection

$$\{\Delta(\mathrm{bc}(\Delta'_0), \mathrm{bc}(\Delta'_1), \dots, \mathrm{bc}(\Delta'_m)) \mid \Delta'_0 < \Delta'_1 < \dots < \Delta'_m \leq \Delta\}$$

of simplices. The collection of simplices in the barycentric subdivision of Δ is denoted by $\mathscr{S}_{\mathrm{bc}}(\Delta)$. •

It takes some time to parse the definition of the barycentric subdivision, so let us try to help things along by considering the definition for the standard 2-simplex.

1.9.33 Example (The barycentric subdivision of Δ_2) We take the standard 2-simplex $\Delta_2 = \Delta(\mathbf{0}, \boldsymbol{e}_1, \boldsymbol{e}_2)$. The faces of Δ_2 are the relative interiors of the following simplices.

1. 0-dimensional faces: $\Delta'_{01} = \Delta(\mathbf{0})$, $\Delta'_{02} = \Delta(\boldsymbol{e}_1)$, $\Delta'_{03} = \Delta(\boldsymbol{e}_2)$.
2. 1-dimensional faces: $\Delta'_{11} = \Delta(\mathbf{0}, \boldsymbol{e}_1)$, $\Delta'_{12} = \Delta(\mathbf{0}, \boldsymbol{e}_2)$, $\Delta'_{13} = \Delta(\boldsymbol{e}_1, \boldsymbol{e}_2)$.
3. 2-dimensional faces: $\Delta'_{21} = \Delta(\mathbf{0}, \boldsymbol{e}_1, \boldsymbol{e}_2)$.

Let us list the barycentres of the simplices which define the faces.

1. 0-dimensional faces: $\mathrm{bc}(\Delta'_{01}) = (0, 0)$, $\mathrm{bc}(\Delta'_{02}) = (1, 0)$, $\mathrm{bc}(\Delta'_{03}) = (1, 0)$.
2. 1-dimensional faces: $\mathrm{bc}(\Delta'_{11}) = (0, \frac{1}{2})$, $\mathrm{bc}(\Delta'_{12}) = (0, \frac{1}{2})$, $\mathrm{bc}(\Delta'_{13}) = (\frac{1}{2}, \frac{1}{2})$.
3. 2-dimensional faces: $\mathrm{bc}(\Delta'_{21}) = (\frac{1}{3}, \frac{1}{3})$.

The possible ways to order these simplices by our order "<" are as follows:

$$\Delta'_{01}, \quad \Delta'_{02}, \quad , \Delta'_{03},$$

$$\Delta'_{01} < \Delta'_{11}, \quad \Delta'_{01} < \Delta'_{12}, \quad \Delta'_{02} < \Delta'_{11}, \quad \Delta'_{02} < \Delta'_{13}, \quad \Delta'_{03} < \Delta'_{12}, \quad \Delta'_{03} < \Delta'_{13},$$

$$\Delta'_{01} < \Delta'_{11} < \Delta'_{21}, \quad \Delta'_{01} < \Delta'_{12} < \Delta'_{21}, \quad \Delta'_{02} < \Delta'_{11} < \Delta'_{21}, \quad \Delta'_{02} < \Delta'_{13} < \Delta'_{21},$$
$$\Delta'_{03} < \Delta'_{12} < \Delta'_{21}, \quad \Delta'_{03} < \Delta'_{13} < \Delta'_{21}.$$

With this data we can then determine the simplices in the barycentric subdivision. We show these in Figure 1.17. •

Now that we maybe have some intuition about the barycentric subdivision, let us list some of its properties.

1.9.34 Theorem (Properties of the barycentric subdivision) *For a simplex $\Delta = \Delta(\mathrm{x}_0, \mathrm{x}_1, \dots, \mathrm{x}_k)$ with barycentric subdivision $\mathscr{S}_{\mathrm{bc}}(\Delta)$ the following statements hold:*

(i) the number of r*-dimensional simplices in $\mathscr{S}_{\mathrm{bc}}(\Delta)$ is*

$$\sum_{\substack{d_0, d_1, \dots, d_r \in \\ \{0,1,\dots,k\}, \\ d_0 < d_1 < \dots < d_r}} \prod_{j=0}^{r} \binom{d_{j+1} + 1}{d_j + 1},$$

where we take $d_{r+1} = k$;

(ii) $\Delta = \mathring{\bigcup}_{\Delta' \in \mathscr{S}_{\mathrm{bc}}(\Delta)} \operatorname{rel\,int}(\Delta')$;

(iii) $\max\{\operatorname{diam}(\Delta') \mid \Delta' \in \mathscr{S}_{\mathrm{bc}}(\Delta)\} \leq \frac{k}{k+1} \operatorname{diam}(\Delta)$.

Proof (i) Let us give some counting lemmata.

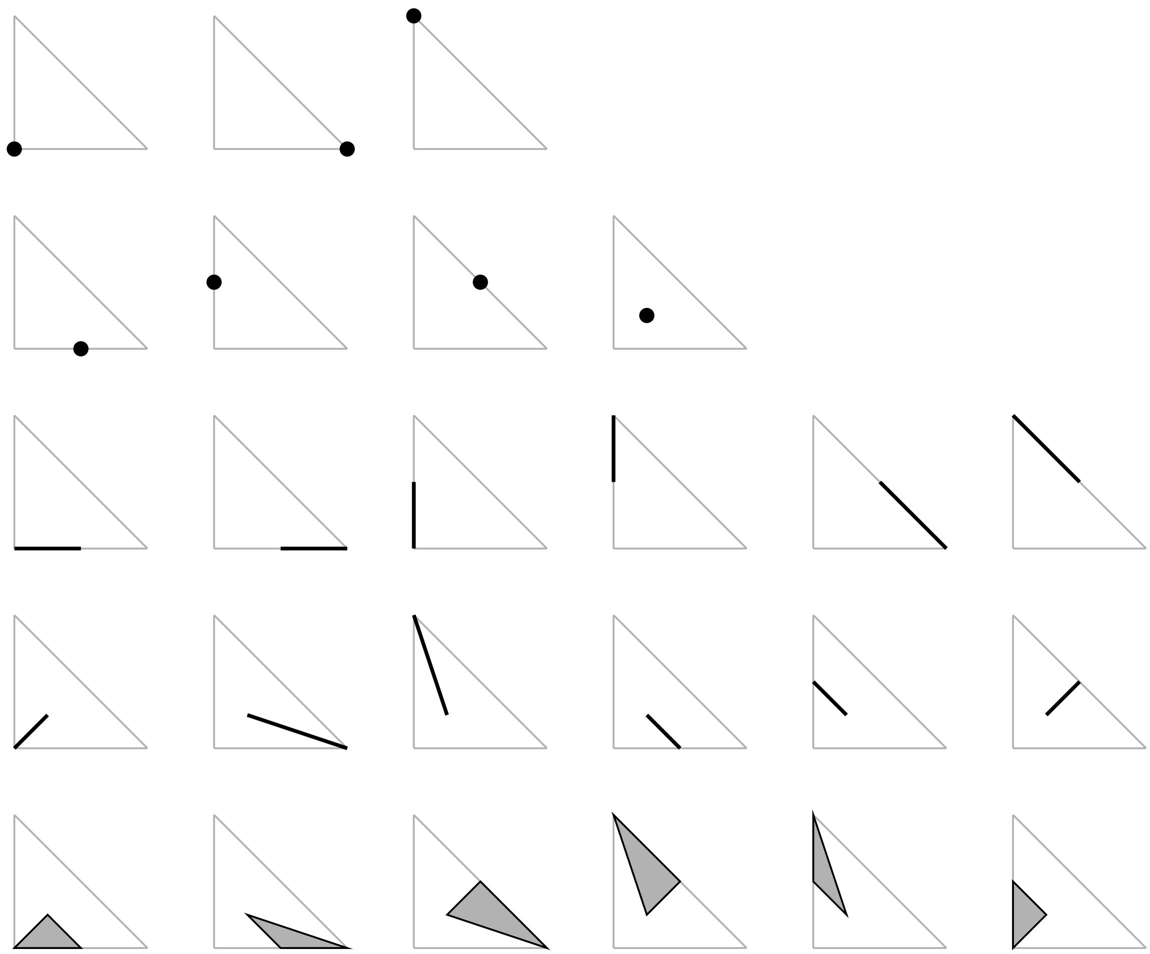

Figure 1.17 The 0-dimensional (top two rows), 1-dimensional (next two rows), and 2-dimensional (bottom row) simplices in the barycentric subdivision of Δ_2

1 Lemma *The number of* r*-dimensional faces of a* k*-simplex is* $\binom{\mathrm{k}+1}{\mathrm{r}+1} = \frac{(\mathrm{k}+1)!}{(\mathrm{r}+1)!(\mathrm{k}-\mathrm{r})!}$.

Proof By definition an r-dimensional face is specified by a subset $J = \{j_0, j_1, \dots, j_r\} \subseteq \{0, 1, \dots, k\}$. The number of such subsets is exactly $\binom{k+1}{r+1}$; this is Exercise I-1.1.4. ▼

2 Lemma *If* Δ *is a* k*-simplex and if* $\mathrm{F} \in \mathscr{F}_{\mathrm{r}}(\Delta)$ *then the number of* s*-dimensional faces* F' *for which* $\mathrm{cl}(\mathrm{F}') < \mathrm{cl}(\mathrm{F})$ *is* $\binom{\mathrm{r}+1}{\mathrm{s}+1}$.

Proof For an r-dimensional face F there are indices $J = \{j_0, j_1, \dots, j_r\} \subseteq \{0, 1, \dots, k\}$ such that every point $\boldsymbol{x} \in F$ is written as

$$\boldsymbol{x} = \sum_{l=0}^{r} \lambda_l \boldsymbol{x}_{j_l}$$

for unique numbers $\lambda_0, \lambda_1, \dots, \lambda_r \in \mathbb{R}_{>0}$ summing to 1. Now, if $F' \in \mathscr{F}_s(\Delta)$ and if

$\mathrm{cl}(F') < \mathrm{cl}(F)$ then F' is a face of $\mathrm{cl}(F)$. If we write a typical point $\boldsymbol{x}' \in F'$ as

$$\boldsymbol{x}' = \sum_{l'=0}^{s} \lambda'_{l'} \boldsymbol{x}_{j'_{l'}},$$

for $J' = \{j'_0, j'_1, \dots, j'_s\}$ then it follows that $J' \subseteq J$. Thus there are as many s-dimensional faces F' such that $\mathrm{cl}(F') < \mathrm{cl}(F)$ as there are subsets of cardinality $s+1$ of a set of cardinality $r+1$, i.e., $\binom{r+1}{s+1}$. ▼

Every r-dimensional simplex $\Delta' \in \mathscr{S}_{\mathrm{bc}}(\Delta)$ has the form

$$\Delta' = \Delta(\mathrm{bc}(\Delta'_0), \mathrm{bc}(\Delta'_1), \dots, \mathrm{bc}(\Delta'_r))$$

for $\Delta'_0 < \Delta'_1 < \dots < \Delta'_r$. Suppose that the dimension of Δ'_j is d_j so that $d_0 < d_1 < \dots < d_r$. By the first lemma above there are $\binom{k+1}{d_r+1}$ possible faces of dimension d_r. Now, fixing one of these d_r-dimensional faces—denote it by F_r—by the second lemma above there are $\binom{d_r+1}{d_{r-1}+1}$ possible d_{r-1}-dimensional faces F_{r-1} for which $F_{r-1} < F_r$. Continuing in this way, upon fixing the dimensions $d_0 < d_1 < \dots < d_r$, we have

$$\binom{d_1+1}{d_0+1} \cdots \binom{d_r+1}{d_{r-1}+1} \binom{k+1}{d_r+1}$$

possible combinations of faces $F_0, F_1, \dots, F_r$ of dimensions $d_0, d_1, \dots, d_r$, respectively and such that $F_0 < F_1 < \dots < F_r$. Summing over all possible dimensions $d_0, d_1, \dots, d_r$ gives the number of r-dimensional simplices in $\mathscr{S}_{\mathrm{bc}}(\Delta)$ as

$$\sum_{\substack{d_0, d_1, \dots, d_r \in \\ \{0,1,\dots,k\}, \\ d_0 < d_1 < \dots < d_r}} \prod_{j=0}^{r} \binom{d_{j+1}+1}{d_j+1},$$

as stated upon taking $d_{r+1} = k$.

(ii) Let $\boldsymbol{x} \in \operatorname{rel\,int}(\Delta')$ where $\Delta' \in \mathscr{S}_{\mathrm{bc}}(\Delta)$. Suppose that Δ' has dimension m so that there exists $J = \{j_0, j_1, \dots, j_m\} \subseteq \{0, 1, \dots, k\}$ such that

$$\boldsymbol{x} = \sum_{l=0}^{m} \lambda_l \mathrm{bc}(\Delta'_l)$$

for some $\Delta'_0, \Delta'_1, \dots, \Delta'_m \in \mathscr{S}_{\mathrm{bc}}(\Delta)$ satisfying $\Delta'_0 < \Delta'_1 < \dots < \Delta'_m$. For each $l \in \{0, 1, \dots, m\}$ the simplex Δ'_l is given by

$$\Delta'_l = \Delta(\boldsymbol{x}_{l,j_0}, \boldsymbol{x}_{l,j_1}, \dots, \boldsymbol{x}_{l,j_{m_l}})$$

for $\boldsymbol{x}_{l,j_s} \in \{\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$, $s \in \{0, 1, \dots, m_l\}$. Therefore,

$$\mathrm{bc}(\Delta'_l) = \sum_{s=0}^{m_l} \frac{1}{m_l+1} \boldsymbol{x}_{l,j_s},$$

and so

$$x = \sum_{l=0}^{m} \sum_{s=0}^{m_l} \frac{\lambda_l}{m_l + 1} x_{l,j_s}.$$

This implies that x is a linear combination of the vectors $\{x_0, x_1, \dots, x_k\}$ with positive coefficients summing to 1. That is $x \in \Delta$ and so

$$\cup_{\Delta' \in \mathscr{S}_{\mathrm{bc}}(\Delta)} \operatorname{rel\,int}(\Delta') \subseteq \Delta.$$

To show the opposite inclusion we use induction on k. The result is trivially true for $k = 0$; assume it true for $k \in \{0, 1, \dots, r\}$ and suppose that Δ is an $(r+1)$-simplex. Let $x \in \Delta$. If

$$x = \sum_{j=0}^{r+1} \lambda_j x_j \tag{1.50}$$

where any one of the coefficients λ_j is zero, then x is in some face F of Δ of dimension less than $r+1$, and the induction hypothesis implies that

$$x \in \Delta(\mathrm{bc}(\Delta'_0), \mathrm{bc}(\Delta'_1), \dots, \mathrm{bc}(\Delta'_m))$$

where $\Delta'_0 < \Delta'_1 < \dots < \Delta'_m \leq \mathrm{cl}(F)$. Since $\mathscr{S}_{\mathrm{bc}}(\mathrm{cl}(F)) \subseteq \mathscr{S}_{\mathrm{bc}}(\Delta)$ (do you see why?) it follows that $x \in \cup_{\Delta' \in \mathscr{S}_{\mathrm{bc}}(\Delta)} \operatorname{rel\,int}(\Delta')$. If $x \in \operatorname{rel\,int}(\Delta)$ (i.e., if none of the coefficients in (1.50) are zero) then we make use of the following lemma.

3 Lemma *If the vectors* $\{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k\}$ *are affinely independent, if* $\Delta = \Delta(\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_k)$*, and if* $\mathbf{y} \in \operatorname{rel\,int}(\Delta)$ *then*

$$\Delta = \{(1-s)\mathbf{y} + s\mathbf{x} \mid \mathbf{x} \in \operatorname{rel\,bd}(\Delta),\ s \in [0,1]\}.$$

Moreover, if $\mathbf{x}'_1, \mathbf{x}'_2 \in \operatorname{rel\,bd}(\Delta)$ *are distinct then*

$$\{(1-s)\mathbf{y} + s\mathbf{x}'_1 \mid s \in [0,1]\} \cap \{(1-s)\mathbf{y} + s\mathbf{x}'_2 \mid s \in [0,1]\} = \{\mathbf{y}\}.$$

Proof This is a restatement of Lemma 1 in the proof of Theorem 1.9.14. ▼

Now we take $x \in \operatorname{rel\,int}(\Delta)$. If $x = \mathrm{bc}(\Delta)$ then it immediately follows that $x \in \cup_{\Delta' \in \mathscr{S}_{\mathrm{bc}}(\Delta)} \operatorname{rel\,int}(\Delta')$. Otherwise, by the lemma, there exists a unique $s \in (0,1)$ and $x' \in \operatorname{rel\,bd}(\Delta)$ such that $x = (1-s)\mathrm{bc}(\Delta) + sx'$. Since $x' \in \operatorname{rel\,bd}(\Delta)$, by Proposition 1.9.31(iv) and by the induction hypothesis, there exists $\Delta'_0, \Delta'_1, \dots, \Delta'_m$ such that

$$\Delta'_0 < \Delta'_1 < \dots < \Delta'_m < \Delta$$

and

$$x' = \sum_{l=0}^{m} \lambda_l \mathrm{bc}(\Delta'_l).$$

Then

$$x = (1-s)\mathrm{bc}(\Delta) + \sum_{l=0}^{m} s\lambda_l \mathrm{bc}(\Delta'_l) = \sum_{l=0}^{m+1} \mu_l \mathrm{bc}(\Delta'_l)$$

upon taking

$$\mu_l = s\lambda_l,\ l \in \{0,1,\dots,m\}, \quad \mu_{m_1} = (1-s), \quad \Delta'_{m+1} = \Delta.$$

Since $\sum_{l=0}^{m+1} \mu_l = 1$ it follows that $x \in \cup_{\Delta' \in \mathscr{S}_{\mathrm{bc}}(\Delta)} \operatorname{rel\,int}(\Delta')$.

Now we need to show that if $\Delta'_1, \Delta'_2 \in \mathscr{S}_{\mathrm{bc}}(\Delta)$ are distinct then $\operatorname{rel\,int}(\Delta'_1) \cap \operatorname{rel\,int}(\Delta'_2) = \emptyset$. We prove this by induction on k. It is clearly true that the result holds for $k = 0$. Assume the result holds for $k \in \{0,1,\dots,r\}$ and let Δ be an $(r+1)$-simplex. Let $\Delta'_1, \Delta'_2 \in \mathscr{S}_{\mathrm{bc}}(\Delta)$. If $\Delta'_1, \Delta'_2 \subseteq \operatorname{rel\,bd}(\Delta)$ then the result holds by Proposition 1.9.31(iv) and the induction hypothesis. If $\Delta'_1 \subseteq \operatorname{rel\,bd}(\Delta)$ and $\Delta'_2 \not\subset \operatorname{rel\,bd}(\Delta)$ then Δ'_2 is $(r+1)$-dimensional and so it immediately follows that $\operatorname{rel\,int}(\Delta'_1) \cap \operatorname{rel\,int}(\Delta'_2) = \emptyset$. The remaining case to consider is when Δ'_1 and Δ'_2 are both $(r+1)$-dimensional. In this case, if $x_1 \in \operatorname{rel\,int}(\Delta'_1)$ and $x_2 \in \operatorname{rel\,int}(\Delta'_2)$ then, by the lemma above, there exists unique $x'_1, x'_2 \in \operatorname{rel\,bd}(\Delta)$ and $s_1, s_2 \in (0,1)$ such that

$$x_1 = (1-s_1)\mathrm{bc}(\Delta) + s_1 x'_1, \quad x_2 = (1-s_2)\mathrm{bc}(\Delta) + s_2 x'_2.$$

Now $x'_1 \in \Delta''_1$ and $x'_2 \in \Delta''_2$ where $\Delta''_1, \Delta''_2 \in \operatorname{rel\,bd}(\Delta)$ are given by

$$\Delta''_1 = \Delta(\mathrm{bc}(\Delta''_{10}), \mathrm{bc}(\Delta''_{11}), \dots, \mathrm{bc}(\Delta''_{1,m_1})),$$
$$\Delta''_2 = \Delta(\mathrm{bc}(\Delta''_{20}), \mathrm{bc}(\Delta''_{21}), \dots, \mathrm{bc}(\Delta''_{2,m_2}))$$

for simplices

$$\Delta''_{10} < \Delta''_{11} < \cdots < \Delta''_{1m_1}, \quad \Delta''_{20} < \Delta''_{21} < \cdots < \Delta''_{2m_2}.$$

Now note that if $x_1 = x_2$ then $x'_1 = x'_2$ and so $\operatorname{rel\,int}(\Delta''_1) \cap \operatorname{rel\,int}(\Delta''_2) \neq \emptyset$. By the induction hypothesis this means that $\Delta''_1 = \Delta''_2$. This in turn means that $\Delta'_1 = \Delta'_2$. This shows that if $\operatorname{rel\,int}(\Delta'_1) \cap \operatorname{rel\,int}(\Delta'_2) \neq \emptyset$ then $\Delta'_1 = \Delta'_2$, giving this part of the result.

(iii) Let us first reduce ourselves to considering only lengths of 1-dimensional faces.

4 **Lemma** *If* $\Delta = \Delta(\mathrm{x}_0, \mathrm{x}_1, \dots, \mathrm{x}_\mathrm{k})$ *is a* k*-simplex then there exists* $\mathrm{j}_1, \mathrm{j}_2 \in \{0,1,\dots,\mathrm{k}\}$ *such that*

$$\operatorname{diam}(\Delta) = \|\mathrm{x}_{\mathrm{j}_1} - \mathrm{x}_{\mathrm{j}_2}\|_{\mathbb{R}^\mathrm{n}}.$$

Proof Let $j_1, j_2 \in \{0,1,\dots,k\}$ be such that

$$\|x_{j_1} - x_{j_2}\|_{\mathbb{R}^n} = \max\{\|x_{j'_1} - x_{j'_2}\|_{\mathbb{R}^n} \mid j'_1, j'_2 \in \{0,1,\dots,k\}\}.$$

Now let $x, y \in \Delta$. Let $j_0 \in \{0,1,\dots,k\}$ have the property that

$$\|x - x_{j_0}\|_{\mathbb{R}^n} = \max\{\|x - x_j\|_{\mathbb{R}^n} \mid j \in \{0,1,\dots,k\}\},$$

and denote $r = \|x - x_{j_0}\|_{\mathbb{R}^n}$. Then it follows that $\overline{\mathsf{B}}^n(r,x)$ contains all vertices x_j, $j \in \{0,1,\dots,k\}$. Moreover, $\overline{\mathsf{B}}^n(r,x)$ is convex by Exercise 1.9.2. Therefore, $\Delta \subseteq \overline{\mathsf{B}}^n(r,x)$ since Δ is the convex hull of the vertices. Thus $\|x-y\|_{\mathbb{R}^n} \le \|x - x_{j_0}\|_{\mathbb{R}^n}$. Next let $j'_0 \in \{0,1,\dots,k\}$ be such that

$$\|x_{j_0} - x_{j'_0}\|_{\mathbb{R}^n} \le \max\{\|x_{j_0} - x_j\|_{\mathbb{R}^n} \mid j \in \{0,1,\dots,k\}\}$$

and let $r' = \|x_{j_0} - x_{j'_0}\|_{\mathbb{R}^n}$. By the same argument as above, $\|x - x_{j_0}\|_{\mathbb{R}^n} \le \|x_{j_0} - x_{j'_0}\|_{\mathbb{R}^n}$. Thus

$$\|x - y\|_{\mathbb{R}^n} \le \|x - x_{j_0}\|_{\mathbb{R}^n} \le \|x_{j_0} - x_{j'_0}\|_{\mathbb{R}^n}.$$

Thus $\|x - y\|_{\mathbb{R}^n} \le x_{j_1} - x_{j_2}$. ▼

Now let Δ' be a 1-dimensional simplex from $\mathscr{S}_{\mathrm{bc}}(\Delta)$ and let $\mathrm{bc}(\Delta'_1)$ and $\mathrm{bc}(\Delta'_2)$ be the vertices of Δ', supposing that $\Delta'_1 < \Delta'_2$. Suppose that Δ'_1 is m_1-dimensional and that Δ'_2 is m_2-dimensional. Then we can write

$$\begin{aligned}\Delta'_1 &= \Delta(\boldsymbol{x}_{j_0}, \boldsymbol{x}_{j_1}, \dots, \boldsymbol{x}_{j_{m_1}}),\\ \Delta'_2 &= \Delta(\boldsymbol{x}_{j_0}, \boldsymbol{x}_{j_1}, \dots, \boldsymbol{x}_{j_{m_1}}, \boldsymbol{x}_{j_{m_1+1}}, \dots, \boldsymbol{x}_{j_{m_2}})\end{aligned}$$

for suitable $j_0, j_1, \dots, j_{m_2} \in \{0, 1, \dots, k\}$. We then compute

$$\begin{aligned}\mathrm{bc}(\Delta'_1) - \mathrm{bc}(\Delta'_2) &= \frac{1}{m_1+1}\sum_{l=0}^{m_1} \boldsymbol{x}_{j_l} - \frac{1}{m_2+1}\sum_{l=0}^{m_2} \boldsymbol{x}_{j_l}\\ &= \Big(\frac{1}{m_1+1} - \frac{1}{m_2+1}\Big)\sum_{l=0}^{m_1} \boldsymbol{x}_{j_l} - \frac{1}{m_2+1}\sum_{l=m_1+1}^{m_2} \boldsymbol{x}_{j_l}\\ &= \frac{m_2-m_1}{m_2+1}\Big(\frac{1}{m_1+1}\sum_{l=1}^{m_1} \boldsymbol{x}_{j_l} - \frac{1}{m_2-m_1}\sum_{l=m_1+1}^{m_2} \boldsymbol{x}_{j_l}\Big).\end{aligned}$$

Note that since

$$\sum_{l=0}^{m_1} \frac{m_1+1}{=}1, \quad \sum_{l=m_1+1}^{m_2} \frac{1}{m_2-m_1} = 1$$

it follows that

$$\frac{1}{m_1+1}\sum_{l=1}^{m_1} \boldsymbol{x}_{j_l}, \frac{1}{m_2-m_1}\sum_{l=m_1+1}^{m_2} \boldsymbol{x}_{j_l} \in \Delta'_2.$$

Since $\Delta'_2 \subseteq \Delta$ it follows that

$$\Big\|\frac{1}{m_1+1}\sum_{l=1}^{m_1} \boldsymbol{x}_{j_l} - \frac{1}{m_2-m_1}\sum_{l=m_1+1}^{m_2} \boldsymbol{x}_{j_l}\Big\|_{\mathbb{R}^n} < \mathrm{diam}(\Delta).$$

Therefore,

$$\|\mathrm{bc}(\Delta'_1) - \mathrm{bc}(\Delta'_2)\|_{\mathbb{R}^n} \le \tfrac{m_2-m_1}{m_2+1}\mathrm{diam}(\Delta) \le \tfrac{m_2}{m_2+1}\mathrm{diam}(\Delta) \le \tfrac{k}{k+1}\mathrm{diam}(\Delta),$$

giving the result. ■

The most interesting of the numbers from part (i) of the theorem is when $r = k$. In this case we get the number of k-dimensional simplices in the barycentric subdivision as

$$\sum_{j=0}^{k} \binom{2}{1} \cdots \binom{k}{k-1}\binom{k+1}{k} = (k+1)!.$$

It is the number of simplices of dimension k is, in some sense, the most indicative of the character of the barycentric subdivision.

Having now divided a simplex into a bunch of smaller simplices in a systematic way, and having acquired an estimate (in the form of part (iii) of the preceding theorem) on the size of the resulting simplices, we can now repeat this process on the $(k+1)!$ simplices of the barycentric subdivision having the highest dimension. This process can then be repeated inductively, giving the following definition.

1.9.35 Definition (mth barycentric subdivision) Let Δ be a k-simplex. For $m \in \mathbb{Z}_{>0}$ define the **m*th barycentric subdivision*** inductively as follows:

1. let $\mathscr{S}^1_{\mathrm{bc}}(\Delta) = \mathscr{S}_{\mathrm{bc}}(\Delta)$;
2. assuming $\mathscr{S}^j_{\mathrm{bc}}(\Delta)$, $j \in \{1, \dots, m-1\}$, defined, let

$$\mathscr{S}^m_{\mathrm{bc}}(\Delta) = \{\Delta' \mid \Delta' \in \mathscr{S}_{\mathrm{bc}}(\Delta''),\ \Delta'' \in \mathscr{S}^{m-1}_{\mathrm{bc}}(\Delta)\}.$$ •

It is evident that

$$\lim_{m\to\infty} \max\{\mathrm{diam}(\Delta') \mid \Delta' \in \mathscr{S}^m_{\mathrm{bc}}(\Delta)\} = 0. \tag{1.51}$$

Thus the diameter of the simplices gets smaller and smaller. We shall use this construction in our proof of the Kakutani Fixed Point Theorem.

1.9.8 Duality

The subject of duality is an important one in the study of convexity. In Section I-5.7 we studied duality in the context of linear algebra. There is some relationship with this discussion and our discussion here of duality in convexity. The main idea is that one studies an object by studying functions related to it. In the case of vector spaces, the functions are linear functions. For convex sets, the appropriate objects to study are affine functions, i.e., functions $\alpha\colon \mathbb{R}^n \to \mathbb{R}$ that are affine as per Definition 1.3.15. Note that such an affine function comes in the form

$$\alpha(\boldsymbol{x}) = \langle \boldsymbol{a}, \boldsymbol{x}\rangle_{\mathbb{R}^n} + b \tag{1.52}$$

for $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$. We shall often write α in this form without comment. Let us denote the set of affine functions on $\mathbb{R}^n$ by $A(\mathbb{R}^n; \mathbb{R})$. Note that $A(\mathbb{R}^n; \mathbb{R})$ is a vector space isomorphic to $\mathbb{R}^n \oplus \mathbb{R}$. Thus it makes sense to talk about convexity in $A(\mathbb{R}^n; \mathbb{R})$, and this is something we shall do.

With this in mind, we make the following definition.

1.9.36 Definition (Polar cone, polar) For $S \subseteq \mathbb{R}^n$,

(i) the ***polar cone*** of S is

$$S^* = \{\alpha \in A(\mathbb{R}^n; \mathbb{R}) \mid \alpha(\boldsymbol{x}) \le 0 \text{ for all } \boldsymbol{x} \in S\}$$

and

(ii) the ***polar*** of S is

$$S^\circ = \{\boldsymbol{a} \in \mathbb{R}^n \mid \langle \boldsymbol{a}, \boldsymbol{x}\rangle_{\mathbb{R}^n} \le 1 \text{ for all } \boldsymbol{x} \in S\}.$$ •

For $b \in \mathbb{R}$ we denote $S^*_b \subseteq \mathbb{R}^n$ by

$$S^*_b = \{\boldsymbol{a} \in \mathbb{R}^n \mid (\boldsymbol{a}, b) \in S^*\}.$$ •

The following result contains some useful elementary properties of polar cones and polars.

1.9.37 Proposition (Elementary properties of polar cones and polars) *For* $\mathsf{S}, \mathsf{T} \subseteq \mathbb{R}^n$, *the following statements hold:*

(i) $({}^*\mathsf{S}) = ({}^*\operatorname{cl}(\operatorname{conv}(\mathsf{S})))$ *and* $({}^\circ\mathsf{S}) = ({}^\circ\operatorname{cl}(\operatorname{conv}(\mathsf{S})))$;

(ii) *if* $\mathsf{S} \subseteq \mathsf{T}$, *then* $\mathsf{T}^* \subseteq \mathsf{S}^*$ *and* $\mathsf{T}^\circ \subseteq \mathsf{S}^\circ$;

(iii) $(\mathsf{S} \cup \mathsf{T})^* = \mathsf{S}^* \cap \mathsf{T}^*$ *and* $(\mathsf{S} \cup \mathsf{T})^\circ = \mathsf{S}^\circ \cap \mathsf{T}^\circ$;

(iv) *if* $\lambda \in \mathbb{R} \setminus \{0\}$ *then*

$$(\lambda \mathsf{S})^* = \{(\mathbf{a}, \mathrm{b}) \in \mathsf{A}(\mathbb{R}^n; \mathbb{R}) \mid (\lambda \mathbf{a}, \mathrm{b}) \in \mathsf{S}^*\}$$

(identifying $\mathsf{A}(\mathbb{R}^n; \mathbb{R})$ *with* $\mathbb{R}^n \oplus \mathbb{R}$ *as in (1.52)) and* $(\lambda \mathsf{S})^\circ = \lambda^{-1}(\mathsf{S}^\circ)$;

(v) *if* $\mathbf{f}\colon \mathbb{R}^n \to \mathbb{R}^n$ *is the affine map* $\mathbf{f}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{d}$ *for* $\mathbf{A} \in \mathsf{L}(\mathbb{R}^n; \mathbb{R}^n)$ *and* $\mathbf{d} \in \mathbb{R}^n$, *then*

$$\begin{aligned}\mathbf{f}(\mathsf{S})^* &= \{(\mathbf{a}', \mathrm{b}') \in \mathbb{R}^n \times \mathbb{R} \mid (\mathbf{A}^T\mathbf{a}', \langle \mathbf{a}', \mathbf{d}\rangle_{\mathbb{R}^n} + \mathrm{b}') \in \mathsf{S}^*\} \\ &= \{(\mathbf{A}^{-T}\mathbf{a}, \mathrm{b} - \langle \mathbf{A}^{-T}\mathbf{a}, \mathbf{d}\rangle_{\mathbb{R}^n}) \in \mathbb{R}^n \times \mathbb{R} \mid (\mathbf{a}, \mathrm{b}) \in \mathsf{S}^*\}\end{aligned}$$

and

$$\begin{aligned}\mathbf{f}(\mathsf{S})^\circ &= \{\mathbf{a}' \in \mathbb{R}^n \mid \mathbf{A}^T\mathbf{a}' \in \mathsf{S}^*_{\langle \mathbf{a}', \mathbf{d}\rangle_{\mathbb{R}^n} - 1}\} \\ &= \{\mathbf{A}^{-T}\mathbf{a} \in \mathbb{R}^n \mid \mathbf{a} \in \mathsf{S}^*_{\langle \mathbf{A}^{-T}\mathbf{a}, \mathbf{d}\rangle_{\mathbb{R}^n} - 1}\}.\end{aligned}$$

Proof We leave this to the reader as Exercise 1.9.9. ■

One can verify that S^* is a convex cone and that S° is convex, regardless of what S might look like.

1.9.38 Proposition (The polar (cone) is a closed convex set (cone)) *If* $\mathsf{S} \subseteq \mathbb{R}^n$ *then*

(i) S^* *is a closed convex cone with* $\mathbf{0}$ *as vertex and*

(ii) S° *is a closed convex set containing* $\mathbf{0}$.

Moreover, S^* *and* S° *are related in the following ways:*

(iii) *if* $\mathrm{b} \in \mathbb{R}_{<0}$, $\mathsf{S}^*_{\mathrm{b}} = |\mathrm{b}|(\mathsf{S}^\circ)$;

(iv) *if* $\mathrm{b} \in \mathbb{R}_{>0}$ *then* $\mathsf{S}^*_{\mathrm{b}} = \varnothing$ *if and only if* $\mathbf{0} \notin \operatorname{cl}(\operatorname{conv}(\mathsf{S}))$;

(v) *if* $\mathrm{b} \in \mathbb{R}_{>0}$ *and if* $\mathsf{S}^*_{\mathrm{b}} \neq \varnothing$, *then* $\mathsf{S}^*_{\mathrm{b}} = \mathrm{b}(\mathsf{S}^\circ)$.

Proof (i) First of all, if $\alpha \in S^*$ is given by

$$\alpha(\boldsymbol{x}) = \langle \boldsymbol{a}, \boldsymbol{x}\rangle_{\mathbb{R}^n} + b$$

then

$$(\lambda\alpha)(\boldsymbol{x}) = \lambda\alpha(\boldsymbol{x})$$

for $\lambda \in \mathbb{R}$. Thus, if $\lambda \in \mathbb{R}_{\geq 0}$ we have $(\lambda\alpha)(\boldsymbol{x}) \leq 0$ for all $\boldsymbol{x} \in S$, and so $\lambda\alpha \in S^*$. Thus S^* is a cone with vertex $\mathbf{0}$. Next suppose that $\alpha_1, \alpha_2 \in S^*$. Then, for $s \in [0, 1]$,

$$((1-s)\alpha_1 + s\alpha_2)(\boldsymbol{x}) = (1-s)\langle \boldsymbol{a}_1, \boldsymbol{x}\rangle_{\mathbb{R}^n} + (1-s)b_1 + s\langle \boldsymbol{a}_2, \boldsymbol{x}\rangle_{\mathbb{R}^n} + sb_2 \leq 0,$$

showing that S^* is convex. Finally, note that, for $x \in S$, the function $f_x\colon A(\mathbb{R}^n;\mathbb{R}) \to \mathbb{R}$ defined by $f_x(\alpha) = \alpha(x)$ is itself affine, and so continuous by Proposition 1.3.16. Therefore, for each $x \in S$, $f_x(\mathbb{R}_{\le 0})$ is closed by Corollary 1.3.4. Since $S^* = \cap_{x\in S} f_x^{-1}(\mathbb{R}_{\le 0})$, it follows from Proposition 1.2.19 that S^* is closed.

(ii) We assume part (iii) proven below. By this result we have $S^\circ = S^* \cap (\mathbb{R}^n \times \{-1\})$. Since S^* is convex and since $\mathbb{R}^n \times \{-1\}$ is convex (it is a hyperplane), it follows from Proposition 1.2.19 and Exercise 1.9.4 that S° is closed and convex.

(iii) Let us prove the last assertion of the proposition. First suppose that $b \in \mathbb{R}_{<0}$. Then

$$\begin{aligned} (a,b) \in S^* \quad &\iff \quad \langle a, x\rangle_{\mathbb{R}^n} + b \le 0, \qquad x \in S \\ &\iff \quad -b^{-1}\langle a, x\rangle_{\mathbb{R}^n} \le 1, \qquad x \in S \\ &\iff \quad \langle a, y\rangle_{\mathbb{R}^n} \le 1, \qquad y \in -b^{-1}S \\ &\iff \quad a \in (-b^{-1}S)^\circ \quad \iff \quad a \in -b(S^\circ), \end{aligned}$$

using Proposition 1.9.37.

(iv) Let $b \in \mathbb{R}_{>0}$. By Proposition 1.9.37(i) we may assume that S is closed and convex.

If $\mathbf{0} \in S$ and if $a \in \mathbb{R}^n$ then

$$\langle a, \mathbf{0}\rangle_{\mathbb{R}^n} + b = b > 0,$$

showing that there cannot be any $a \in \mathbb{R}^n$ such that $\langle a, x\rangle_{\mathbb{R}^n} + b \le 0$ for every $x \in S$.

Conversely, suppose that $\mathbf{0} \notin S$. Since S is closed, there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\overline{\mathsf{B}}^n(\epsilon, \mathbf{0}) \cap S = \varnothing$. By , let $a' \in \mathbb{R}^n$ and $b' \in \mathbb{R}_{>0}$ be such that $\langle a', x\rangle_{\mathbb{R}^n} + b' < 0$ for every $x \in S$ and $\langle a', x\rangle_{\mathbb{R}^n} + b' > 0$ for every $x \in \overline{\mathsf{B}}^n(\epsilon, \mathbf{0})$. Then, taking $a = \frac{b}{b'}a'$, we see that $\langle a, x\rangle_{\mathbb{R}^n} + b < 0$ for every $x \in S$, as desired.

closed and compact strictly separated

(v) If $b \in \mathbb{R}_{>0}$,

$$\begin{aligned} (a,b) \in S^* \quad &\iff \quad \langle a, x\rangle_{\mathbb{R}^n} + b \le 0, \qquad x \in S \\ &\iff \quad b^{-1}\langle a, x\rangle_{\mathbb{R}^n} \le -1, \qquad x \in S \\ &\iff \quad \langle a, y\rangle_{\mathbb{R}^n} \le 1, \qquad y \in b^{-1}S \\ &\iff \quad a \in (b^{-1}S)^\circ \quad \iff \quad a \in b(S^\circ), \end{aligned}$$

using Proposition 1.9.37. ■

The following theorem gives the most interesting property of the polar cone.

1.9.39 Theorem (The Polar (Cone) Theorem) *If* $S \subseteq \mathbb{R}^n$ *then*

(i) $\{x \in \mathbb{R}^n \mid \alpha(x) \le 0$ *for all* $\alpha \in S^*\} = \mathrm{cl}(\mathrm{conv}(S))$ *and*

(ii) $\{x \in \mathbb{R}^n \mid \langle a, x\rangle_{\mathbb{R}^n} \le 1$ *for all* $a \in S^\circ\} = \mathrm{cl}(\mathrm{conv}(S \cup \{\mathbf{0}\}))$.

Proof (i) First let $x \in \mathrm{conv}(S)$. Then, by Proposition 1.9.4, there exists $x_1, \dots, x_k \in S$ and $\lambda_1, \dots, \lambda_k \in (0,1)$ satisfying $\sum_{j=1}^k \lambda_j = 1$ such that $x = \sum_{j=1}^k \lambda_j x_j$. Then, for $\alpha \in S^*$,

$$\alpha(x) = \sum_{j=1}^k \lambda_j \langle a, x_j\rangle_{\mathbb{R}^n} + b\sum_{j=1}^k \lambda_j = \sum_{j=1}^k \lambda_j(\langle a, x_j\rangle_{\mathbb{R}^n} + b) \le 0.$$

Thus

$$\mathrm{conv}(S) \subseteq \{y \in \mathbb{R}^n \mid \alpha(y) \le 0 \text{ for all } \alpha \in S^*\}.$$

Next let $x \in \mathrm{cl}(\mathrm{conv}(S))$. Then, $x = \lim_{j\to\infty} x_j$ for a sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathrm{conv}(S)$. As in the preceding paragraph, if $\alpha \in S^*$, then $\alpha(x_j) \le 0$ for each $j \in \mathbb{Z}_{>0}$. Therefore, by continuity of α and Theorem 1.3.2,

$$\alpha(x) = \lim_{j\to\infty} \alpha(x_j) \le 0,$$

showing that

$$\mathrm{cl}(\mathrm{conv}(S)) \subseteq \{y \in \mathbb{R}^n \mid \alpha(y) \le 0 \text{ for all } \alpha \in S^*\}.$$

For the converse inclusion, suppose that $x \notin \mathrm{cl}(\mathrm{conv}(S))$. By Corollary 1.9.17 there exists $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that

$$\langle a, x\rangle_{\mathbb{R}^n} > -b, \quad \langle a, y\rangle_{\mathbb{R}^n} < -b, \qquad y \in \mathrm{cl}(\mathrm{conv}(S)).$$

Thus, if $\alpha \in A(\mathbb{R}^n; \mathbb{R})$ is defined by

$$\alpha(x) = \langle a, x\rangle_{\mathbb{R}^n} + b,$$

then $\alpha(y) < 0$ for all $y \in S$ and so $\alpha \in S^*$. Since $\alpha(x) > 0$ it follows that

$$x \notin \{y \in \mathbb{R}^n \mid \alpha(y) \le 0 \text{ for all } \alpha \in S^*\}.$$

This gives

$$\{y \in \mathbb{R}^n \mid \alpha(y) \le 0 \text{ for all } \alpha \in S^*\} \subseteq \mathrm{cl}(\mathrm{conv}(S)),$$

so proving this part of the theorem.

(ii) Let $x \in \mathrm{conv}(S \cup \{0\})$. By Proposition 1.9.4, there exists $x_1, \dots, x_k \in S \cup \{\mathbf{0}\}$ and $\lambda_1, \dots, \lambda_k \in (0, 1)$ satisfying $\sum_{j=1}^k \lambda_j = 1$ such that $x = \sum_{j=1}^k \lambda_j x_j$. Then, if $a \in S^\circ$,

$$\langle a, x\rangle_{\mathbb{R}^n} = \sum_{j=1}^k \lambda_j \langle a, x_j\rangle_{\mathbb{R}^n} \le 0.$$

Thus

$$\mathrm{conv}(S) \subseteq \{x \in \mathbb{R}^n \mid \langle a, x\rangle_{\mathbb{R}^n} \le 0 \text{ for all } a \in S^\circ\}.$$

If $x \in \mathrm{cl}(\mathrm{conv}(S \cup \{\mathbf{0}\}))$ then let $x = \lim_{j\to\infty} x_j$ for a sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathrm{conv}(S)$. As above, if $a \in S^\circ$, then $\langle a, x_j\rangle_{\mathbb{R}^n} \le 0$ for each $j \in \mathbb{Z}_{>0}$. By continuity of the inner product (see) and Theorem 1.3.2, what?

$$\langle a, x\rangle_{\mathbb{R}^n} = \lim_{j\to\infty} \langle a, x_j\rangle_{\mathbb{R}^n} \le 0,$$

giving

$$\mathrm{cl}(\mathrm{conv}(S)) \subseteq \{x \in \mathbb{R}^n \mid \langle a, x\rangle_{\mathbb{R}^n} \le 0 \text{ for all } a \in S^\circ\}.$$

For the converse inclusion, let $x \notin \mathrm{cl}(\mathrm{conv}(S \cup \{\mathbf{0}\}))$. By Corollary 1.9.17 there exists $a' \in \mathbb{R}^n$ and $b' \in \mathbb{R}$ such that

$$\langle a', x\rangle_{\mathbb{R}^n} > b', \quad \langle a', y\rangle_{\mathbb{R}^n} < b', \qquad y \in \mathrm{cl}(\mathrm{conv}(S \cup \{\mathbf{0}\})).$$

Note that $b' \neq 0$ since $\mathbf{0} \in S \cup \{\mathbf{0}\}$. Thus, define $\boldsymbol{a} = \frac{1}{b'}\boldsymbol{a}'$ and note that $\langle \boldsymbol{a}, \boldsymbol{y}\rangle_{\mathbb{R}^n} < 1$ for every $\boldsymbol{y} \in \mathrm{cl}(\mathrm{conv}(S \cup \{\mathbf{0}\}))$ and that $\langle \boldsymbol{a}, \boldsymbol{x}\rangle_{\mathbb{R}^n} > 1$. Thus $\boldsymbol{a} \in S^\circ$ and

$$\boldsymbol{x} \notin \{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{a}, \boldsymbol{y}\rangle_{\mathbb{R}^n} \leq 1 \text{ for all } \boldsymbol{a} \in S^\circ\}.$$

Thus

$$\{\boldsymbol{y} \in \mathbb{R}^n \mid \langle \boldsymbol{a}, \boldsymbol{y}\rangle_{\mathbb{R}^n} \leq 1 \text{ for all } \boldsymbol{a} \in S^\circ\} \subseteq \mathrm{cl}(\mathrm{conv}(S \cup \{\mathbf{0}\})),$$

as desired. ■

The second part of the result immediately gives the following result, illustrating the importance to the polar of the origin being contained in a convex set.

1.9.40 Corollary (The polar of the polar is identity (sometimes)) *If* $\mathsf{C} \subseteq \mathbb{R}^n$ *is a closed convex set for which* $\mathbf{0} \in \mathsf{C}$*, then* $\mathsf{C}^{\circ\circ} = \mathsf{C}$*.*

Now, understanding a few fundamental properties of polars and polar cones, let us give some examples of these objects to get some insight into how the dual constructions work.

1.9.41 Examples (Polars and polar cones)

1. The reader may verify in Exercise 1.9.10 that $\{\mathbf{0}\}^* = \mathbb{R}^n \times \mathbb{R}_{\leq 0}$ and $\{\mathbf{0}\}^\circ = \mathbb{R}^n$ and that $(\mathbb{R}^n)^* = \{\mathbf{0}\} \times \mathbb{R}_{\leq 0}$ and $(\mathbb{R}^n)^\circ = \{\mathbf{0}\}$. We identify $A(\mathbb{R}^n; \mathbb{R})$ with $\mathbb{R}^n \oplus \mathbb{R}$ in the above expressions.
2. Let $\boldsymbol{x}_0 \in \mathbb{R}^n$. Let us first describe $\{\boldsymbol{x}_0\}^\circ$. If $\boldsymbol{x}_0 = \mathbf{0}$ then we already have $\mathsf{P} = \mathbb{R}^n$, so let us suppose that $\boldsymbol{x}_0 \neq \mathbf{0}$. We first note that $\{\boldsymbol{x}_0\}^* \cap (\mathbb{R}^n \times \{b\})$ is in bijective correspondence with the half-space

$$\{\boldsymbol{a} \in \mathbb{R}^n \mid \langle \boldsymbol{a}, \boldsymbol{x}_0\rangle_{\mathbb{R}^n} \leq -b\},$$

by definition. Let us describe this half space explicitly, first by describing its bounding hyperplane

$$\mathsf{P} = \{\boldsymbol{a} \in \mathbb{R}^n \mid \langle \boldsymbol{a}, \boldsymbol{x}_0\rangle_{\mathbb{R}^n} = -b\}.$$

By Proposition I-5.4.49 we have that P is an affine subspace whose linear part is $\{\boldsymbol{x}_0\}^\perp$. Thus, to determine P, it suffices to find one vector in P. Since $\boldsymbol{x}_0 \neq \mathbf{0}$ we have $-b\frac{\boldsymbol{x}_0}{\|\boldsymbol{x}_0\|^2_{\mathbb{R}^n}} \in \mathsf{P}$ and so

$$\mathsf{P} = \left\{-b\frac{\boldsymbol{x}_0}{\|\boldsymbol{x}_0\|^2_{\mathbb{R}^n}}\right\} + \{\boldsymbol{x}_0\}^\perp.$$

From this, one directly characterises $\{\boldsymbol{x}_0\}^*$, and we depict the situation in Figure 1.18. Much insight into polar cones and polars can be gained by thinking about the polar cones and polars of points, so the pictures in Figure 1.18 should be studied a little.

The next few examples will deal with sequentially more complicated ellipsoids.

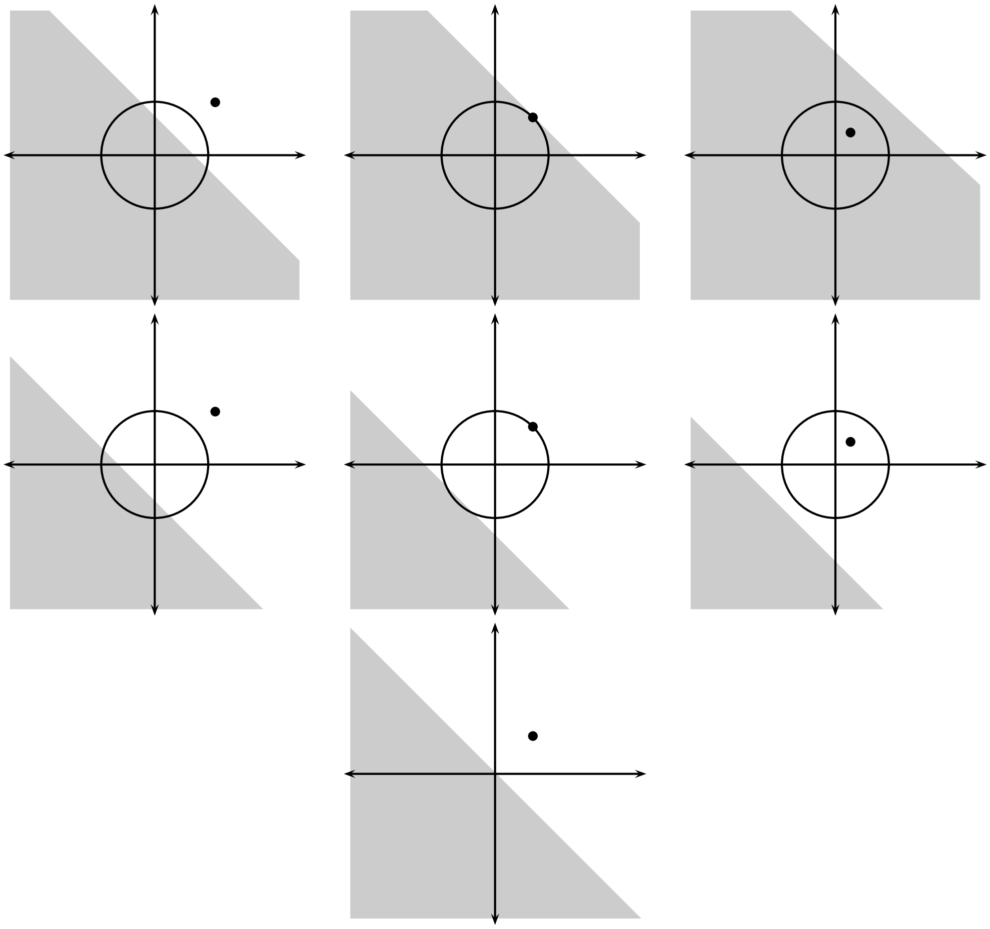

Figure 1.18 We depict $\{x_0\}^* \cap (\mathbb{R}^n \times \{b\})$. In the top row, $b \in \mathbb{R}_{<0}$, in the middle row $b \in \mathbb{R}_{>0}$, and in the bottom row $b = 0$. In the first two rows we have $\|x_0\|_{\mathbb{R}^n} < |b|$ (left), $\|x_0\|_{\mathbb{R}^n} = |b|$ (centre), and $\|x_0\|_{\mathbb{R}^n} > |b|$ (right). The circle is that of radius $|b|$.

3. First let us take (for brevity) $B_r = \overline{\mathsf{B}}^n(r, \mathbf{0})$ for $r \in \mathbb{R}_{>0}$. By Proposition 1.9.38, $(B_r)^*_b = \emptyset$ for $b \in \mathbb{R}_{<0}$. If $b \in \mathbb{R}_{>0}$ then, by direct consideration of example 2 above, $(B_r)^*_b = \overline{\mathsf{B}}^n(\frac{|b|}{r}, \mathbf{0})$, and, in particular, $B_r{}^\circ = \overline{\mathsf{B}}^n(r^{-1}, \mathbf{0})$. Note, then, that $\overline{\mathsf{B}}^n(1, \mathbf{0})^\circ = \overline{\mathsf{B}}^n(1, \mathbf{0})$.
4. Now let us consider an ellipsoid centred at $\mathbf{0}$. Thus we take $E = \boldsymbol{A}(\overline{\mathsf{B}}^n(1, \mathbf{0}))$ for $\boldsymbol{A}$ symmetric and positive-definite. By Proposition 1.9.38, $E^*_b = \emptyset$ for $b \in \mathbb{R}_{<0}$. By Proposition 1.9.37(v) and the preceding example, $E^*_b = \boldsymbol{A}^{-1}(\overline{\mathsf{B}}^n(|b|, \mathbf{0}))$ for $b \in \mathbb{R}_{<0}$. In particular, $E^\circ = \boldsymbol{A}^{-1}(\overline{\mathsf{B}}^n(1, \mathbf{0}))$.
5. Next we consider a general ellipsoid

$$E = \{\boldsymbol{A}\boldsymbol{u} + \boldsymbol{d} \in \mathbb{R}^n \mid \boldsymbol{u} \in \overline{\mathsf{B}}^n(1, \mathbf{0})\}$$

for $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}^n)$ symmetric and positive-definite and for $\boldsymbol{d} \in \mathbb{R}^n$. Let us abbreviate $B = \overline{\mathsf{B}}^n(1, \mathbf{0})$ so that, as above,

$$B_b^* = \begin{cases} \overline{\mathsf{B}}^n(|b|, \mathbf{0}), & b \in \mathbb{R}_{\le 0}, \\ \varnothing, & b \in \mathbb{R}_{>0}. \end{cases}$$

By Proposition 1.9.37(v),

$$E^* = \{(\boldsymbol{A}^{-1}\boldsymbol{a}, b - \langle \boldsymbol{A}^{-1}\boldsymbol{a}, \boldsymbol{d}\rangle_{\mathbb{R}^n}) \in \mathbb{R}^n \mid |\boldsymbol{a}| \le |b|, b \in \mathbb{R}_{\le 0}\}.$$

finish and get polars of cubes from Webster

1.9.9 Convex functions

In Sections I-3.1.6 and I-3.2.6 we considered in some detail the notion of a convex function defined on an interval in $\mathbb{R}$. In this section we extend this to multivariable functions.

We begin with the definition.

1.9.42 Definition (Convex function) For a convex set $C \subseteq \mathbb{R}^n$, a function $f : C \to \mathbb{R}$ is ***convex*** if

$$f((1-s)\boldsymbol{x}_1 + s\boldsymbol{x}_2) \le (1-s)f(\boldsymbol{x}_1) + sf(\boldsymbol{x}_2)$$

for every $\boldsymbol{x}_1, \boldsymbol{x}_2 \in C$ and every $s \in [0,1]$. If $-f$ is convex, then f is ***concave***. •

It turns out that the notion of convexity for single variable functions is, in an appropriate sense, sufficient to determine convexity of a multivariable function.

1.9.43 Proposition (Convexity equals convexity on lines) *Let $C \subseteq \mathbb{R}^n$ be convex. For a function $f : C \to \mathbb{R}$ the following statements are equivalent:*

(i) f is convex;

(ii) if $\mathbf{x}_1, \mathbf{x}_2 \in C$ are distinct, then the function $f_{\mathbf{x}_1,\mathbf{x}_2} : [0, \|\mathbf{x}_2 - \mathbf{x}_1\|_{\mathbb{R}^n}] \to \mathbb{R}$ given by

$$f_{\mathbf{x}_1,\mathbf{x}_2}(x) = f\Big(\Big(1 - \frac{x}{\|\mathbf{x}_2 - \mathbf{x}_1\|_{\mathbb{R}^n}}\Big)\mathbf{x}_1 + \frac{x}{\|\mathbf{x}_2 - \mathbf{x}_1\|_{\mathbb{R}^n}}\mathbf{x}_2\Big)$$

is convex, i.e., f restricted to the line connecting any two points in C is convex.

Proof (i) $\Longrightarrow$ (ii) Let $\boldsymbol{x}_1, \boldsymbol{x}_2 \in C$ and let $x_1, x_2 \in [0, \|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}]$. If $s \in [0,1]$ then

$$\begin{aligned} f_{\boldsymbol{x}_1,\boldsymbol{x}_2}((1-s)x_1 + sx_2) &= f\Big(\Big(1 - \frac{(1-s)x_1 + sx_2}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}}\Big)\boldsymbol{x}_1 + \frac{(1-s)x_1 + sx_2}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}}\boldsymbol{x}_2\Big) \\ &= f\Big((1-s)\Big(\Big(1 - \frac{x_1}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}}\Big)\boldsymbol{x}_1 + \frac{x_1}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}}\boldsymbol{x}_2\Big) \\ &\quad + s\Big(\Big(1 - \frac{x_2}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}}\Big)\boldsymbol{x}_1 + \frac{x_2}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}}\boldsymbol{x}_2\Big)\Big) \\ &\le (1-s)f\Big(1 - \Big(\frac{x_1}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}}\Big)\boldsymbol{x}_1 + \frac{x_1}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}}\boldsymbol{x}_2\Big) \\ &\quad + sf\Big(\Big(1 - \frac{x_2}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}}\Big)\boldsymbol{x}_1 + \frac{x_2}{\|\boldsymbol{x}_2 - \boldsymbol{x}_1\|_{\mathbb{R}^n}}\boldsymbol{x}_2\Big) \\ &= (1-s)f_{\boldsymbol{x}_1,\boldsymbol{x}_2}(x_1) + sf_{\boldsymbol{x}_1,\boldsymbol{x}_2}(x_2), \end{aligned}$$

as desired.

(ii) $\Longrightarrow$ (i) Let $x_1, x_2 \in C$ and $s \in [0,1]$. Then

$$\begin{aligned} f((1-s)x_1 + sx_2) &= f\Big(\Big(1 - \frac{s\|x_2 - x_1\|_{\mathbb{R}^n}}{\|x_2 - x_1\|_{\mathbb{R}^n}}\Big)x_1 + \frac{s\|x_2 - x_1\|_{\mathbb{R}^n}}{\|x_2 - x_1\|_{\mathbb{R}^n}}x_2\Big) \\ &= f_{x_1,x_2}((1-s)0 + s\|x_2 - x_1\|_{\mathbb{R}^n}) \\ &\leq (1-s)f_{x_1,x_2}(0) + sf_{x_1,x_2}(\|x_2 - x_1\|_{\mathbb{R}^n}) \\ &= (1-s)f(x_1) + sf(x_2), \end{aligned}$$

as desired. ■

As with convex functions defined on intervals, multivariable convex functions are continuous.

1.9.44 Proposition (Convex functions are continuous) *If* $C \subseteq \mathbb{R}^n$ *is open and convex and if* $f\colon C \to \mathbb{R}$ *is convex, then* f *is continuous.*

Proof ■

1.9.10 Integration over convex sets

In this section we give a few definitions and results concerning integration over convex domains. The crucial idea in getting started with this is the following result.

1.9.45 Theorem (Convex sets are Jordan measurable) *If* $C \subseteq \mathbb{R}^n$ *is convex then it is Jordan measurable.*

Proof If $\mathrm{int}(C) = \emptyset$ this means that C is contained in an affine subspace of $\mathbb{R}^n$ and so necessarily has measure zero, essentially by Exercise 1.2.13. Then we suppose that $\mathrm{int}(C) \neq \emptyset$.

We begin assuming that C is bounded. By a scaling argument using the change of variables formula for the Riemann integral we will assume, for simplicity and without loss of generality, that C is contained in the cube R with sides of length one and centred at $\mathbf{0} \in \mathbb{R}^n$. For $k \in \mathbb{Z}_{>0}$ let $\boldsymbol{P}_k$ be the partition of R into cubes whose sides have equal length 2^{-k}. There are thus 2^k subcubes from R. Let $\mathscr{R}_0$ be the subcubes from $\boldsymbol{P}$ that intersect the interior of C and let $\mathscr{R}_0$ be the subcubes of $\boldsymbol{P}$ that intersect the boundary of C. A ***vertex*** of R is a point in R having all components equal to $\pm\frac{1}{2}$. There are 2^n vertices and we denote the collection of vertices by $V(R)$. A ***face*** of R is given by the collection of all points in R where one of the components has a fixed value of $\pm\frac{1}{2}$. For a given vertex v there are n faces which contain v and we denote the set of faces containing v by $F_v(R)$. Let $v \in V(R)$ and let R' be a subcube of $\boldsymbol{P}_k$ intersecting one of the n faces containing v and let $x' \in R'$ be the centre of R'. Since there are n faces containing v and since each face intersects $2^{k(n-1)}$ subcubes, there are $n2^{k(n-1)}$ such points x'. Now consider the ray $\rho_{x',v}\colon \mathbb{R}_{\geq 0} \to \mathbb{R}^n$

$$\rho_{x',v}(s) = x' - sv$$

emanating from $\boldsymbol{x}'$ in the direction of $-\boldsymbol{v}$. By considering all such rays associated with all vertices we have a total of $n2^n 2^{k(n-1)}$ rays constructed in this way. Let us denote by $\rho(R)$ the collection of these rays. The following lemma records a useful property of the rays constructed in this way.

1 Lemma *If* R' *is a subcube of* $\mathbf{P}_\mathrm{k}$ *then the set*

$$\{\rho_{\mathbf{x}',\mathbf{v}} \in \rho(\mathrm{V}) \mid \ \mathrm{image}(\rho_{\mathbf{x}',\mathbf{v}}) \textit{ intersects the centre of } \mathrm{R}'\}$$

has cardinality 2^n.

Proof We just sketch the idea, leaving the elementary and notation-filled details to the reader. Let R' be a subcube and let $\boldsymbol{x}_{R'}$ be the centre of R'. For $\boldsymbol{v} \in V(R)$ let $r_{\boldsymbol{v}} \colon \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ be defined by

$$r_{\boldsymbol{v}}(s) = \boldsymbol{x}_{R'} + s\boldsymbol{v},$$

i.e., the ray through $\boldsymbol{x}'$ in the direction of $\boldsymbol{v}$. Since $\boldsymbol{x}_{R'} \in \mathrm{int}(R)$ this ray will intersect the boundary of R and so, before it does, it will pass through the centre $\boldsymbol{x}'$ (why and why the centre?) of one of the subcubes intersecting $F_{\boldsymbol{v}}(R)$. There will be exactly one such subcube for each ray $r_{\boldsymbol{v}}$. Thus the ray $\rho_{\boldsymbol{x}',\boldsymbol{v}}$ will pass through the centre of R', and this holds for every vertex, and so we arrive at the 2^n rays $\rho_{\boldsymbol{x}',\boldsymbol{v}}$ passing through the centre of R'. ▼

For a subcube R' of $\boldsymbol{P}_k$ denote by $\rho(R')$ the 2^n rays passing through the centre of R'. For subcubes $R' \in \mathscr{R}_1$ we have the following observation.

2 Lemma *Let* $\mathrm{R}' \in \mathscr{R}_1$ *and for* $\rho_{\mathbf{x}',\mathbf{v}} \in \rho(\mathrm{R}')$ *let* $\mathrm{s}_1(\rho_{\mathbf{x}',\mathbf{v}}) \in \mathbb{R}_{\geq 0}$ *be such that* $\rho_{\mathbf{x}',\mathbf{v}}(\mathrm{s}_1(\rho_{\mathbf{x}',\mathbf{v}}))$ *is equal to the centre of* R' *and let* $\mathrm{s}_0(\rho_{\mathbf{x}',\mathbf{v}}) \in \mathbb{R}_{\geq 0}$ *be smallest number such that* $\rho_{\mathbf{x}',\mathbf{v}}(\mathrm{s}_0(\rho_{\mathbf{x}',\mathbf{v}}))$ *is equal to the centre of a subcube, distinct from* R'*, from* $\mathscr{R}_0$*. Then the set*

$$\{\rho_{\mathbf{x}'} \in \rho(\mathrm{R}') \mid \ \mathrm{s}_1(\rho_{\mathbf{x}',\mathbf{v}}) < \mathrm{s}_0(\rho_{\mathbf{x}',\mathbf{v}})\}$$

is nonempty.

Proof Suppose the conclusions of the lemma do not hold so that $s_1(\rho_{\boldsymbol{x}',\boldsymbol{v}}) > s_0(\rho_{\boldsymbol{x}',\boldsymbol{v}})$ for every $\rho_{\boldsymbol{x}',\boldsymbol{v}} \in \rho(R')$. Let $R'(\rho_{\boldsymbol{x}',\boldsymbol{v}}) \in \mathscr{R}_0 \setminus \{R'\}$ be the subcube whose centre $\rho_{\boldsymbol{x}',\boldsymbol{v}}$ passes through before it passes through the centre of R'. From Lemma 1 there are 2^n such subcubes. Choose $\boldsymbol{x}(\rho_{\boldsymbol{x}',\boldsymbol{v}}) \in \mathrm{int}(\mathrm{C}) \cap R'(\rho_{\boldsymbol{x}',\boldsymbol{v}})$. We claim that

$$R' \subseteq \mathrm{conv}(\{\boldsymbol{x}(\rho_{\boldsymbol{x}',\boldsymbol{v}}) \mid \ \rho_{\boldsymbol{x}',\boldsymbol{v}} \in \rho(R')\}.$$

Indeed, let $\boldsymbol{x} \in R'$. As we saw in the proof of Lemma 1, the 2^n rays from $\rho(R')$ pass through the centre of R', one in the direction (from $\mathbf{0}$) of each of the vertices. It follows, therefore, that the 2^n points of the form $\boldsymbol{x}(\rho_{\boldsymbol{x}',\boldsymbol{v}})$ lie on rays through the centre of R' in the direction (from $\mathbf{0}$) of one of the vertices. Moreover, each of the points $\boldsymbol{x}(\rho_{\boldsymbol{x}',\boldsymbol{v}})$ does not lie in R'. Therefore, we have the following situation. Let $\mathsf{P}_1, \dots, \mathsf{P}_n$ be the hyperplanes passing through $\boldsymbol{x}$ and parallel to the coordinate hyperplanes (i.e., the hyper planes where exactly one of the coordinates is fixed to be zero). Let H_j^- and H_j^+ be the half-spaces defined by the hyperplane P_j, $j \in \{1, \dots, n\}$. Thus there are 2^n such half-spaces and their intersections partition $\mathbb{R}^n$ into 2^n regions. The points $\boldsymbol{x}(\boldsymbol{x}', \boldsymbol{v})$ have the property that there is exactly one of these in each of the 2^n regions. Using this fact, all we have to do is prove the following statement:

If R^n is partitioned by the 2^n half-spaces defined by the coordinate hyperplanes, and if $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{2^n}$ are chosen so that exactly one of these points lies in the interior of each region, then $\boldsymbol{0} \in \mathrm{conv}(\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{2^n}\})$.

We prove this assertion by induction on n. For $n = 1$ it is trivial since it says that $0 \in \mathbb{R}$ is in the convex hull of two points, one being negative the other being positive. Now suppose the result true for $n = m$ and consider the case when $n = m + 1$. From the collection of 2^{m+1} points $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{2^{m+1}}$ let $\boldsymbol{x}_1^-, \dots, \boldsymbol{x}_{2^m}^-$ be those points for which the first coordinate is negative. Also denote by $\boldsymbol{y}_1, \dots, \boldsymbol{y}_{2^m}$ the points in the first coordinate hyper plane obtained by setting the first component to zero in each of the vectors $\boldsymbol{x}_1^-, \dots, \boldsymbol{x}_{2^m}^-$. By the induction hypothesis,

$$\boldsymbol{0} \in \mathrm{conv}(\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_{2^m}\}).$$

Thus $\mathrm{conv}(\{\boldsymbol{x}_1^-, \dots, \boldsymbol{x}_{2^m}^-\})$ contains a point of the form $(a, 0, \dots, 0)$ for $a < 0$. In a similar manner we denote by $\boldsymbol{x}_1^+, \dots, \boldsymbol{x}_{2^m}^+$ those points for which the first coordinate is positive and show that $\mathrm{conv}(\{\boldsymbol{x}_1^+, \dots, \boldsymbol{x}_{2^m}^+\})$ contains a point of the form $(b, 0, \dots, 0)$ for $b > 0$. Since

$$\mathrm{conv}(\{\{\boldsymbol{x}_1^-, \dots, \boldsymbol{x}_{2^m}^-\}) \cup \mathrm{conv}(\{\boldsymbol{x}_1^+, \dots, \boldsymbol{x}_{2^m}^+\}) \subseteq \mathrm{conv}(\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{2^{m+1}}\})$$

the line segment

$$\{((1-s)a + sb, 0, \dots, 0) \mid s \in [0,1]\}$$

is contained in $\mathrm{conv}(\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{2^{m+1}}\})$ and so $\boldsymbol{0}$ is also contained in this convex hull.

Thus we have shown that if the conclusions of the lemma do not hold then $R' \notin \mathscr{R}_1$, i.e., the hypotheses of the lemma also do not hold. ▼

This shows that the number of subcubes in $\mathscr{R}_1$ cannot exceed $n 2^n 2^{k(n-1)}$. The total volume of these cubes is then bounded above by $2^{-kn}(n 2^n 2^{k(n-1)}) = n 2^n 2^{-k}$. Since this goes to zero as $k \to \infty$ this means that we can cover $\mathrm{bd}(C)$ with cubes whose total volume is arbitrarily small, and so C is Jordan measurable by Theorem 1.6.16.

If C is not bounded than we argue as follows. Consider the countable family of cubes

$$R_j = [-j, j] \times \cdots \times [-j, j], \qquad j \in \mathbb{Z}_{>0},$$

and define $C_j = C \cap R_j$. Then C_j and bounded (obviously) and convex (by Exercise 1.9.3) and so $\mathrm{vol}(\mathrm{bd}(C_j))$. Thus $\mathrm{bd}(C_j)$ has zero measure. Now note that $\mathrm{bd}(C) \subseteq \cup_{j \in \mathbb{Z}_{>0}} \mathrm{bd}(C_j)$, the latter being a countable union of sets of measure zero, and so having measure zero by Exercise 1.2.12. Thus $\mathrm{bd}(C)$ is contained in a set of measure zero, and so has measure zero. ■

An interesting idea that can now be associated with a bounded convex set is the following.

1.9.46 Definition (Centre of a convex set) If $C \subseteq \mathbb{R}^n$ is a bounded n-dimensional convex set the ***centre*** of C is

$$\boldsymbol{x}_C = \frac{1}{\mathrm{vol}(C)} \int_C \boldsymbol{x} \, d\boldsymbol{x}. \qquad \bullet$$

The following property of the centre gives some insight into how one should interpret it.

1.9.47 Proposition (Property of the centre) *For a bounded* n*-dimensional convex set* C *we have*

$$\int_{\mathrm{C}} (\mathbf{x} - \mathbf{x}_{\mathrm{C}})\,\mathrm{d}\mathbf{x} = \mathbf{0}.$$

Proof This is trivial:

$$\int_C (\boldsymbol{x} - \boldsymbol{x}_C)\,\mathrm{d}\boldsymbol{x} = \int_C \boldsymbol{x}\,\mathrm{d}\boldsymbol{x} - \boldsymbol{x}_C \mathrm{vol}(C) = \boldsymbol{x}_C \mathrm{vol}(C) - \boldsymbol{x}_C \mathrm{vol}(C) = \mathbf{0}. \qquad \blacksquare$$

The idea is that the integral

$$\int_C (\boldsymbol{x} - \boldsymbol{x}_C)\,\mathrm{d}\boldsymbol{x}$$

measures the sums of the moments about $\boldsymbol{x}_0$ of uniformly distributed masses located in the body C. Since these moments sum to zero according to the proposition, this means that $\boldsymbol{x}_0$ should be interpreted as the "centre of balance" of the body. This gives the best interpretation of the barycentre.

1.9.48 Proposition (Barycentre equals centre) *If* $\Delta \subseteq \mathbb{R}^{\mathrm{k}}$ *is an* n*-dimensional simplex then the centre of* Δ *is its barycentre.*

Proof Suppose that $\Delta = \Delta(\boldsymbol{x}_0, \boldsymbol{x}_1, \dots, \boldsymbol{x}_n)$ and denote $\boldsymbol{y}_j = \boldsymbol{x}_j - \boldsymbol{x}_0$ for $j \in \{1, \dots, n\}$. Since Δ is a translation by $\boldsymbol{x}_0$ of the simplex $\Delta(\mathbf{0}, \boldsymbol{y}_1, \dots, \boldsymbol{y}_n)$, by the change of variables formula (why?) we can suppose without loss of generality that $\boldsymbol{x}_0 = \mathbf{0}$. That is to say, we take Δ to be an n-simplex of the form $\Delta(\mathbf{0}, \boldsymbol{y}_1, \dots, \boldsymbol{y}_n)$ for linearly independent vectors $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_n\}$.

Define a linear map $\mathsf{L}_\Delta \colon \mathbb{R}^n \to \mathbb{R}^n$ by asking that $\mathsf{L}_\Delta(\boldsymbol{e}_j) = \boldsymbol{y}_j$, $j \in \{1, \dots, n\}$. This indeed defines a linear map by Theorem I-4.5.24, and is also invertible since $\mathrm{rank}(\mathsf{L}_\Delta) = n$. We claim that L_Δ maps the standard n-simplex bijectively onto Δ. Indeed, a point in

$$\sum_{j=1}^{n} x_j \boldsymbol{e}_j, \qquad x_1, \dots, x_n \in \mathbb{R}_{\geq 0},\ \sum_{j=1}^{n} x_j \leq 1$$

in Δ_n is mapped to the point

$$\sum_{j=1}^{n} x_j \boldsymbol{y}_j \in \Delta.$$

Therefore, by the change of variables formula,

$$\int_\Delta \boldsymbol{y}\,\mathrm{d}\boldsymbol{y} = \det \mathsf{L}_\Delta \int_{\Delta_n} \mathsf{L}_\Delta(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = (\det \mathsf{L}_\Delta)\mathsf{L}_\Delta\Big(\int_{\Delta_n} \boldsymbol{x}\,\mathrm{d}\boldsymbol{x}\Big), \tag{1.53}$$

using linearity of the integral for the last equality.

Let us compute some integrals.

1 Lemma $\int_{\Delta_n} dx = \frac{1}{n!}$.

Proof We prove this by induction on n. For $n = 1$ we have

$$\int_{\Delta_1} d\boldsymbol{x} = \int_0^1 dx_1 = 1,$$

which gives the result in this case. Now suppose the result holds for $n = k - 1$ and consider the integral for $n = k$. Let us partition the coordinates for $\mathbb{R}^k$ as

$$((x_1, \dots, x_{k-1}), x_k) = (\boldsymbol{x}', x_k) \in \mathbb{R}^{k-1} \times \mathbb{R}.$$

For $\xi \in [0, 1]$ define the hyperplane

$$\mathsf{P}(\xi) = \{((x_1, \dots, x_{k-1}), x_k) \mid x_k = \xi\}.$$

Then take $\Delta_k(\xi) = \Delta_k \cap \mathsf{P}(\xi)$, noting that

$$((x_1, \dots, x_{k-1}), x_k) = (\boldsymbol{x}', x_k) \in \Delta_k(\xi) \quad \Longleftrightarrow \quad x_k = \xi,\ x_1 + \dots + x_{k-1} = 1 - \xi.$$

Thus the points in $\Delta_k(\xi)$ are in 1–1 correspondence with a copy of Δ_{k-1} scaled by the factor $1 - \xi$. By the change of variables formula we have

$$\int_{\Delta_{k,j}(\xi)} d\boldsymbol{x}' = (1 - \xi)^{k-1} \mathrm{vol}(\Delta_{k-1}).$$

Thus, by Fubini's Theorem and using the induction hypothesis,

$$\begin{aligned} \int_{\Delta_k} d\boldsymbol{x} &= \int_0^1 \int_{\Delta_k(x_k)} d\boldsymbol{x}' dx_k = \int_0^1 \frac{(1 - x_k)^{k-1}}{(k-1)!} dx_k \\ &= \int_0^1 \frac{\xi^{k-1}}{(k-1)!} d\xi = \frac{1}{k} \frac{1}{(k-1)!} = \frac{1}{k!}, \end{aligned}$$

as desired. ▼

2 Lemma $\int_{\Delta_n} \mathbf{x}\, d\mathbf{x} = \left(\frac{1}{(n+1)!}, \dots, \frac{1}{(n+1)!}\right)$.

Proof The integral we are to compute has a vector argument and so we integrate it componentwise. Let us fix $j \in \{1, \dots, n\}$ and partition the coordinates for $\mathbb{R}^k$ as

$$(x_j, (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k)) = (x_j, \boldsymbol{x}') \in \mathbb{R} \times \mathbb{R}^{k-1}.$$

For $\xi \in [0, 1]$ define the hyperplane

$$\mathsf{P}_j(\xi) = \{(x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k) \mid x_j = \xi\}.$$

Then take $\Delta_{k,j}(\xi) = \Delta_k \cap \mathsf{P}_j(\xi)$, noting that

$$\begin{aligned} &(x_j, (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_k)) \in \Delta_{k,j}(\xi) \\ &\qquad \Longleftrightarrow \quad x_j = \xi,\ x_1 + \dots + x_{j-1} + x_{j+1} \dots + x_k = 1 - \xi. \end{aligned}$$

Thus the points in $\Delta_{k,j}(\xi)$ are in 1–1 correspondence with a copy of Δ_{k-1} scaled by the factor $1-\xi$. By the change of variables formula we have

$$\int_{\Delta_{k,j}(\xi)} \mathrm{d}\boldsymbol{x}' = (1-\xi)^{k-1}\mathrm{vol}(\Delta_{k-1}).$$

Now, by Fubini's Theorem and Lemma 1 we have

$$\begin{aligned}\int_{\Delta_k} x_j\,\mathrm{d}\boldsymbol{x} &= \int_0^1 \int_{\Delta_{k,j}(x_j)} x_j\,\mathrm{d}\boldsymbol{x}'\mathrm{d}x_j = \int_0^1 \frac{x_j(1-x_j)^{k-1}}{(k-1)!}\,\mathrm{d}x_j \\ &= \int_0^1 \frac{(1-\xi)\xi^{k-1}}{(k-1)!}\,\mathrm{d}\xi = \frac{1}{(k-1)!}\frac{1}{k(k+1)} = \frac{1}{(k+1)!},\end{aligned}$$

as desired. ▼

By Theorem 1.6.38 and Lemma 1 we have

$$\det \mathsf{L}_\Delta = \frac{\mathrm{vol}(\mathsf{L}_\Delta(\Delta_n))}{\mathrm{vol}(\Delta_n)} = n!\mathrm{vol}(\Delta).$$

Using this formula and (1.53) we arrive at the formula

$$\frac{1}{\mathrm{vol}(\Delta)}\int_\Delta \boldsymbol{y}\,\mathrm{d}\boldsymbol{y} = n!\mathsf{L}_\Delta\big(\tfrac{1}{(n+1)!},\dots,\tfrac{1}{(n+1)!}\big) = \mathsf{L}_\Delta\big(\tfrac{1}{n+1},\dots,\tfrac{1}{n+1}\big),$$

which gives the result since the last expression is exactly the barycentre of Δ. ■

1.9.11 Convex polyhedra

In this section we describe a particular interesting class of convex sets.

Let us define the various sets in which we are interested. First let us talk about convex polytopes. Notationally it will be convenient, for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, to write $\boldsymbol{x} \le \boldsymbol{y}$ if $x_j \le y_j$, $j \in \{1,\dots,n\}$.

1.9.49 Definition (Convex polyhedron, convex polytope, finitely generated convex cones and sets) A nonempty subset $C \subseteq \mathbb{R}^n$ is

(i) a ***convex polyhedron*** if there exists $\boldsymbol{A} \in \mathrm{L}(\mathbb{R}^n;\mathbb{R}^k)$ and $\boldsymbol{b} \in \mathbb{R}^k$ such that

$$C = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \le \boldsymbol{b}\}$$

and is

(ii) a ***convex polytope*** if it is a compact convex polyhedron.

We also make the following definitions:

(iii) a convex cone K is ***finitely generated*** if $K = \mathrm{conv\,cone}(\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k\},\boldsymbol{x}_0)$ for some $\boldsymbol{x}_0,\boldsymbol{x}_1,\dots,\boldsymbol{x}_k \in \mathbb{R}^n$;

(iv) a convex set C is ***finitely generated*** if $C = \mathrm{conv}(\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k\})$ for some $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k \in \mathbb{R}^n$. •

It is a straightforward exercise to show that a convex polyhedron is, in fact, convex. The reader is invited to check this in Exercise 1.9.11.

Thus a convex polyhedron is the intersection of the solutions to a finite number of linear inequalities, i.e., an intersection of a finite number of half-spaces. Let us denote the half-spaces by $H_1, \dots, H_k$ and the boundary hyperplanes by $\mathsf{P}_1, \dots, \mathsf{P}_k$. Thus

$$C = \mathrm{cl}(H_1) \cap \cdots \cap \mathrm{cl}(H_k).$$

The intersection F_j of a convex polyhedron C with the hyperplane P_j, $j \in \{1, \dots, k\}$, is a ***face of dimension*** $\mathbf{n-1}$. Thus there are as many faces of dimension $n-1$ as there are noncollinear rows in the matrix $\boldsymbol{A}$. Let $F_1, \dots, F_k$ denote the faces of dimension $n-1$. For fixed $j_1, j_2 \in \{1, \dots, k\}$ the set $F_{(j_1,j_2)} = C \cap \mathsf{P}_{j_1} \cap \mathsf{P}_{j_2}$ is a ***face of dimension*** $\mathbf{n-2}$. Thus we can write the set of faces of dimension $n-2$ as $F_{(j_{11},j_{12})}, \dots, F_{(j_{l1},j_{l2})}$ for suitable pairs $(j_{11}, j_{12}), \dots, (j_{l1}, j_{l2}) \in \{1, \dots, k\}^2$. One can proceed in this way, defining faces of dimensions $n-1, n-2, \dots, 1, 0$. A face of dimension 0 is called a ***vertex*** and a face of dimension 1 is sometimes called a ***rib***.

With the preceding discussion as backdrop, let us characterise convex polyhedra and convex polytopes.

1.9.50 Theorem (Farkas[6]–Minkowski[7]–Weyl[8] Theorem) *For a convex set* $\mathrm{C} \subseteq \mathbb{R}^{\mathrm{n}}$ *and a convex cone* $\mathrm{K} \subseteq \mathbb{R}^{\mathrm{n}}$*, the following statements hold:*

(i) C *is a convex polytope if and only if it is finitely generated;*

(ii) K *is a convex polyhedron if and only if it is finitely generated.*

Proof Let us begin with a general discussion. Let $C \subseteq \mathbb{R}^n$ be convex. Let us call a point $\boldsymbol{x} \in C$ an ***extreme*** point of C if, for any line segment

$$\ell_{\boldsymbol{x}_1,\boldsymbol{x}_2} \triangleq \{(1-s)\boldsymbol{x}_1 + s\boldsymbol{x}_2 \mid s \in [0,1]\}$$

contained in C for which $\boldsymbol{x} \in \ell_{\boldsymbol{x}_1,\boldsymbol{x}_2}$, it holds that either $\boldsymbol{x} = \boldsymbol{x}_1$ or $\boldsymbol{x} = \boldsymbol{x}_2$. Thus an extreme point is one not contained in the relative interior of any line segment contained in C. Let $E(C)$ denote the set of extreme points of C.

We now have a lemma.

1 Lemma *If* C *is a compact convex set, then* $\mathrm{C} = \mathrm{conv}(\mathrm{E}(\mathrm{C}))$.

Proof It is clear that $\mathrm{conv}(E(C)) \subseteq C$. The proof of the converse inclusion is by induction on $\dim(C)$. If $\dim(C) = 0$ the result is vacuous. Thus suppose that $\dim(C) = d > 0$ and let us assume without loss of generality that $C \subseteq \mathbb{R}^d$, since otherwise we

[6]Gyula (Julius) Farkas (1847–1930) was a Hungarian mathematician and physicist. His mathematical contributions, apart from the convex analysis contributions to which we refer here, were to function theory. In physics, he also wrote on mechanics, thermodynamics, and electrodynamics.

[7]Hermann Minkowski (1864–1909) was a Russian-born mathematician who spent his professional career in what is now Germany. He made mathematical contributions to were geometric in nature. He gave a mathematical formulation of Einstein's work on special relativity.

[8]Hermann Klaus Hugo Weyl (1885–1955) was a Germaan mathematician whose interests were in algebra and geometry.

can replace $\mathbb{R}^d$ with $\mathrm{aff}(C)$. The induction hypotheses is that all compact convex sets of dimension $\{0, 1, \dots, d-1\}$ are the convex hull of their extreme points.

Let $x \in C$.

If $x \in \mathrm{bd}(C)$ then, by Corollary 1.9.17, $x \in \mathsf{P}$ where P is a hyperplane separating $\{x\}$ and C. Note that $C \cap \mathsf{P}$ is compact and convex (see Exercise 1.9.4) and that $C \cap \mathsf{P} \subseteq \mathrm{bd}(C)$. We also claim that $E(C \cap \mathsf{P}) \subseteq E(C)$. Indeed, suppose that $y \in E(C \cap \mathsf{P})$ and let ℓ_{x_1,x_2} be a line segment in C containing y. We have to show that, if $y \notin \{x_1, x_2\}$, then $y \notin \mathrm{rel\,int}(\ell_{x_1,x_2})$. First note that if $y \in \mathrm{rel\,int}(\ell_{x_1,x-2})$ then $\ell_{x_1,x_2} \subseteq C \cap \mathsf{P}$. Indeed, were this not the case then this would contradict the fact that $y \in C \cap \mathsf{P} \subseteq \mathrm{bd}(C)$. Now, if $\ell_{x_1,x_2} \subseteq C \cap \mathsf{P}$ it follows since $y \in E(C \cap \mathsf{P})$ that $y \notin \mathrm{rel\,int}(\ell_{x_1,x-2}$. This gives $E(C \cap \mathsf{P}) \subseteq E(C)$, as claimed. It follows by the induction hypothesis that $x \in \mathrm{conv}(E(C \cap \mathsf{P})) \subseteq E(C)$.

Now suppose that $x \in \mathrm{int}(C)$. Let $u \in \mathbb{R}^n$ and define

$$L = \{x + su \mid s \in \mathbb{R}\} \cap C$$

as the intersection of a line through x with C. Note that L is compact since C is compact. Then, since C is compact, L is connected and so we have

$$L = \{(1-s)x_1 + sx_2 \mid s \in [0, 1]\}$$

for $x_1, x_2 \in \mathrm{bd}(C)$. From the preceding paragraph, $x_1, x_2 \in \mathrm{conv}(E(C))$, and, since $x \in L$, x is a convex combination of x_1 and x_2. Thus $x \in \mathrm{conv}(E(C))$. ▼

(i) Suppose that

$$C = \mathrm{cl}(H_1) \cap \cdots \cap \mathrm{cl}(H_k)$$

for half-spaces $H_1, \dots, H_k$. If $C = \{x\}$ then C is obviously finitely generated. Thus we assume that C is comprised of more than one point. We claim that if $x \in E(C)$ then there exists $j_1, \dots, j_m \in \{1, \dots, k\}$ such that

$$\{x\} = \mathrm{cl}(H_{j_1}) \cap \cdots \cap \mathrm{cl}(H_{j_m}). \tag{1.54}$$

Indeed, let $x \in E(C)$ and define

$$J(x) = \{j \in \{1, \dots, k\} \mid x \in \mathrm{bd}(H_j)\}$$

and

$$\mathsf{P}(x) = \{y \in \mathbb{R}^n \mid y \in \mathrm{bd}(H_j),\ j \in J(x)\}.$$

Since $x \notin \mathrm{bd}(H_j)$ for $j \notin J(x)$, it follows that $x \in \mathrm{rel\,int}(C \cap \mathsf{P}(x))$. However, since $x \in E(C)$ we also have $x \in E(C \cap \mathsf{P}(x))$. Thus $x \in \mathrm{rel\,bd}(C \cap \mathsf{P}(x))$. The only convex sets for which the relative interior and relative boundary can intersect are the zero dimensional convex sets. Thus $\mathsf{P} = \{x\}$. Thus (1.54) holds. Since there are only finitely many possible intersections of the form $\mathrm{cl}(H_{j_1}) \cap \cdots \cap \mathrm{cl}(H_{j_m})$, it follows that there can only be finitely many extreme points. By the lemma above, C is the convex hull of this finite collection of extreme points, and so is finitely generated.

Next let $C = \mathrm{conv}(\{x_1, \dots, x_k\})$. We shall assume without loss of generality that $\mathrm{aff}(C) = \mathbb{R}^n$. For each $j \in \{1, \dots, k\}$ let $f_j \colon \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be given by $f_j(\alpha) = \alpha(x_j)$. If C^* is the polar cone of C (see Definition 1.9.36) then, since

$$C^* = \{(a, b) \in \mathbb{R}^n \times \mathbb{R} \mid \langle a, x_j\rangle_{\mathbb{R}^n} + b) \le 0,\ j \in \{1, \dots, k\}\}$$

(noting that $A(\mathbb{R}^n;\mathbb{R})$ is isomorphic to $\mathbb{R}^n \oplus \mathbb{R}$), we see that C^* is a polyhedron. Since the function

$$f\colon \alpha \mapsto \sum_{j=1}^{k} \alpha(\boldsymbol{x}_j)$$

is linear, the set $f^{-1}(-1)$ is an affine subspace of $A(\mathbb{R}^n;\mathbb{R})$. Thus $C^* \cap f^{-1}(-1)$ is a polyhedron in the affine subspace $f^{-1}(-1)$. We claim that $C^* \cap f^{-1}(-1)$ is compact. Since we are assuming that $\mathrm{aff}(C) = \mathbb{R}^n$, by Proposition 1.9.23 there exists distinct $j_0, j_1, \dots, j_n \in \{1, \dots, k\}$ such that the set

since intersections of polyhedra are polyhedra and affine subspaces are polyhedra

$$\{\boldsymbol{x}_{j_1} - \boldsymbol{x}_{j_0}, \dots, \boldsymbol{x}_{j_n} - \boldsymbol{x}_{j_0}\}$$

is linearly independent. By Proposition 1.9.26 the map

$$A(\mathbb{R}^n;\mathbb{R}) \ni \alpha \mapsto (\alpha(\boldsymbol{x}_{j_0}), \alpha(\boldsymbol{x}_{j_1}), \dots, \alpha(\boldsymbol{x}_{j_n})) \in \mathbb{R}^{n+1}$$

is a bijection. Moreover, the map is easily seen to be affine. Therefore, the map and its inverse map bounded subsets to bounded subsets. In particular, the preimage of $[-1,0]^{n+1}$ is bounded. Since $C^* \cap f^{-1}(-1)$ is a subset of this preimage, we conclude that it is bounded. Since C^* is closed by Proposition 1.9.38 and $f^{-1}(-1)$ is closed by Corollary 1.3.4, it follows from Proposition 1.2.19 that $C^* \cap f^{-1}(-1)$ is closed, and so compact.

Next we claim that

$$C = \{\boldsymbol{y} \in \mathbb{R}^n \mid \alpha(\boldsymbol{y}) \le 0 \text{ for all } \alpha \in C^* \cap f^{-1}(-1)\}. \tag{1.55}$$

First let $\boldsymbol{x} \in C$. Then if $\alpha \in C^* \cap f^{-1}(-1)$ we have $\alpha(\boldsymbol{x}) \le 0$ by Theorem 1.9.39. Thus

$$C \subseteq \{\boldsymbol{y} \in \mathbb{R}^n \mid \alpha(\boldsymbol{y}) \le 0 \text{ for all } \alpha \in C^* \cap f^{-1}(-1)\}.$$

Next let

$$\boldsymbol{x} \in \{\boldsymbol{y} \in \mathbb{R}^n \mid \alpha(\boldsymbol{y}) \le 0 \text{ for all } \alpha \in C^* \cap f^{-1}(-1)\}.$$

Let $\alpha \in C^* \setminus \{0\}$ and let $\lambda \in \mathbb{R}_{>0}$ be such that $f(\lambda\alpha) = -1$, this being possible since $f(\alpha) \in \mathbb{R}_{<0}$. Thus $\lambda\alpha \in C^* \cap f^{-1}(-1)$. Then we have

$$\alpha(\boldsymbol{x}) = \lambda^{-1}(\lambda\alpha(\boldsymbol{x})) \le 0.$$

Since this holds for every $\alpha \in C^*$ we have $\boldsymbol{x} \in C$ by Theorem 1.9.39. Thus (1.55) holds, as desired.

As we showed above, $C^* \cap f^{-1}(-1)$ is a polytope, and so, according to the first part of the proof, is finitely generated. Thus there exists $\alpha_1, \dots, \alpha_m \in C^* \cap f^{-1}(-1)$ such that

$$C^* \cap f^{-1}(-1) = \mathrm{conv}(\{\alpha_1, \dots, \alpha_k\}).$$

We next claim that

$$C = \{\boldsymbol{y} \in \mathbb{R}^n \mid \alpha_j(\boldsymbol{y}) \le 0,\ j \in \{1, \dots, m\}\}. \tag{1.56}$$

From the previous paragraph,

$$C \subseteq \{\boldsymbol{y} \in \mathbb{R}^n \mid \alpha_j(\boldsymbol{y}) \le 0,\ j \in \{1, \dots, m\}\}.$$

Conversely, let

$$x \in \{y \in \mathbb{R}^n \mid \alpha_j(y) \leq 0,\ j \in \{1, \dots, m\}\}.$$

Let $\alpha \in C^* \cap f^{-1}(-1)$ and write $\alpha = \sum_{j=1}^m \lambda_j \alpha_j$ for $\lambda_j [0, 1]$, $j \in \{1, \dots, m\}$, satisfying $\sum_{j=1}^m \lambda_j = 1$. Then we have

$$\alpha(x) = \sum_{j=1}^{m} \alpha_j(x) \leq 0.$$

Since this holds for every $\alpha \in C^* \cap f^{-1}(-1)$, from the previous paragraph we conclude that $x \in C$. Thus (1.56) indeed holds.

change the above argument to use the polar and not the polar cone

being the intersection of two polyehra

Note that (1.56) gives C as a polytope, as desired.

(ii) Let K be a convex polyhedron with x_0 as vertex. Let C be a convex polytope with $x_0 \in \operatorname{int}(C)$. Note that $K \cap C$ is a polytope and so is finitely generated by part (i) of the theorem. Let us write $K \cap C = \operatorname{conv}(\{x_1, \dots, x_k\})$. We shall show that $K = \operatorname{conv}\operatorname{cone}(\{x_1, \dots, x_k\}, x_0)$. It is evident that

$$\operatorname{conv}\operatorname{cone}(\{x_1, \dots, x_k\}, x_0) \subseteq K$$

since K is a convex cone. Let $x \in K$ and note that there exists $\lambda \in \mathbb{R}_{>0}$ such that $x_0 + \lambda(x - x_0) \in C$ since $x_0 \in \operatorname{int}(C)$. Therefore,

$$x_0 + \lambda(x - x_0) \in \operatorname{conv}(\{x_1, \dots, x_k\}) \subseteq \operatorname{conv}\operatorname{cone}(\{x_1, \dots, x_k\}, x_0).$$

Since

$$x = x_0 + \lambda^{-1}\lambda(x - x_0) \in \operatorname{conv}\operatorname{cone}(\{x_1, \dots, x_k\}, x_0),$$

we have that

$$K = \operatorname{conv}\operatorname{cone}(\{x_1, \dots, x_k\}, x_0),$$

showing that K is finitely generated.

Next suppose that $K = \operatorname{conv}\operatorname{cone}(\{x_1, \dots, x_k\}, x_0)$. Note that, by part (i) of the theorem,

$$\operatorname{conv}(\{x_0, x_1, \dots, x_k\})$$

is the intersection of the closure of half-spaces $H_1, \dots, H_m$. Let

$$J = \{j \in \{1, \dots, m\} \mid x_0 \in \operatorname{bd}(H_j)\}.$$

We shall show that $K = \cap_{j \in J} \operatorname{cl}(H_j)$.

First we claim that $\cap_{j \in J} \operatorname{cl}(H_j)$ is a convex cone with vertex x_0. It is convex, being a polyhedron. To show that it is a cone, let $\lambda_j \in \mathbb{R}^n$ and $a_j \in \mathbb{R}$, $j \in J$, be such that

$$H_j = \{x \in \mathbb{R}^n \mid \langle \lambda_j, x \rangle_{\mathbb{R}^n} > a_j\}$$

for each $j \in J$. Since $x_0 \in \operatorname{bd}(H_j)$, $\langle \lambda_j, x_0 \rangle_{\mathbb{R}^n} = a_j$ for each $j \in J$. Now let $x \in \cap_{j \in J} \operatorname{cl}(H_j)$ and let $\lambda \in \mathbb{R}_{\geq 0}$. Then

$$\langle \lambda_j, x_0 + \lambda(x - x_0) \rangle_{\mathbb{R}^n} = a_j + \lambda(\langle \lambda_j, x \rangle_{\mathbb{R}^n} - a_j) \geq a_j$$

for each $j \in J$, showing that $\cap_{j \in J} \operatorname{cl}(H_j)$ is indeed a convex cone with vertex x_0.

Now, note that we obviously have

$$x_1, \dots, x_k \in \cap_{j \in J} \operatorname{cl}(H_j)$$

and so $K \subseteq \cap_{j \in J} \operatorname{cl}(H_j)$ since $\cap_{j \in J} \operatorname{cl}(H_j)$ is a convex cone with vertex x_0.

For the opposite inclusion, let $x \in \cap_{j \in J} \operatorname{cl}(H_j)$ and note that, for all $\lambda \in \mathbb{R}_{\geq 0}$,

$$x_0 + \lambda(x - x_0)x \in \operatorname{cl}(H_j), \qquad j \in J.$$

Let $j \in \{1, \dots, m\} \setminus J$. Then

$$x_0 + \lambda(x - x_0) = (1 - \lambda)x_0 + \lambda x \in \operatorname{cl}(H_j)$$

for $\lambda \in (0, 1)$. Therefore, for $\lambda \in (0, 1)$ we have

$$x_0 + \lambda(x - x_0) \in \operatorname{cl}(H_j) \subseteq K, \qquad j \in \{1, \dots, m\}.$$

Since

$$x = x_0 + \lambda^{-1}\lambda(x - x_0) \in K$$

thee theorem follows. ■

1.9.12 Linear programming

In this section we provide, just for fun, an application of elementary convex analysis. It turns out that in many applications it is useful to know how to minimise a linear function on a convex polytope (to be defined shortly). This is a well studied problem, going under the general name of ***linear programming***. In this section we shall define all the terminology needed in this problem, and give the main result in linear programming that we shall make use of.

Let us define precisely the fundamental problem of linear programming.

1.9.51 Problem (Linear programming) The ***linear programming problem*** is: For a convex polyhedron

$$C = \{x \in \mathbb{R}^n \mid Ax \leq b\}$$

and for $c \in \mathbb{R}^n$, minimise the function $x \mapsto \langle c, x\rangle_{\mathbb{R}^n}$ over C. A ***solution*** to the linear programming problem is thus a point $x_0 \in C$ such that $\langle c, x_0\rangle_{\mathbb{R}^n} \leq \langle c, x\rangle_{\mathbb{R}^n}$ for every $x \in C$. •

The following result describes the solutions to the linear programming problem.

1.9.52 Theorem (Solutions to linear programming problem) *Let* $\mathbf{c} \in \mathbb{R}^n$*, let*

$$C = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \leq \mathbf{b}\}$$

be a convex polyhedron, and consider the linear programming problem for the function $\mathbf{x} \mapsto \langle \mathbf{c}, \mathbf{x}\rangle_{\mathbb{R}^n}$*. Then the following statements hold:*

(i) *the linear programming problem has a solution if and only if* $\mathbf{x} \mapsto \langle \mathbf{c}, \mathbf{x} \rangle_{\mathbb{R}^n}$ *is bounded below on* C*;*

(ii) *if* C *is a convex polytope then the linear programming problem has a solution;*

(iii) *if* $\mathbf{x} \mapsto \langle \mathbf{c}, \mathbf{x} \rangle_{\mathbb{R}^n}$ *is not constant on* C *then any solution of the linear programming problem lies in* $\mathrm{rel\,bd}(C)$.

Proof We assume, without loss of generality, that the normals of the boundary hyperplanes of the defining half-spaces are not collinear. This amounts to saying that no two rows of the matrix A are collinear. We can also assume, by restricting to the affine hull of C if necessary, that $\mathrm{int}(C) \neq \emptyset$. This simplifies the discussion.

(i) Certainly if the linear programming problem has a solution, then $\boldsymbol{x} \mapsto \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n}$ is bounded below on C. So suppose that $\boldsymbol{x} \mapsto \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n}$ is bounded below on C. Specifically, suppose that $\langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n} \geq -M$ for some $M \in \mathbb{R}_{>0}$. If $\boldsymbol{c} = \boldsymbol{0}$ then the linear programming problem obviously has a solution, indeed many of them. So we suppose that $\boldsymbol{c} \neq \boldsymbol{0}$. Let $\boldsymbol{x}_0 \in C$ and define

$$A = \{\boldsymbol{x} \in \mathbb{R}^n \mid \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n} \in [-M, \langle \boldsymbol{c}, \boldsymbol{x}_0 \rangle_{\mathbb{R}^n}]\}.$$

Note that A is nonempty, closed (since $\boldsymbol{x} \mapsto \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n}$ is continuous), and bounded (since linear functions are proper by). Thus $A \cap C$ is nonempty and compact. The function $\boldsymbol{x} \mapsto \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n}$, restricted to $A \cap C$, therefore achieves its minimum on $A \cap C$ at some point, say $\bar{\boldsymbol{x}}$. It holds that $\langle \boldsymbol{c}, \bar{\boldsymbol{x}} \rangle_{\mathbb{R}^n} \leq \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n}$ for every $\boldsymbol{x} \in C$ since, it clearly holds for $\boldsymbol{x} \in A \cap C$, and if $\boldsymbol{x} \notin A \cap C$ then $\langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n} \geq \langle \boldsymbol{c}, \boldsymbol{x}_0 \rangle_{\mathbb{R}^n}$. Thus the point $\bar{\boldsymbol{x}}$ solves the linear programming problem. what?

(ii) This follows immediately since $\boldsymbol{x} \mapsto \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n}$ is bounded on C if C is a convex polytope.

(iii) That $\boldsymbol{x} \mapsto \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n}$ is not constant on C is equivalent to $\boldsymbol{x} \mapsto \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n}$ not being constant on $\mathrm{aff}(C)$. This is in turn equivalent to the subspace $\mathsf{U}(C)$ of $\mathbb{R}^n$ associated to $\mathrm{aff}(C)$ not being contained in $\boldsymbol{c}^\perp$. Let $\boldsymbol{x} \in \mathrm{rel\,int}(C)$. Let $\boldsymbol{u_c} \in \mathsf{U}(C)$ be the unit vector such that

$$\langle \boldsymbol{u_c}, \boldsymbol{c} \rangle_{\mathbb{R}^n} = \inf\{\langle \boldsymbol{u}, \boldsymbol{c} \rangle_{\mathbb{R}^n} \mid \boldsymbol{u} \in \mathsf{U}(C),\ \|\boldsymbol{u}\|_{\mathbb{R}^n} = 1\}.$$

That such a $\boldsymbol{u_c}$ exists since $\boldsymbol{u} \mapsto \langle \boldsymbol{u}, \boldsymbol{c} \rangle_{\mathbb{R}^n}$ is a continuous function on the compact set $\mathsf{U}(C) \cap \mathbb{S}^{n-1}$. Moreover, since $\langle -\boldsymbol{u}, \boldsymbol{c} \rangle_{\mathbb{R}^n} = -\langle \boldsymbol{u}, \boldsymbol{c} \rangle_{\mathbb{R}^n}$, it follows that $\langle \boldsymbol{u_c}, \boldsymbol{c} \rangle_{\mathbb{R}^n} \in \mathbb{R}_{<0}$. Since $\boldsymbol{x} \in \mathrm{rel\,int}(C)$ there exists $r \in \mathbb{R}_{>0}$ such that $\boldsymbol{x} + r\boldsymbol{u_c} \in C$. Then

$$\langle \boldsymbol{c}, \boldsymbol{x} + r\boldsymbol{u_c} \rangle_{\mathbb{R}^n} = \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n} + r\langle \boldsymbol{c}, \boldsymbol{u_c} \rangle_{\mathbb{R}^n} > \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n},$$

showing that $\boldsymbol{c} \mapsto \langle \boldsymbol{c}, \boldsymbol{x} \rangle_{\mathbb{R}^n}$ cannot achieve its minimum at $\boldsymbol{x} \in \mathrm{rel\,int}(C)$. Thus it must achieve its minimum on $\mathrm{rel\,bd}(C)$. ■

The upshot of the theorem is that in searching for solutions to the linear programming problem, one can restrict one's attention to the boundary of C. Under certain "genericity" assumptions, the search can further be restricted to the vertices of C. This, note, is very significant since it reduces the search for minima to a finite number of computations. Moreover, there is an efficient algorithm, called the ***simplex method***, for searching over the vertices to find the minimum. To describe the simplex method would take us slightly beyond the fun excursion rôle we envision for our presentation of linear programming, so we refer to the notes at the end of the chapter for references.

1.9.13 Notes

The fact that a convex set is Jordan measurable is due to Minkowski [1903], and the proof we give is from the paper of Szabó [1997].

That a finitely generated convex set or cone is polyhedral was proved by Weyl [1934]. The converse statement is proved by Minkowski [1897]. Contributions to the problem are also made by Farkas [1902].

Exercises

1.9.1 Prove that convex sets and cones are connected.

1.9.2 Show that $\mathbb{D}^n = \{x \in \mathbb{R}^{n+1} \mid \|x\|_{\mathbb{R}^n} \leq 1\}$ is convex.
Hint: *Use the triangle inequality.*

1.9.3 If the following statements are true, prove them true. If they are false, give a counterexample to demonstrate this.

(a) The intersection of two convex sets is convex.
(b) The intersection of two cones with vertex x_0 is a cone with vertex x_0.
(c) The union of two convex sets is a convex set.
(d) The union of two cones with vertex x_0 is a cone with vertex x_0.
(e) The intersection of two affine subspaces is an affine subspace.
(f) The union of two affine subspaces is an affine subspace.

1.9.4 Let $(C_a)_{a\in A}$ be a family of convex subsets of $\mathbb{R}^n$ with the property that $\cap_{a\in A} C_a \neq \emptyset$. Show that $\cap_{a\in A} C_a$ is convex.

1.9.5 If $C \subseteq \mathbb{R}^n$ is convex, show that $\operatorname{conv}(C \cup \{x_0\})$ is comprised of the set of lines from x_0 to points in C.

1.9.6 Let $C \subseteq \mathbb{R}^m$ and $D \subseteq \mathbb{R}^n$. Show that $C \times D \subseteq \mathbb{R}^{m+n}$ is convex if and only if both C and D are convex.

1.9.7 Show that the image of a convex set (resp. cone) under an affine map is a convex set (resp. cone).

1.9.8 Prove Proposition 1.9.27.

1.9.9 Prove Proposition 1.9.37.

1.9.10 Show that $\{\mathbf{0}\}^* = \mathbb{R}^n \times \mathbb{R}_{\leq 0}$ and $\{\mathbf{0}\}^\circ = \mathbb{R}^n$ and that $(\mathbb{R}^n)^* = \{\mathbf{0}\} \times \mathbb{R}_{\leq 0}$ and $(\mathbb{R}^n)^\circ = \{\mathbf{0}\}$, where we identify $A(\mathbb{R}^n; \mathbb{R})$ with $\mathbb{R}^n \oplus \mathbb{R}$.

1.9.11 Show that a convex polyhedron is convex.

Section 1.10

Some advanced topics on multivariable functions

In this section we collect some results on multivariable functions that are of a more sophisticated character than the results of Sections 1.3 and 1.4. We choose to collection these results in a separate section in order to minimise the distraction that would arise from their being embedded among the more pedestrian results of the preceding section. It is true, nonetheless, that we will call on results from this section when we need them. It is also the case that reading Sections 1.10.1–1.3.7 will be *very* useful in understanding some of the more abstract material in Chapter III-1.

Do I need to read this section? Obviously this section can be bypassed until specific results are needed in the sequel. It also might be a nice idea to read this section before studying general topology. •

1.10.1 Open, closed, and proper maps

One of the comments one will often hear from students on a first encounter with the statement "a function is continuous if the preimage of open sets are open" is, "Why is it not the *image* of open sets that should be open?" The answer to this question, of course, is that the requirement that preimages are open is right; for example, it agrees with the $\epsilon - \delta$ definition. However, this does lead to the question of whether it is interesting to think about maps sending open sets to open sets. It turns out that there is some value to such maps, although they are far less important than continuous maps.

Let us give the definition.

1.10.1 Definition (Open map, closed map) If $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$, a map $f: A \to B$ is

(i) ***open*** if $f(U)$ is relatively open in B for every relatively open subset U of A, and is

(ii) ***closed*** if $f(C)$ is relatively open in B for every relatively open subset C of A. •

Let us give some examples that discriminate between "open," "closed," and "continuous."

1.10.2 Examples (Open and closed maps) We shall simply state the examples, leaving it to the reader the make the easy verifications of the statements we make.

1. Let $A = \mathbb{Q} \setminus \{0\} \subseteq \mathbb{R}$ and let $B = \mathbb{R}$. Define $f: A \to B$ by

$$f(q) = \begin{cases} q-1, & q \in \mathbb{Q}_{<0}, \\ q+1, & q \in \mathbb{Q}_{>0}. \end{cases}$$

Then f is neither continuous, open, nor closed.

2. Let $A = [0, 1) \subseteq \mathbb{R}$ and let $B = \mathbb{R} \subseteq \mathbb{R}$. The inclusion map $i_A : A \to B$ is continuous, but neither open nor closed.
3. Let $A = \mathbb{R} \subseteq \mathbb{R}$ and let $B = \mathbb{R} \subseteq \mathbb{R}$. The map $f : A \to B$ given by $f(x) = \mathrm{sign}(x)$ (here sign is the signum function defined following Definition I-2.2.8) is closed, but neither continuous nor open.
4. Let us take $A = (-1, 1) \subseteq \mathbb{R}$ and $B = (-1, 0] \cup (1, 2)$. If we define $f : A \to B$ by

$$f(x) = \begin{cases} x, & x \in (-1, 0], \\ x + 1, & x \in (0, 1), \end{cases}$$

then f is open, but neither continuous nor closed.

5. Let $A = \mathbb{R} \subseteq \mathbb{R}$ and $B = \mathbb{R} \subseteq \mathbb{R}$. The function $f : A \to B$ given by $f(x) = \lfloor x \rfloor$ (here $\lfloor \cdot \rfloor$ is the floor function defined following Definition I-2.2.8) is closed and open, but not continuous.
6. Let $A = [0, 1] \subseteq \mathbb{R}$, let $B = \mathbb{R}$, and let $i_A : A \to B$ be the inclusion map. This map is closed and continuous, but not open.
7. Let $A = (0, 1) \subseteq \mathbb{R}$ and $B = \mathbb{R} \subseteq \mathbb{R}$. Then the inclusion map $i_A : A \to B$ is open and continuous, but not closed.
8. Let $A = B = \mathbb{R} \subseteq \mathbb{R}$. The identity map is continuous, closed, and open. •

Now that we have established through examples the impossibility of any general connection between openness, closedness, and continuity of a map, let us examine some connections that *do* exist.

1.10.3 Proposition (Bijective maps are closed if and only if they are open) *If* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *and* $\mathrm{B} \subseteq \mathbb{R}^{\mathrm{m}}$ *then a map* $\mathbf{f} : \mathrm{A} \to \mathrm{B}$ *is open if and only if it is closed.*

Proof Suppose $\boldsymbol{f}$ is an open bijection. Let $C \subseteq A$ be relatively closed so that $A \setminus C$ is relatively open. Then $\boldsymbol{f}(C) = B \setminus \boldsymbol{f}(A \setminus C)$ since $\boldsymbol{f}$ is a bijection. Moreover, $\boldsymbol{f}(A \setminus C)$ is relatively open since $\boldsymbol{f}$ is open. Thus $\boldsymbol{f}(C)$ is relatively closed and so $\boldsymbol{f}$ is closed. A similar argument shows that a closed bijection is open. ■

An important class of open maps are projections.

1.10.4 Proposition (Projections are continuous) *Let* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *and* $\mathrm{B} \subseteq \mathbb{R}^{\mathrm{m}}$*, and define* $\pi_{\mathrm{A}} : \mathrm{A} \times \mathrm{B} \to \mathrm{A}$ *and* $\pi_{\mathrm{B}} : \mathrm{A} \times \mathrm{B} \to \mathrm{B}$ *by*

$$\pi_{\mathrm{A}}(\mathbf{x}, \mathbf{y}) = \mathbf{x}, \quad \pi_{\mathrm{B}}(\mathbf{x}, \mathbf{y}) = \mathbf{y}.$$

Then π_{A} *and* π_{B} *are continuous.*

Proof Let $U \subseteq A \times B$ be relatively open. Then, for $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in U$ there exists $\epsilon \in \mathbb{R}_{>0}$ such that $\mathsf{B}^{n+m}(2\epsilon, (\boldsymbol{x}_0, \boldsymbol{y}_0)) \cap (A \times B) \subseteq U$. By (1.9) this means that $(\mathsf{B}^n(\epsilon, \boldsymbol{x}_0) \times \mathsf{B}^m(\epsilon, \boldsymbol{y}_0)) \cap (A \times B) \subseteq U$. Therefore, $\mathsf{B}^n(\epsilon, \boldsymbol{x}_0) \subseteq \pi_A(U)$ and so $\pi_A(U)$ is open. ■

We have already seen that open maps may not be closed. It is further true that projections, which are open, need not be closed.

1.10.5 Example (Projections are not generally closed) Let $A = \mathbb{R} \subseteq \mathbb{R}$ and $B = \mathbb{R} \subseteq \mathbb{R}$ and consider the projection $\pi_A(x, y) = x$. We claim that π_A is not closed. Indeed, the subset

$$C = \{(x, \tan^{-1}(x)) \mid x \in (-\tfrac{\pi}{2}, \tfrac{\pi}{2})\} \subseteq A \times B$$

which is closed. We have $\pi(C) = (-\frac{\pi}{2}, \frac{\pi}{2})$ which is open. •

Possibly the best characterisations of open and closed maps require some methodology that is most naturally presented in the general context of general topology. Thus we postpone these nice characterisations until . what?

We now turn to a notion that can be quite useful, although it is not as easy to see why until one actually needs to use it.

1.10.6 Definition (Proper map) Let $A \subseteq \mathbb{R}^n$ and $B \subseteq \mathbb{R}^m$. A map $\boldsymbol{f}\colon A \to B$ is ***proper*** if $\boldsymbol{f}^{-1}(K)$ is a compact subset of A for every compact subset K of B. •

Before we get to some examples which exhibit the notion of properness, let us give an alternative characterisation of properness. In order to do this we introduce the following notion about sequences.

1.10.7 Definition (Sequence diverging to infinity) If $A \subseteq \mathbb{R}^n$ a sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ ***diverges to infinity*** if, for every compact subset $K \subseteq A$, there exists $N \in \mathbb{Z}_{>0}$ such that $\boldsymbol{x}_j \in A \setminus K$ for all $j \geq N$. •

We then have the following characterisation of a proper map.

1.10.8 Proposition (Characterisation of proper maps) *If $A \subseteq \mathbb{R}^n$, if $B \subseteq \mathbb{R}^m$, and if $\mathbf{f}\colon A \to B$, then the following statements are equivalent:*

(i) *$\mathbf{f}$ is proper;*

(ii) *for every sequence $(\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ diverging to infinity, the sequence $(\mathbf{f}(\mathbf{x}_j))_{j\in\mathbb{Z}_{>0}}$ diverges to infinity;*

(iii) *if a sequence $(\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ has the property that the sequence $(\mathbf{f}(\mathbf{x}_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\mathbf{y}_0 \in B$, then there exists a subsequence $(\mathbf{x}_{j_k})_{k\in\mathbb{Z}_{>0}}$ converging to a point $\mathbf{x}_0 \in A$ for which $\mathbf{f}(\mathbf{x}_0) = \mathbf{y}_0$.*

Proof (i) $\Longrightarrow$ (ii) Suppose there exists a sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ diverging to infinity for which $(\boldsymbol{f}(\boldsymbol{x}_j))_{j\in\mathbb{Z}_{>0}}$ does not diverge to infinity. Thus there is a compact set $K \subseteq B$ such that for any $N \in \mathbb{Z}_{>0}$ there exists some $j \geq N$ such that $\boldsymbol{f}(\boldsymbol{x}_j) \in K$. In other words, there exists a subsequence $(\boldsymbol{f}(\boldsymbol{x}_{j_k}))_{k\in\mathbb{Z}_{>0}}$ which is a subset of K. The subsequence $(\boldsymbol{x}_{j_k})_{k\in\mathbb{Z}_{>0}}$ still diverges to infinity and so every subsequence of $(\boldsymbol{x}_{j_k})_{k\in\mathbb{Z}_{>0}}$ also diverges to infinity. Therefore, there can be no compact set containing the subsequence $(\boldsymbol{x}_{j_k})_{k\in\mathbb{Z}_{>0}}$. In particular, $\boldsymbol{f}^{-1}(K)$ is not compact and so $\boldsymbol{f}$ is not proper.

(ii) $\Longrightarrow$ (i) Suppose that $\boldsymbol{f}$ is not proper and let $K \subseteq B$ be such that $\boldsymbol{f}^{-1}(K)$ is not compact. Thus there exists a sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in $\boldsymbol{f}^{-1}(K)$ having no convergent subsequence. This means that for any compact subset $K' \subseteq \boldsymbol{f}^{-1}(K)$ and for any $N \in \mathbb{Z}_{>0}$ there exists $j \geq N$ such that $\boldsymbol{x}_j \notin K'$. (Indeed, were this not the case, then there would exist a compact set K' containing $\boldsymbol{x}_j$ for all sufficiently large j, and this would imply the existence of a convergent subsequence.) This implies that there is a subsequence $(\boldsymbol{x}_{j_k})_{k\in\mathbb{Z}_{>0}}$ which diverges to infinity. But the sequence $(\boldsymbol{f}(\boldsymbol{x}_{j_k}))_{k\in\mathbb{Z}_{>0}}$ is then a sequence in K, and so does not diverge to infinity. ■

This allows a convenient and insightful characterisation of properness for a certain sort of function.

1.10.9 Corollary (Properness of $\mathbb{R}$-valued functions on $\mathbb{R}^n$) *A map* $f\colon \mathbb{R}^n \to \mathbb{R}$ *is proper if and only if* $\lim_{\|\mathbf{x}\|_{\mathbb{R}^n}\to\infty}|f(\mathbf{x})| = \infty$.

Proof First suppose that f is proper and let $(\boldsymbol{x}_j)_{j\in\infty}$ be a sequence such that the sequence $(\|\boldsymbol{x}_j\|_{\mathbb{R}^n})_{j\in\mathbb{Z}_{>0}}$ diverges to ∞ (in the sense of Definition I-2.3.2). Then the sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ clearly diverges to infinity (in the sense of Definition 1.10.7). By Proposition 1.10.8 it follows that $(f(\boldsymbol{x}_j))_{j\in\infty}$ diverges to infinity (in the sense of Definition 1.10.7). Thus, for any $M \in \mathbb{R}_{>0}$ there exists $N \in \mathbb{Z}_{>0}$ so that $f(\boldsymbol{x}_j) \notin [-M, M]$ for all $j \geq N$. Thus $\lim_{\|\boldsymbol{x}\|_{\mathbb{R}^n}\to\infty}|f(\boldsymbol{x})| = \infty$, as desired.

Conversely, suppose that $\lim_{\|\boldsymbol{x}\|_{\mathbb{R}^n}\to\infty}|f(\boldsymbol{x})| = \infty$. This means that for every sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ for which the sequence $(\|\boldsymbol{x}_j\|_{\mathbb{R}^n})_{j\in\mathbb{Z}_{>0}}$ converges to ∞ (in the sense of Definition I-2.3.2) the sequence $(|f(\boldsymbol{x}_j)|)_{j\in\mathbb{Z}_{>0}}$ diverges to ∞ (in the sense of Definition I-2.3.2). Now let $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{R}^n$ converging to infinity (in the sense of Definition 1.10.7). Such a sequence also clearly diverges to ∞ in the sense of Definition I-2.3.2. Now let $K \subseteq \mathbb{R}$ be compact. Take $M \in \mathbb{R}_{>0}$ sufficiently large that $K \in [-M, M]$. Since the sequence $(|f(\boldsymbol{x}_j)|)_{j\in\mathbb{Z}_{>0}}$ diverges to ∞ in the sense of Definition I-2.3.2 let $N \in \mathbb{Z}_{>0}$ be sufficiently large that $|f(\boldsymbol{x}_j)| \geq M$ for $j \geq N$. Then $f(\boldsymbol{x}_j) \notin K$ for $j \geq N$, and so the sequence $(f(\boldsymbol{x}_j))_{j\in\infty}$ diverges to infinity in the sense of Definition 1.10.7. Thus f is proper. ■

Let us give some examples of proper maps.

1.10.10 Examples (Proper maps)

1. By Corollary 1.10.9 the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = \mathrm{Ev}_{\mathbb{R}}(P)(x)$ is proper for any polynomial $P \in \mathbb{R}[\xi]$. That is to say (for those not wanting algebra baggage at a moment like this), polynomial functions are proper.
2. By Corollary 1.10.9 the function $f\colon \mathbb{R} \to \mathbb{R}$ given by $f(x) = \tan^{-1}(x)$ is not proper.
3. Properness and continuity are not generally connected. For example, the function $f\colon \mathbb{R} \to \mathbb{R}$ given by
$$f(x) = \begin{cases} x+1, & x \geq 0, \\ x-1, & x < 0 \end{cases}$$
is proper and discontinuous, while the function $g\colon \mathbb{R} \to \mathbb{R}$ defined by $g(x) = \mathrm{sign}(x)$ is not proper and discontinuous. •

1.10.11 Proposition (Properties of proper maps) *For* $A \subseteq \mathbb{R}^n$*, for* $B \subseteq \mathbb{R}^m$*, and for* $\mathbf{f}\colon A \to B$*, the following statements hold:*

(i) if A *is compact and* $\mathbf{f}$ *is continuous then* $\mathbf{f}$ *is proper and closed;*

(ii) if $\mathbf{f}$ *is continuous and proper then it is closed.*

Proof (i) If f is continuous and $K \subseteq B$ is compact then $f^{-1}(K)$ is relatively closed by Exercise 1.3.3. Therefore, $f^{-1}(K) = C \cap A$ for a closed subset C by Proposition 1.2.50. Thus $f^{-1}(K)$ is closed (being the intersection of closed sets) and so compact by Corollary 1.2.36. Thus f is proper.

Also let $C \subseteq A$ be relatively closed, and so C is compact by Corollary 1.2.36. Therefore, $f(C)$ is compact by Proposition 1.3.29, and so closed. Thus f is closed.

(ii) Let $C \subseteq A$ be relatively closed. Then $C = C' \cap A$ for a closed subset C' of $\mathbb{R}^n$. Let $K \subseteq B$ be compact so that $f^{-1}(K)$ is compact by properness of f. Thus $C \cap f^{-1}(K) = C' \cap f^{-1}(K)$ is compact by Exercise 1.2.4 and so

$$f(C \cap f^{-1}(K)) = f(C) \cap f(f^{-1}(K)) = f(C) \cap K$$

(why the last equality?) is compact since f is continuous. Therefore, by Proposition 1.3.29 it follows that $f(C)$ is closed since K is an arbitrary compact set. ■

1.10.2 Semicontinuity

In this section we study useful generalisations of the notion of continuity. The definitions and their characterisations are simple enough. The difficult thing is to accept why the concepts are worth introducing, and we give a glimpse into this via Proposition 1.10.17 below.

But first let us give the definitions.

1.10.12 Definition (Upper semicontinuous, lower semicontinuous) Let $A \subseteq \mathbb{R}^n$ and let $f\colon A \to \overline{\mathbb{R}}$.

(i) The map f is ***upper semicontinuous*** at $\boldsymbol{x}_0 \in A$ if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists a relative neighbourhood U of $\boldsymbol{x}_0$ in A such that $f(\boldsymbol{x}) < f(\boldsymbol{x}_0) + \epsilon$ for all $\boldsymbol{x} \in U$.

(ii) The map f is ***lower semicontinuous*** at $\boldsymbol{x}_0 \in A$ if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists a relative neighbourhood U of $\boldsymbol{x}_0$ in A such that $f(\boldsymbol{x}) > f(\boldsymbol{x}_0) - \epsilon$ for all $\boldsymbol{x} \in U$.

(iii) The map f is ***upper semicontinuous*** if it is upper semicontinuous at each $\boldsymbol{x} \in A$.

(iv) The map f is ***lower semicontinuous*** if it is lower semicontinuous at each $\boldsymbol{x} \in A$. •

Before we get to trying to understand these notions of semicontinuity and give illustrative examples, let us first give some useful equivalent characterisations. First we give the results for upper semicontinuity.

1.10.13 Proposition (Characterisations of upper semicontinuity) *For* $A \subseteq \mathbb{R}^n$ *and for* $f\colon A \to \overline{\mathbb{R}}$ *the following statements are equivalent:*

(i) f *is upper semicontinuous;*

(ii) $-f$ *is lower semicontinuous;*

(iii) $f^{-1}([-\infty, \alpha))$ *is relatively open in* A *for every* $\alpha \in \mathbb{R}$*;*

(iv) $f^{-1}([\alpha, \infty])$ *is relatively closed in* A *for every* $\alpha \in \mathbb{R}$*;*

(v) *for every* $\mathbf{x}_0 \in A$*, if* $(\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ *is a sequence in* A *converging to* $\mathbf{x}_0 \in A$*, then* $\limsup_{j\to\infty} f(\mathbf{x}_j) \leq f(\mathbf{x}_0)$*.*

Proof (i) $\iff$ (ii) This follows immediately from the definitions.

(i) $\implies$ (iii) Let $\alpha \in \mathbb{R}$ and let $x_0 \in f^{-1}([-\infty, \alpha))$. Let $\epsilon \in \mathbb{R}_{>0}$ be such that $[-\infty, f(x_0)+\epsilon) \subseteq [-\infty, \alpha)$, this being possible by openness of $[-\infty, \alpha)$. By upper semicontinuity of f at x_0 there exists $\delta \in \mathbb{R}_{>0}$ such that $f(x) < f(x_0) + \epsilon$ for all $x \in A \cap \mathsf{B}^n(\delta, x_0)$. Thus $f(x) \in [-\infty, f(x_0) + \epsilon)$ and so $x \in f^{-1}([-\infty, \alpha))$. This gives the desired openness of $f^{-1}([-\infty, \alpha))$.

(iii) $\iff$ (iv) This is clear since $f^{-1}([\alpha, \infty]) = A \setminus f^{-1}([-\infty, a))$.

(iii) $\implies$ (v) Let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in A converging to $x_0 \in A$. If $f(x_0) = \infty$ the result is trivially true, so we may as well suppose that $f(x_0) \in \mathbb{R}$. Let $\alpha > f(x_0)$. By assumption the set $f^{-1}([-\infty, \alpha))$ is open. Moreover, $x_0 \in f^{-1}([-\infty, \alpha))$. Openness of $f^{-1}([-\infty, \alpha))$ and convergence of $(x_j)_{j\in\mathbb{Z}_{>0}}$ to x_0 ensures that there exists $N \in \mathbb{Z}_{>0}$ such that $x_j \in f^{-1}([-\infty, \alpha))$ for all $j \geq N$. Therefore,

$$\begin{aligned}\limsup_{j\to\infty} f(x_j) &= \inf\{\sup\{f(x_j) \mid j \geq k\} \mid k \in \mathbb{Z}_{>0}\}\\ &\leq \sup\{f(x_j) \mid j \geq N\} \leq \alpha.\end{aligned}$$

Thus $\limsup_{j\to\infty} f(x_j) \leq \alpha$ for every $\alpha > f(x_0)$. Thus $\limsup_{j\to\infty} f(x_j) \leq f(x_0)$ as desired.

(v) $\implies$ (i) Suppose that f is not upper semicontinuous at $x_0 \in A$. Then there exists $\epsilon \in \mathbb{R}_{>0}$ such that for any relative neighbourhood U of x_0 in A there exists $x \in U$ such that $f(x) \geq f(x_0)+\epsilon$. Now let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in A converging to x_0. It, therefore, follows that for any $N \in \mathbb{Z}_{>0}$ there exists $j \geq N$ such that $f(x_j) \geq f(x_0) + \epsilon$. Therefore,

$$\liminf_{j\to\infty} f(x_j) \geq f(x_0) + \epsilon > f(x_0),$$

as desired. ■

The following characterisation of lower semicontinuity is proved by a suitable modification of the preceding proof.

1.10.14 Proposition (Characterisations of lower semicontinuity) *For* $A \subseteq \mathbb{R}^n$ *and for* $f\colon A \to \overline{\mathbb{R}}$ *the following statements are equivalent:*

(i) f *is lower semicontinuous;*

(ii) $-f$ *is upper semicontinuous;*

(iii) $f^{-1}((\alpha, \infty]))$ *is relatively open in* A *for every* $\alpha \in \mathbb{R}$*;*

(iv) $f^{-1}([-\infty, \alpha]))$ *is relatively closed in* A *for every* $\alpha \in \mathbb{R}$;

(v) *for every* $\mathbf{x}_0 \in A$, *if* $(\mathbf{x}_j)_{j \in \mathbb{Z}_{>0}}$ *is a sequence in* A *converging to* $\mathbf{x}_0$, *then* $\liminf_{j\to\infty} f(\mathbf{x}_j) \geq f(\mathbf{x}_0)$.

Let us also state the following useful joint property of upper and lower semicontinuity.

1.10.15 Proposition (Continuity equals upper and lower semicontinuity) *If* $A \subseteq \mathbb{R}^n$ *then a map* $f\colon A \to \mathbb{R}$ *is continuous at* $\mathbf{x}_0 \in A$ *if and only if it is both upper and lower semicontinuous at* $\mathbf{x}_0$.

Proof First suppose that f is continuous at $\boldsymbol{x}_0$. Let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta \in \mathbb{R}_{>0}$ such that if $\boldsymbol{x} \in A \cap \mathsf{B}^n(\delta, \boldsymbol{x}_0)$ then $f(\boldsymbol{x}) \in \mathsf{B}^1(\epsilon, f(\boldsymbol{x}_0))$. Then $f(\boldsymbol{x}) < f(\boldsymbol{x}_0) + \epsilon$ and $f(\boldsymbol{x}) > f(\boldsymbol{x}_0) - \epsilon$ for every $\boldsymbol{x} \in \mathsf{B}^n(\delta, \boldsymbol{x}_0)$. Thus f is both upper and lower semicontinuous at $\boldsymbol{x}_0$.

Now suppose that f is both upper and lower semicontinuous at $\boldsymbol{x}_0$. For $\epsilon \in \mathbb{R}_{>0}$ let U_1 be a relative neighbourhood of $\boldsymbol{x}_0$ such that $\boldsymbol{x} \in U_1$ implies that $f(\boldsymbol{x}) < f(\boldsymbol{x}_0) + \epsilon$, this being possible by upper semicontinuity of f. Let U_2 be a relative neighbourhood of $\boldsymbol{x}_0$ such that $\boldsymbol{x} \in U_2$ implies that $f(\boldsymbol{x}) > f(\boldsymbol{x}_0) - \epsilon$, this being possible by lower semicontinuity of f. Now let $\delta \in \mathbb{R}_{>0}$ be such that $A \cap \mathsf{B}^n(\delta, \boldsymbol{x}_0) \subseteq U_1 \cap U_2$. Then $f(\boldsymbol{x}) \in \mathsf{B}^1(\epsilon, f(\boldsymbol{x}_0))$ for $\boldsymbol{x} \in A \cap \mathsf{B}^n(\delta, \boldsymbol{x}_0)$, giving continuity of f at $\boldsymbol{x}_0$. ■

Now that we understand some basic characterisations of upper and lower semicontinuity, let us give some examples that, we hope, will help in understanding these ideas.

1.10.16 Examples (Upper and lower semicontinuous functions)

1. The floor function $x \mapsto \lfloor x \rfloor$ mapping a real number x to the largest integer less than or equal to x is upper semicontinuous. Indeed, let $x_0 \in \mathbb{R}$ and let $\epsilon \in \mathbb{R}_{>0}$. If x_0 is not an integer then there exists $\delta \in \mathbb{R}_{>0}$ such that $\mathsf{B}^1(\delta, x_0)$ does not contain an integer. In this case we have $f(x) = f(x_0)$ for all $x \in \mathsf{B}^1(\delta, x_0)$ and so, in particular, $f(x) < f(x_0) + \epsilon$. If x_0 is an integer then $f(x_0) = x_0$. If $\delta = \frac{1}{2}$ and if $x \in [x_0, x_0 + \delta)$ then $f(x) = x_0 = f(x_0)$. If $x \in (x_0 - \delta, x_0)$ then $f(x) = x_0 - 1 = f(x_0) - 1$. In either case $f(x) < f(x_0) + \epsilon$. Thus f is indeed upper semicontinuous.
2. An entirely similar analysis shows that the ceiling function $x \mapsto \lceil x \rceil$ which assigns to a real number x the smallest integer not less that x is lower semicontinuous.

In Figure 1.19 we show the graphs of the floor and ceiling functions. Looking at these graphs, one might be inclined to say that upper and lower semicontinuity of piecewise continuous functions has to do with whether the functions takes the value of the left or right limit at points of discontinuity. This is not correct. At points of discontinuity of a piecewise continuous function, an upper semicontinuous function takes the larger of the left and right limit, and a lower semicontinuous function takes the smaller of the left and right limit. This assertion follows immediately from the definition.

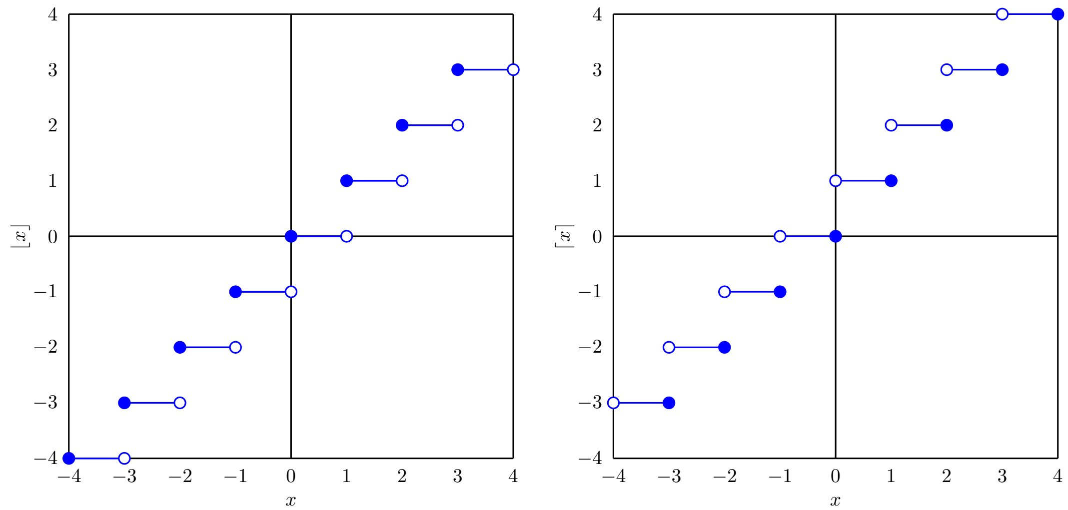

Figure 1.19 The floor function (left) and ceiling function (right)

3. If $U \subseteq \mathbb{R}^n$ is open then we claim that the characteristic function χ_U of U is lower semicontinuous. To see this first recall from Proposition 1.2.29 that $\mathbb{R}^n = \mathrm{int}(U) \cup \mathrm{bd}(U) \cup \mathrm{int}(\mathbb{R}^n \setminus U)$. If $x_0 \in \mathrm{int}(U)$ then there exists $\delta \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(\delta, x_0) \subseteq U$. Then $\chi_U(x) = 1$ for all $x \in \mathsf{B}^n(\delta, x_0)$ and so x_0 is a point of continuity of χ_U. Similarly if $x_0 \in \mathrm{int}(\mathbb{R}^n \setminus U)$ then χ_U takes the value 0 in a neighbourhood of x_0. For the remaining case where $x_0 \in \mathrm{bd}(U)$ we note that $\chi_U(x_0) = 0$ and so $\chi_U(x) > \chi_U(x_0 - \epsilon$ for every $\epsilon \in \mathbb{R}_{>0}$ and for every $x \in \mathbb{R}^n$. This gives the desired lower semicontinuity. One can easily show that, conversely, U is open only if χ_U is lower semicontinuous.
4. If $C \subseteq \mathbb{R}^n$ is closed then the characteristic function χ_C of C is upper semicontinuous. This is shown in a manner analogous to the preceding example. Moreover, C is closed only if χ_C is upper semicontinuous. •

Let us now consider a result which indicates how upper and lower semicontinuity are preserved by certain limits.

1.10.17 Proposition (Pointwise infimum (supremum) of upper (lower) semicontinuous functions) *Let* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *, let* J *be an index set, and let* $(\mathrm{f}_{\mathrm{j}})_{\mathrm{j} \in \mathrm{J}}$ *be a family of* $\overline{\mathbb{R}}$*-valued functions on* A*. Define* $\mathrm{f}_{\min}, \mathrm{f}_{\max} \colon \mathrm{A} \to \overline{\mathbb{R}}$ *by*

$$\mathrm{f}_{\min}(\mathbf{x}) = \inf\{\mathrm{f}_{\mathrm{j}}(\mathbf{x}) \mid \mathrm{j} \in \mathrm{J}\}, \quad \mathrm{f}_{\max}(\mathbf{x}) = \sup\{\mathrm{f}_{\mathrm{j}}(\mathbf{x}) \mid \mathrm{j} \in \mathrm{J}\}.$$

Then the following statements hold:

(i) *if each of the functions* f_{j}, $\mathrm{j} \in \mathrm{J}$*, is upper semicontinuous then* $\mathrm{f}_{\min}$ *is upper semicontinuous;*

(ii) *if each of the functions* f_{j}, $\mathrm{j} \in \mathrm{J}$*, is lower semicontinuous then* $\mathrm{f}_{\max}$ *is lower semicontinuous.*

Proof We give the proof in the lower semicontinuous case; the upper semicontinuous case follows similarly.

Let $\alpha \in \mathbb{R}$ and define

$$B_\alpha = \{\boldsymbol{x} \in A \mid f_{\max}(x) \leq \alpha\}, \quad B_{j,\alpha} = \{\boldsymbol{x} \in A \mid f_j(x) \leq \alpha\}.$$

We claim that

$$B_\alpha = \cap_{j \in J} B_{j,\alpha}.$$

If $\boldsymbol{x} \in B_\alpha$ then

$$f_j(\boldsymbol{x}) \leq f_{\max}(\boldsymbol{x}) \leq \alpha,$$

implying that $\boldsymbol{x} \in B_{j,\alpha}$ for each $j \in J$. Conversely, let $\boldsymbol{x} \in \cap_{j \in J} B_{j,\alpha}$ so that $f_j(\boldsymbol{x}) \leq \alpha$ for every $j \in J$. Let $\epsilon \in \mathbb{R}_{>0}$. Then there exists $j_0 \in J$ such that $f_{j_0}(\boldsymbol{x}) > f_{\max}(\boldsymbol{x}) - \epsilon$. Thus

$$f_{\max}(\boldsymbol{x}) - \epsilon < f_{j_0}(\boldsymbol{x}) \leq \alpha.$$

This gives $f_{\max}(\boldsymbol{x}) - \epsilon < \alpha$ for every $\epsilon \in \mathbb{R}_{>0}$, and so $f_{\max}(\boldsymbol{x}) \leq \alpha$ and so $\boldsymbol{x} \in B_a$.

The above arguments show that

$$f_{\max}^{-1}((\alpha, \infty]) \cup_{j \in J} f_j^{-1}((\alpha, \infty]),$$

being a union of relatively open sets by lower semicontinuity of the functions f_j, $j \in J$, is relatively open for each $\alpha \in \mathbb{R}$. In particular, for $\epsilon \in \mathbb{R}_{>0}$ the set $f_{\max}^{-1}((f_{\max}(\boldsymbol{x}_0) - \epsilon, \infty])$ is relatively open. Thus there exists a relative neighbourhood U about $\boldsymbol{x}_0$ such that $f_{\max}(\boldsymbol{x}) > f_{\max}(\boldsymbol{x}_0) - \epsilon$ for $\boldsymbol{x} \in U$. ■

It is important to note that the conclusions of the preceding result do not hold if one replaces upper or lower semicontinuity with continuity.

1.10.18 Example (Pointwise infimum of a family of continuous functions) Let us take $A = [0, 1] \subseteq \mathbb{R}$ and consider the family of functions $(f_j)_{j \in \mathbb{Z}_{>0}}$ defined 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 1.20 we depict this sequence of functions. Note that, using the notation of Proposition 1.10.17, we have

$$f_{\min}(x) = \begin{cases} 0, & x \in [0, \frac{1}{2}), \\ 1, & x \in [\frac{1}{2}, 1], \end{cases}$$

which is discontinuous, but nonetheless upper semicontinuous. •

The preceding example is one of the principle instances where the notions of upper and lower semicontinuity are important. The idea is that one has a family of functions and one wishes to minimise, not just a single function, but the family of functions. To this end the following result combines with the preceding discussion to give a valuable tool for solving optimisation problems.

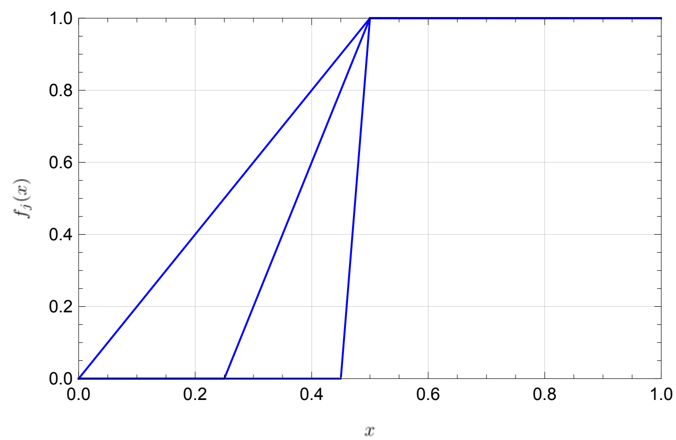

Figure 1.20 A sequence of continuous functions whose pointwise infimum is discontinuous

1.10.19 Proposition (Maxima (minima) for upper (lower) semicontinuous functions) *Let* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *be compact and let* $\mathrm{f}\colon \mathrm{A} \to \mathbb{R}$*. Then the following statements hold:*

(i) if f *is upper semicontinuous then there exists* $\mathbf{x}_0 \in \mathrm{A}$ *such that*

$$\mathrm{f}(\mathbf{x}_0) = \sup\{\mathrm{f}(\mathbf{x}) \mid \mathbf{x} \in \mathrm{A}\};$$

(ii) if f *is lower semicontinuous then there exists* $\mathbf{x}_0 \in \mathrm{A}$ *such that*

$$\mathrm{f}(\mathbf{x}_0) = \inf\{\mathrm{f}(\mathbf{x}) \mid \mathbf{x} \in \mathrm{A}\}.$$

Proof We give the proof in the lower semicontinuous case; the upper semicontinuous case follows similarly.

For $\alpha \in \mathbb{R}$ denote

$$B_\alpha = \{\boldsymbol{x} \in A \mid f(\boldsymbol{x}) \leq \alpha\},$$

noting that B_α is closed since f is lower semicontinuous. By Corollary 1.2.36 it follows that B_α is compact for each $\alpha \in \mathbb{R}$. Denote

$$\Lambda = \{\alpha \in \mathbb{R} \mid B_\alpha \neq \emptyset\}.$$

We claim that the family of relatively closed subsets $(B_\alpha)_{\alpha\in\Lambda}$ has the finite intersection property. Indeed, if $\alpha_1, \dots, \alpha_k \in \Lambda$ then

$$\cap_{j=1}^{k} B_{\alpha_j} = B_\alpha$$

where $\alpha = \min\{\alpha_1, \dots, \alpha_k\}$. Thus, by Proposition 1.2.63, $\cap_{\alpha\in\Lambda} B_\alpha \neq \emptyset$. But

$$\cap_{\alpha\in\Lambda} B_\alpha = \{\boldsymbol{x} \in A \mid f(\boldsymbol{x}) = \inf\{f(\boldsymbol{x}') \mid \boldsymbol{x}' \in A\}\}$$

and so the result follows. ■

1.10.3 Semicontinuity of rank and nullity

One of the other interesting applications of the notions of upper and lower semicontinuity is as they arise in looking at matrix-valued functions. Recall that we can defined what it mean for a matrix-valued function to be continuous using any norm on the set of matrices (i.e., linear maps) discussed in Section 1.1.3. For such functions we have the following useful result.

1.10.20 Proposition (Semicontinuity of rank and nullity) *Let* $\mathsf{S} \subseteq \mathbb{R}^n$ *and let* $\mathbf{A}\colon \mathsf{S} \to \mathrm{End}_{\mathbb{R}}(\mathbb{R}^n; \mathbb{R}^m)$ *be continuous. Then the functions* $\mathbf{x} \mapsto \mathrm{rank}(\mathbf{A}(\mathbf{x}))$ *and* $\mathbf{x} \mapsto \mathrm{nullity}(\mathbf{A}(\mathbf{x}))$ *are lower and upper semicontinuous, respectively.*

Proof Let $\alpha \in \mathbb{R}$ and let $x_0 \in \mathrm{rank}(A)^{-1}((\alpha, \infty))$. Thus $k \triangleq \mathrm{rank}(A(x_0)) > a$. There then exists $j_1, \dots, j_k \in \{1, \dots, n\}$ such that the columns $j_1, \dots, j_k$ of A are linearly independent. Therefore, there exists $i_1, \dots, i_k \in \{1, \dots, m\}$ such that the submatrix

$$\begin{bmatrix} A(i_1, j_1)(x_0) & \cdots & A(i_1, j_k)(x_0) \\ \vdots & \ddots & \vdots \\ A(i_k, j_1)(x_0) & \cdots & A(i_k, j_k)(x_0) \end{bmatrix}$$

has nonzero determinant. Since det is a continuous function on the set of $k \times k$ matrices, it follows that the function

$$x \mapsto \det \begin{bmatrix} A(i_1, j_1)(x) & \cdots & A(i_1, j_k)(x) \\ \vdots & \ddots & \vdots \\ A(i_k, j_1)(x) & \cdots & A(i_k, j_k)(x) \end{bmatrix}$$

is continuous. Therefore, there exists a relative neighbourhood U of x_0 such that the matrix

$$\begin{bmatrix} A(i_1, j_1)(x) & \cdots & A(i_1, j_k)(x) \\ \vdots & \ddots & \vdots \\ A(i_k, j_1)(x) & \cdots & A(i_k, j_k)(x) \end{bmatrix}$$

has nonzero determinant for each $x \in U$. Therefore, $\mathrm{rank}(A(x)) \geq k > a$ for all $x \in U$, and so $U \subseteq \mathrm{rank}(A)^{-1}((\alpha, \infty))$. Thus $\mathrm{rank}(A)^{-1}((\alpha, \infty))$ is relatively open, and so $\mathrm{rank}(A)$ is lower semicontinuous.

from where?

Now let $\alpha \in \mathbb{R}$ and let $x_0 \in \mathrm{nullity}(A)^{-1}((-\infty, \alpha))$. Thus $k \triangleq \mathrm{nullity}(A(x_0)) < a$. Recall () that the kernel of a matrix is equal to the orthogonal complement of the image of its transpose. This means that there exists $i_1, \dots, i_{m-k} \in \{1, \dots, m\}$ such that the rows $i_1, \dots, i_{m-k}$ form a basis for the orthogonal complement to $\ker(A)$. Therefore, there exists $j_1, \dots, j_{m-k} \in \{1, \dots, n\}$ such that the submatrix

$$\begin{bmatrix} A(i_1, j_1)(x_0) & \cdots & A(i_1, j_{m-k})(x_0) \\ \vdots & \ddots & \vdots \\ A(i_{m-k}, j_1)(x_0) & \cdots & A(i_{m-k}, j_{m-k})(x_0) \end{bmatrix}$$

has nonzero determinant. As above, there exists a relative neighbourhood U of x_0 such that this same submatrix has a nonzero determinant. This shows that

$\operatorname{rank}(A^T(x)) \geq k$ in some relative neighbourhood U of x_0. Therefore, $\operatorname{nullity}(A(x)) \leq k$ in the same relative neighbourhood U. This shows that $U \subseteq \operatorname{nullity}(A)^{-1}((-\infty, \alpha))$, and so $\operatorname{nullity}(A)^{-1}((-\infty, \alpha))$ is relatively open, giving upper semicontinuity of $\operatorname{nullity}(A)$. ■

The value of these results is that they allow one to characterise sets on which continuous matrix functions have constant rank. This constancy of rank is important, for example, in understanding the local behaviour of differentiable maps (see). To this end we make a definition. what?

1.10.21 Definition (Regular and singular points) Let $S \subseteq \mathbb{R}^n$ and let $A\colon S \to \operatorname{End}_{\mathbb{R}}(\mathbb{R}^n; \mathbb{R}^m)$ be continuous. A point $x_0 \in S$ is

(i) a ***rank regular point*** for A if there exists a relative neighbourhood U of x_0 such that $\operatorname{rank}(A(x)) = \operatorname{rank}(A(x_0))$ for each $x \in U$, is

(ii) a ***nullity regular point*** for A if there exists a relative neighbourhood U of x_0 such that $\operatorname{nullity}(A(x)) = \operatorname{nullity}(A(x_0))$ for each $x \in U$, is

(iii) a ***rank singular point*** if it is not a rank regular point, and is

(iv) a ***nullity singular point*** if it is not a nullity regular point. •

The sets of rank and nullity regular points have a topological property that is sometimes useful.

1.10.22 Proposition (Regular points are open and dense) *Let* $\mathrm{S} \subseteq \mathbb{R}^{\mathrm{n}}$ *and let* $\mathbf{A}\colon \mathrm{X} \to \operatorname{End}_{\mathbb{R}}(\mathbb{R}^{\mathrm{n}}; \mathbb{R}^{\mathrm{m}})$ *be continuous. Then the sets of rank regular points and nullity regular points are relatively open and dense.*

Proof Let us denote by R_{rank} the set of rank regular points and let $x_0 \in R_{\mathrm{rank}}$. Then, by definition of R_{rank}, there exists a relative neighbourhood U of x_0 such that $U \subseteq R_{\mathrm{rank}}$. Thus R_{rank} is open. Now let $x_0 \in S$ and let U be a relative neighbourhood of x_0. Since the function rank is bounded, there exists a least integer M such that $\operatorname{rank}(A(x)) \leq M$ for each $x \in U$. Moreover, since rank is integer-valued, there exists $x' \in U$ such that $\operatorname{rank}(A(x')) = M$. Now, by lower semicontinuity of rank, there exists a neighbourhood U' of x' such that $\operatorname{rank}(A(x)) \geq M$ for all $x \in U'$. By definition of M we also have $\operatorname{rank}(A(x)) \leq M$ for each $x \in U'$. Thus $x' \in R_{\mathrm{rank}}$, and so $x_0 \in \operatorname{cl}(R_{\mathrm{rank}})$. Therefore R_{rank} is dense.

The argument for the set of nullity regular points proceeds along identical lines except that M is chosen to be the maximal integer such that $\operatorname{nullity}(A(x)) \geq M$ for each $x \in U$, and one instead uses upper semicontinuity of nullity. ■

1.10.4 The distance between sets

In the next section we give a brief introduction to the idea of studying maps that take values in a power set. The basic definition is simple enough, but one can add some structure to this definition, and here we need some structure concerning the distance between sets.

We begin with a fairly simple measure of "closeness" of sets.

1.10.23 Definition (Distance between sets) If $A, B \subseteq \mathbb{R}^n$ are nonempty subsets of $\mathbb{R}^n$ the ***distance*** between A and B is

$$\operatorname{dist}(A, B) = \inf\{\|x - y\|_{\mathbb{R}^n} \mid x \in A,\ y \in B\}.$$

If $A = \{x\}$ for some $x \in \mathbb{R}^n$ then we denote $\operatorname{dist}(x, B) = \operatorname{dist}(\{x\}, B)$ and if $B = \{y\}$ for some $y \in \mathbb{R}^n$ then we denote $\operatorname{dist}(A, y) = \operatorname{dist}(A, \{y\})$. •

Note that $\operatorname{dist}(A, B)$ is always defined, being an infimum of a nonempty subset of $\mathbb{R}$ that is bounded below by zero. For general sets A and B there is not much useful one can say about $\operatorname{dist}(A, B)$. However, if we make some assumptions about the sets, then there is some structure here. Let us explore some of this.

1.10.24 Proposition (Continuity of distance to a set) *If* $\mathrm{B} \subseteq \mathbb{R}^{\mathrm{n}}$ *then the function* $\mathbf{x} \mapsto \operatorname{dist}(\mathbf{x}, \mathrm{B})$ *on* $\mathbb{R}^{\mathrm{n}}$ *is uniformly continuous.*

Proof Let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta = \frac{\epsilon}{2}$. Let $y \in B$ be such that $\|x_1 - y\|_{\mathbb{R}^n} - \operatorname{dist}(x_1, B) < \frac{\epsilon}{2}$. Then, if $\|x_1 - x_2\|_{\mathbb{R}^n} < \delta$,

$$\operatorname{dist}(x_2, B) \le \|x_2 - y\|_{\mathbb{R}^n} \le \|x_2 - x_1\|_{\mathbb{R}^n} + \|x_1 - y\|_{\mathbb{R}^n} \le \operatorname{dist}(x_1, B) + \epsilon.$$

In a symmetric manner one shows that

$$\operatorname{dist}(x_1, B) \le \operatorname{dist}(x_2, B) + \epsilon$$

provided that $\|x_1 - x_2\|_{\mathbb{R}^n} < \delta$. Therefore,

$$|\operatorname{dist}(x_1, B) - \operatorname{dist}(x_2, B)| < \epsilon$$

provided that $\|x_1 - x_2\|_{\mathbb{R}^n} < \delta$, giving uniform continuity, as desired. ■

Now let us consider some properties of the distance function for closed sets.

1.10.25 Proposition (Set distance and closed sets) *If* $\mathrm{A}, \mathrm{B} \subseteq \mathbb{R}^{\mathrm{n}}$ *are closed sets then the following statements hold:*

(i) if $\mathrm{A} \cap \mathrm{B} = \emptyset$ *then* $\operatorname{dist}(\mathbf{x}, \mathrm{B}), \operatorname{dist}(\mathrm{A}, \mathbf{y}) > 0$ *for all* $\mathbf{x} \in \mathrm{A}$ *and* $\mathbf{y} \in \mathrm{B}$*;*

(ii) if A *is compact then there exists* $\mathbf{x}_0 \in \mathrm{A}$ *and* $\mathbf{y}_0 \in \mathrm{B}$ *such that* $\operatorname{dist}(\mathrm{A}, \mathrm{B}) = \|\mathbf{x}_0 - \mathbf{y}_0\|_{\mathbb{R}^{\mathrm{n}}}$.

Proof (i) Suppose that $\operatorname{dist}(x, B) = 0$. Then there exists a sequence $(y_j)_{j \in \mathbb{Z}_{>0}}$ in B such that $\|y_j - x\|_{\mathbb{R}^n} < \frac{1}{j}$ for each $j \in \mathbb{Z}_{>0}$. Thus the sequence $(y_j)_{j \in \mathbb{Z}_{>0}}$ converges to x and so $x \in \operatorname{cl}(B) = B$ by Proposition 1.2.26. Therefore, if $A \cap B = \emptyset$ we can conclude that if $\operatorname{dist}(x, B) = 0$ then $x \notin A$. That is, $\operatorname{dist}(x, B) > 0$ for every $x \in A$, and similarly $\operatorname{dist}(A, y) > 0$ for every $y \in B$.

(ii) By Proposition 1.10.24 the function $x \mapsto \operatorname{dist}(x, B)$ is continuous and so too then is its restriction to the compact set A by Proposition 1.3.24. Thus, by Theorem 1.3.32 it follows that there exists $x_0 \in A$ such that $\operatorname{dist}(A, B) = \operatorname{dist}(x_0, B)$. Now there exists a sequence $(y_j)_{j \in \mathbb{Z}_{>0}}$ in B such that $\|y_j - x_0\|_{\mathbb{R}^n} < \operatorname{dist}(x_0, B) + \frac{1}{j}$ for each $j \in \mathbb{Z}_{>0}$. The sequence $(y_j)_{j \in \mathbb{Z}_{>0}}$ is bounded, being contained in the compact set $\mathsf{B}^n(1, x_0)$. Therefore, by the Bolzano–Weierstrass Theorem, there exists a convergent subsequence $(y_{j_k})_{k \in \mathbb{Z}_{>0}}$

what?

converging to $\boldsymbol{y}_0$. Since B is closed we necessarily have $\boldsymbol{y}_0 \in B$. We claim that $\operatorname{dist}(A, B) = \|\boldsymbol{x}_0 - \boldsymbol{y}_0\|_{\mathbb{R}^n}$. Indeed, continuity of the norm (see and Theorem 1.3.2 ensure that

$$\operatorname{dist}(A, B) = \operatorname{dist}(\boldsymbol{x}_0, B) = \lim_{k \to \infty} \|\boldsymbol{y}_{j_k} - \boldsymbol{x}_0\|_{\mathbb{R}^n} = \|\boldsymbol{y}_0 - \boldsymbol{x}_0\|_{\mathbb{R}^n},$$

as desired. ■

Let us illustrate this notion of distance with a couple of examples.

1.10.26 Examples (Distance between sets)

1. Take $A, B \subseteq \mathbb{R}^2$ defined by

$$A = \{\boldsymbol{x} \in \mathbb{R}^2 \mid \|\boldsymbol{x}\|_{\mathbb{R}^2} \leq 1\}, \quad B = \{\boldsymbol{x} \in \mathbb{R}^2 \mid \|\boldsymbol{x}\|_{\mathbb{R}^2} \geq 2\};$$

see Figure 1.21. We first claim that $\operatorname{dist}(A, B) = 1$. Certainly $\operatorname{dist}(A, B) \leq 1$

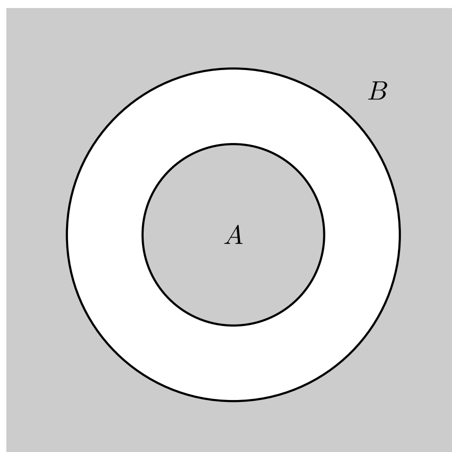

Figure 1.21 Set with many points at which distance achieves its minimum

since, for example, $(1, 0) \in A$, $(2, 0) \in B$, and $\|(2, 0) - (1, 0)\|_{\mathbb{R}^2} = 1$. To see that $\operatorname{dist}(A, B) \geq 1$ notice that if $\boldsymbol{x} \in A$, if $r \in (0, 1]$, and if $\boldsymbol{y} \in B$ then

$$\|\boldsymbol{y} - \boldsymbol{x}\|_{\mathbb{R}^2} \geq |\|\boldsymbol{y}\|_{\mathbb{R}^2} - \|\boldsymbol{x}\|_{\mathbb{R}^2}| \geq 1,$$

using Exercise 1.1.3. Note that in this case any points $\boldsymbol{x}_0 \in A$ and $\boldsymbol{y}_0 \in B$ that satisfy $\|\boldsymbol{x}_0\|_{\mathbb{R}^2} = 1$ and $\boldsymbol{y}_0 = 2\boldsymbol{x}_0$ will have the property that $\operatorname{dist}(A, B) = \|\boldsymbol{x}_0 - \boldsymbol{y}_0\|_{\mathbb{R}^2}$. There are clearly many such points, and so we see that the points at which distance is minimised, if they exist (and they do in this case since A is compact), need not be unique.

2. Next consider $A, B \subseteq \mathbb{R}^2$ given by

$$A = \left\{(x_1, x_2) \in \mathbb{R}^2 \;\middle|\; x_2 \geq \tfrac{1}{x_1}\right\}, \quad B = \left\{(y_1, y_2) \in \mathbb{R}^2 \;\middle|\; y_2 \leq -\tfrac{1}{y_1}\right\};$$

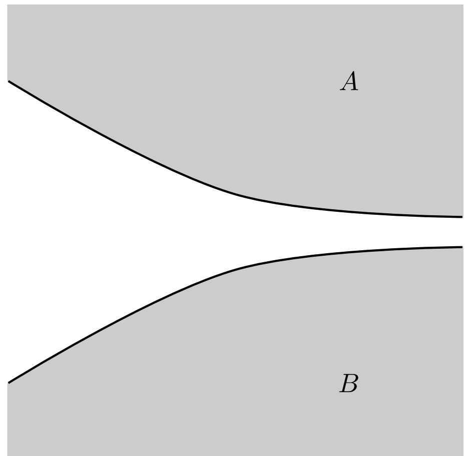

Figure 1.22 Sets where there are no points at which the distance achieves its minimum

see Figure 1.22. It is evident that $\mathrm{dist}(A, B) = 0$ since, for any $j \in \mathbb{Z}_{>0}$ we have $(j, \frac{1}{j}) \in A$, $(j, -\frac{1}{j}) \in B$, and $\|(j, \frac{1}{j}) - (j, -\frac{1}{j})\|_{\mathbb{R}^2} = \frac{2}{j}$. Thus the distance between such points can be made arbitrarily small. Moreover, note that there are no points $\boldsymbol{x}_0 \in A$ and $\boldsymbol{y}_0 \in B$ for which $\mathrm{dist}(A, B) = \|\boldsymbol{x}_0 - \boldsymbol{y}_0\|_{\mathbb{R}^2}$. This shows that compactness of one of the sets A and B is necessary in order that the minimum distance actually be achieved by points in the set. •

1.10.5 A little set-valued analysis

The preceding section had to do with the minimum distance of points in two sets. Next let us turn to a different comparison of sets, one that compares their "shapes." In order to do this the following notion will be useful.

1.10.27 Definition (r-neighbourhood of a set) For $A \subseteq \mathbb{R}^n$ nonempty and for $r \in \mathbb{R}_{>0}$, the set

$$N(r, A) = \cup_{\boldsymbol{x} \in A} \mathsf{B}^n(r, \boldsymbol{x})$$

is called the ***r-neighbourhood of* A**. •

The idea is that one "thickens" the set by r as depicted in Figure 1.23. One now defines the following numbers associated with nonempty subsets $A, B \subseteq \mathbb{R}^n$:

$$\mathsf{h}^*(A, B) = \inf\{r \in \mathbb{R}_{>0} \mid B \subseteq N(r, A)\},$$
$$\mathsf{h}_*(A, B) = \inf\{r \in \mathbb{R}_{>0} \mid A \subseteq N(r, B)\}.$$

Let us record some properties of these numbers.

1.10.28 Lemma (Properties of $\mathsf{h}^*(A, B)$ and $\mathsf{h}_*(A, B)$) *For nonempty subsets* $A, B, C \subseteq \mathbb{R}^n$ *the following statements hold:*

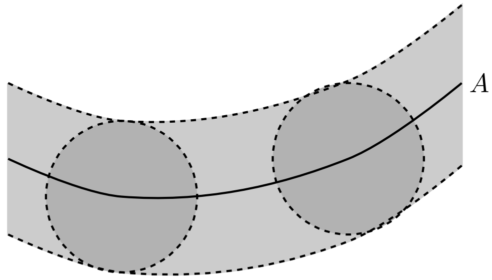

Figure 1.23 The r-neighbourhood of a set

(i) $\mathsf{h}^*(A, B) = \sup\{\mathrm{dist}(A, \mathbf{y}) \mid \mathbf{y} \in B\}$;

(ii) $\mathsf{h}_*(A, B) = \sup\{\mathrm{dist}(\mathbf{x}, B) \mid \mathbf{x} \in A\}$;

(iii) *if* A *is bounded then* $\mathsf{h}_*(A, B) < \infty$;

(iv) *if* B *is bounded then* $\mathsf{h}^*(A, B) < \infty$;

(v) $\mathsf{h}^*(A, B) = \mathsf{h}_*(B, A)$;

(vi) $\mathsf{h}^*(A, B) = 0$ *if and only if* $B \subseteq \mathrm{cl}(A)$;

(vii) $\mathsf{h}_*(A, B) = 0$ *if and only if* $A \subseteq \mathrm{cl}(B)$;

(viii) $\mathsf{h}^*(A, C) \le \mathsf{h}^*(A, B) + \mathsf{h}^*(B, C)$;

(ix) $\mathsf{h}_*(A, C) \le \mathsf{h}_*(A, B) + \mathsf{h}_*(B, C)$.

Proof (i) and (ii) We prove (ii).

Let $\epsilon \in \mathbb{R}_{>0}$. Then, for each $x \in A$ there exists $y_x \in B$ such that

$$\mathrm{dist}(x, B) \le \|x - y_x\|_{\mathbb{R}^n} < \mathsf{h}_*(A, B) + \epsilon$$

(the first inequality follows from the definition of $\mathrm{dist}(x, B)$ and the second follows from the definition of $\mathsf{h}_*(A, B)$). This holding for every $x \in A$ we have

$$\sup\{\mathrm{dist}(x, B) \mid x \in A\} \le \mathsf{h}_*(A, B) + \epsilon.$$

This holding for every $\epsilon \in \mathbb{R}_{>0}$ we have

$$\sup\{\mathrm{dist}(x, B) \mid x \in A\} \le \mathsf{h}_*(A, B).$$

Again let $\epsilon \in \mathbb{R}_{>0}$ and note that for $x \in A$ there exists $y_x \in B$ such that

$$\|x - y_x\|_{\mathbb{R}^n} < \mathrm{dist}(x, B) + \epsilon.$$

Thus

$$\sup\{\|x - y_x\|_{\mathbb{R}^n} \mid x \in A\} \le \sup\{\mathrm{dist}(x, B) \mid x \in A\} + \epsilon.$$

This means that every point in A lies within a distance $\sup\{\mathrm{dist}(x, B) \mid x \in A\} + \epsilon$ of some point in B. This is the definition of the expression

$$\mathsf{h}_*(A, B) \le \sup\{\mathrm{dist}(x, B) \mid x \in A\} + \epsilon$$

or

$$\mathsf{h}_*(A, B) \leq \sup\{\mathrm{dist}(\boldsymbol{x}, B) \mid \boldsymbol{x} \in A\}.$$

Combining the previous two paragraphs gives this part of the proof.

(iii) and (iv) Let us prove, say, (iii). Suppose that $B \subseteq \mathsf{B}^n(M_B, \mathbf{0})$ for some $M_B \in \mathbb{R}_{>0}$. Taking $r \in \mathbb{R}_{>0}$ sufficiently large ensures that $\mathsf{B}^n(M_B, \mathbf{0}) \subseteq \mathsf{B}^n(r, \boldsymbol{x}_0)$ for some $\boldsymbol{x}_0 \in A$. Thus $\mathsf{h}^*(A, B) \leq r$ and so is finite.

(v) This is obvious.

(vi) and (vii) Let us prove, say (vi).

Suppose that $\mathsf{h}^*(A, B) = 0$ and let $\boldsymbol{y} \in B$. Then, for every $r \in \mathbb{R}_{>0}$ there exists $\boldsymbol{x} \in A$ such that $\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} < r$. Thus there exists a sequence $(\boldsymbol{x}_j)_{j \in \mathbb{Z}_{>0}}$ in A such that $\|\boldsymbol{x}_j - \boldsymbol{y}\|_{\mathbb{R}^n} < \frac{1}{j}$ for every $j \in \mathbb{Z}_{>0}$. Thus the sequence converges to $\boldsymbol{y}$ and so $\boldsymbol{y} \in \mathrm{cl}(A)$.

Next suppose that $B \subseteq \mathrm{cl}(A)$ and let $\boldsymbol{y} \in B$. Then there exists a sequence $(\boldsymbol{x}_j)_{j \in \mathbb{Z}_{>0}}$ in A such that $\|\boldsymbol{x}_j - \boldsymbol{y}\|_{\mathbb{R}^n} < \frac{1}{j}$ for every $j \in \mathbb{Z}_{>0}$. Thus $\boldsymbol{y} \in N(\frac{1}{j}, A)$ for every $j \in \mathbb{Z}_{>0}$. Since this holds for every $\boldsymbol{y} \in B$ it follows that $\mathsf{h}^*(A, B) = 0$.

(viii) and (ix) Let us prove, say, (ix). Let $\boldsymbol{x} \in A$, $\boldsymbol{y} \in B$, and $\boldsymbol{z} \in C$. We then have

$$\|\boldsymbol{x} - \boldsymbol{z}\|_{\mathbb{R}^n} \leq \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} + \|\boldsymbol{y} - \boldsymbol{z}\|_{\mathbb{R}^n}.$$

Since this holds for every $\boldsymbol{z} \in C$ we then have

$$\begin{aligned} & \inf\{\|\boldsymbol{x} - \boldsymbol{z}\|_{\mathbb{R}^n} \mid \boldsymbol{z} \in C\} \leq \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} + \inf\{\|\boldsymbol{y} - \boldsymbol{z}\|_{\mathbb{R}^n} \mid \boldsymbol{z} \in C\} \\ \implies \quad & \mathrm{dist}(\boldsymbol{x}, C) \leq \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} + \mathrm{dist}(\boldsymbol{y}, C). \end{aligned}$$

Since this holds for every $\boldsymbol{y} \in B$ we have

$$\begin{aligned} & \mathrm{dist}(\boldsymbol{x}, C) \leq \inf\{\|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^n} \mid \boldsymbol{y} \in B\} + \sup\{\mathrm{dist}(\boldsymbol{y}, C) \mid \boldsymbol{y} \in B\} \\ \implies \quad & \mathrm{dist}(\boldsymbol{x}, C) \leq \mathrm{dist}(\boldsymbol{x}, B) + \mathsf{h}_*(B, C). \end{aligned}$$

Since this holds for every $\boldsymbol{x} \in A$ we have

$$\begin{aligned} & \sup\{\mathrm{dist}(\boldsymbol{x}, C) \mid \boldsymbol{x} \in A\} \leq \sup\{\mathrm{dist}(\boldsymbol{x}, B) \mid \boldsymbol{x} \in A\} + \mathsf{h}_*(B, C) \\ \implies \quad & \mathsf{h}_*(A, C) \leq \mathsf{h}_*(A, B) + \mathsf{h}_*(B, C), \end{aligned}$$

as desired. ■

Note that it is not generally the case that $\mathsf{h}^*(A, B) = \mathsf{h}_*(A, B)$ as we shall see in the examples below. To "symmetrise" this we then make the following definition.

1.10.29 Definition (Hausdorff distance) For nonempty subsets $A, B \subseteq \mathbb{R}^n$ the ***Hausdorff***[9] ***distance*** between A and B is

$$\mathsf{h}(A, B) = \max\{\mathsf{h}^*(A, B), \mathsf{h}_*(A, B)\}. \qquad \bullet$$

From Lemma 1.10.28 we have the following result.

[9]Felix Hausdorff (1868–1942) was German mathematician who made key contributions to the burgeoning fields of topology and set theory in the early 1900's.

1.10.30 Theorem (Properties of the Hausdorff distance) *If* $\mathrm{A}, \mathrm{B}, \mathrm{C} \subseteq \mathbb{R}^{\mathrm{n}}$ *are nonempty and compact then the following statements hold:*

(i) $\mathsf{h}(\mathrm{A}, \mathrm{B}) \geq 0$*;*

(ii) $\mathsf{h}(\mathrm{A}, \mathrm{B}) = 0$ *if and only if* $\mathrm{A} = \mathrm{B}$*;*

(iii) $\mathsf{h}(\mathrm{A}, \mathrm{B}) = \mathsf{h}(\mathrm{B}, \mathrm{A})$*;*

(iv) $\mathsf{h}(\mathrm{A}, \mathrm{C}) \leq \mathsf{h}(\mathrm{A}, \mathrm{B}) + \mathsf{h}(\mathrm{B}, \mathrm{C})$.

As we shall see in Section III-1.1, the properties of the Hausdorff metric from the preceding theorem makes the collection of compact subsets of $\mathbb{R}^n$ into a metric space, and h is called the ***Hausdorff metric***. We refer to Section 1.10.10 for references.

Let us give a few examples to illustrate the Hausdorff distance.

1.10.31 Examples (Hausdorff distance)

1. Let us take $A = (0, 1)$ and $B = [0, 1]$. Note that for any $r \in \mathbb{R}_{>0}$ we have $A \subseteq N(r, B)$ and $B \subseteq N(r, A)$. Thus $\mathsf{h}^*(A, B) = \mathsf{h}_*(A, B)$ and so $\mathsf{h}(A, B) = 0$. However, $A \neq B$. This explains why closedness is an essential property for sets to possess when using the Hausdorff distance.
2. Let us take $A = [0, 1]$ and $B = [0, \infty)$. Since $A \subseteq B$ we have $\mathsf{h}_*(A, B) = 0$. However, there is no finite $r \in \mathbb{R}_{>0}$ for which $B \subseteq N(r, A)$ and so $\mathsf{h}^*(A, B) = \infty$. Thus $\mathsf{h}(A, B) = \infty$. This explains why boundedness is important when using the Hausdorff distance.
3. Define $A, B \subseteq \mathbb{R}^2$ by

$$A = \{\boldsymbol{x} \in \mathbb{R}^2 \mid \|\boldsymbol{x}\|_{\mathbb{R}^n} \leq 1\}, \quad B = \{\boldsymbol{x} \in \mathbb{R}^2 \mid \|\boldsymbol{x}\|_{\mathbb{R}^n} \in [2, 4]\};$$

we show these sets in Figure 1.24. We leave it for the reader to convince themselves that

$$\mathsf{h}^*(A, B) = 3, \quad \mathsf{h}_*(A, B) = 2,$$

and so $\mathsf{h}(A, B) = 3$. •

One of the important (but by no means the only) applications of the Hausdorff distance is in characterising set-valued maps. Let us give the definition so we know what we are talking about. First let us define some notation associated to general set-valued maps.

1.10.32 Definition (Set-valued map) Let S and T be sets. A ***set-valued map*** from S to T is a map $F \colon S \to \mathbf{2}^T$. If F is a set-valued map from S to T we shall write $F \colon S \twoheadrightarrow T$. The ***graph*** of a set-valued map $F \colon S \twoheadrightarrow T$ is

$$\operatorname{graph}(F) = \{(x, y) \in S \times T \mid y \in F(x)\}$$ •

Sometimes what we call a set-valued map is called a ***correspondence***.

Now let us assign some properties to set-valued maps between Euclidean spaces.

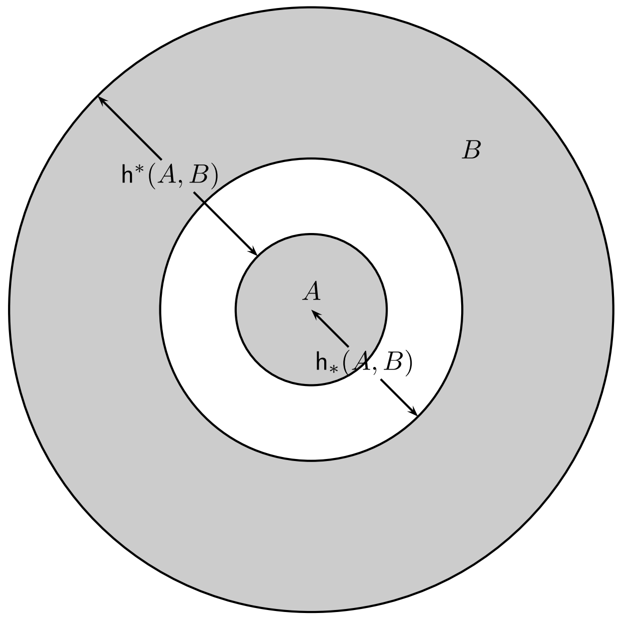

Figure 1.24 An illustration of Hausdorff distance

1.10.33 Definition (Upper hemicontinuous, lower hemicontinuous, Hausdorff continuous, bounded, locally bounded) Let $A \subseteq \mathbb{R}^n$ and let $\boldsymbol{F}\colon A \twoheadrightarrow \mathbb{R}^m$ be a set-valued map for which $\boldsymbol{F}(\boldsymbol{x}) \neq \emptyset$ for each $\boldsymbol{x} \in A$.

(i) The set-valued map $\boldsymbol{F}$ is ***upper hemicontinuous*** at $\boldsymbol{x}_0 \in A$ if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that $\mathsf{h}^*(\boldsymbol{F}(\boldsymbol{x}_0), \boldsymbol{F}(\boldsymbol{x})) < \epsilon$ for every $\boldsymbol{x} \in \mathsf{B}^n(\delta, \boldsymbol{x}_0) \cap A$.

(ii) The set-valued map $\boldsymbol{F}$ is ***lower hemicontinuous*** at $\boldsymbol{x}_0$ if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that $\mathsf{h}_*(\boldsymbol{F}(\boldsymbol{x}_0), \boldsymbol{F}(\boldsymbol{x})) < \epsilon$ for every $\boldsymbol{x} \in \mathsf{B}^n(\delta, \boldsymbol{x}_0) \cap A$.

(iii) The set-valued map $\boldsymbol{F}$ is ***Hausdorff continuous*** at $\boldsymbol{x}_0 \in A$ if it is both upper and lower hemicontinuous at $\boldsymbol{x}_0$. That is, $\boldsymbol{F}$ is Hausdorff continuous at $\boldsymbol{x}_0$ if, for every $\epsilon \in \mathbb{R}_{>0}$, there exists $\delta \in \mathbb{R}_{>0}$ such that $\mathsf{h}(\boldsymbol{F}(\boldsymbol{x}), \boldsymbol{F}(\boldsymbol{x}_0)) < \epsilon$ for every $\boldsymbol{x} \in \mathsf{B}^n(\delta, \boldsymbol{x}_0) \cap A$.

(iv) The set-valued map $\boldsymbol{F}$ is ***upper hemicontinuous*** if it is upper hemicontinuous at every point in A.

(v) The set-valued map $\boldsymbol{F}$ is ***lower hemicontinuous*** if it is lower hemicontinuous at every point in A.

(vi) The set-valued map $\boldsymbol{F}$ is ***Hausdorff continuous*** if it is Hausdorff continuous at every point in A.

(vii) The set-valued map $\boldsymbol{F}$ is ***bounded*** if there exists $M \in \mathbb{R}_{>0}$ such that $\boldsymbol{F}(\boldsymbol{x}) \subseteq \mathsf{B}^n(M, \boldsymbol{0})$ for each $\boldsymbol{x} \in A$.

(viii) The set-valued map $\boldsymbol{F}$ is ***locally bounded*** if $\boldsymbol{F}|K$ is bounded for every compact subset K of A. •

There are related notions of continuity of set-valued maps that can be made using sequences. Let us present these, and then explore the relationships between all of the notions.

1.10.34 Definition (Upper semicontinuous, lower semicontinuous, continuous) Let $A \subseteq \mathbb{R}^n$ and let $\boldsymbol{F}\colon A \twoheadrightarrow \mathbb{R}^m$ be a set-valued map for which $\boldsymbol{F}(\boldsymbol{x}) \neq \emptyset$ for each $\boldsymbol{x} \in A$.

(i) The set-valued map $\boldsymbol{F}$ is ***upper semicontinuous*** at $\boldsymbol{x}_0 \in A$ if, for every sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in A converging to $\boldsymbol{x}_0$ and for every open subset $V \subseteq \mathbb{R}^m$ for which $\boldsymbol{F}(\boldsymbol{x}_0) \subseteq V$, there exists $N \in \mathbb{Z}_{>0}$ such that $\boldsymbol{F}(\boldsymbol{x}_j) \subseteq V$ for $j \geq N$.

(ii) The set-valued map $\boldsymbol{F}$ is ***lower semicontinuous*** at $\boldsymbol{x}_0 \in A$ if, for every sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in A converging to $\boldsymbol{x}_0$ and for every open subset $V \subseteq \mathbb{R}^m$ for which $\boldsymbol{F}(\boldsymbol{x}_0) \cap V \neq \emptyset$, there exists $N \in \mathbb{Z}_{>0}$ such that $\boldsymbol{F}(\boldsymbol{x}_j) \cap V \neq \emptyset$ for $j \geq N$.

(iii) The set-valued map $\boldsymbol{F}$ is ***continuous*** at $\boldsymbol{x}_0$ if it is both upper and lower semicontinuous at $\boldsymbol{x}_0$.

(iv) The set-valued map $\boldsymbol{F}$ is ***upper semicontinuous*** if it is upper semicontinuous at every point in A.

(v) The set-valued map $\boldsymbol{F}$ is ***lower semicontinuous*** if it is lower semicontinuous at every point in A.

(vi) The set-valued map $\boldsymbol{F}$ is ***continuous*** if it is continuous at every point in A. •

The terminology we use here is very much not standardised. Different authors will use the words "hemicontinuous" and "semicontinuous" to mean different things. The reader should be alert to this.

Let us now establish the relationships between the notions of hemicontinuity and semicontinuity. The following result should be read carefully; the implications have a potentially confusing asymmetry.

1.10.35 Theorem (Relationship between hemicontinuity and semicontinuity) *Let* $A \subseteq \mathbb{R}^n$ *and let* $\mathbf{F}\colon A \twoheadrightarrow \mathbb{R}^m$ *be a set-valued map for which* $\mathbf{F}(\mathbf{x}) \neq \emptyset$ *for each* $\mathbf{x} \in A$*. Then the following statements hold:*

(i) if $\mathbf{F}$ *is upper semicontinuous at* $\mathbf{x}_0 \in A$ *then it is upper hemicontinuous at* $\mathbf{x}_0$*;*

(ii) if $\mathbf{F}$ *is lower hemicontinuous at* $\mathbf{x}_0 \in A$ *then it is lower semicontinuous at* $\mathbf{x}_0$*;*

(iii) if $\mathbf{F}$ *is upper hemicontinuous at* $\mathbf{x}_0 \in A$ *and if* $\mathbf{F}(\mathbf{x}_0)$ *is compact, then* $\mathbf{F}$ *is upper semicontinuous at* $\mathbf{x}_0$*;*

(iv) if $\mathbf{F}$ *is lower semicontinuous at* $\mathbf{x}_0 \in A$ *and if* $\mathbf{F}(\mathbf{x}_0)$ *is totally bounded, then* $\mathbf{F}$ *is lower hemicontinuous at* $\mathbf{x}_0$*.*

Proof (i) Suppose that $\boldsymbol{F}$ is not upper hemicontinuous at $\boldsymbol{x}_0$. Then there exists $\epsilon \in \mathbb{R}_{>0}$ such that, for every $\delta \in \mathbb{R}_{>0}$, $\mathsf{h}^*(\boldsymbol{F}(\boldsymbol{x}_0), \boldsymbol{F}(\boldsymbol{x})) \geq \epsilon$ for some $\boldsymbol{x} \in \mathsf{B}^n(\delta, \boldsymbol{x}_0) \cap A$. Therefore, for each $j \in \mathbb{Z}_{>0}$ there exists $\boldsymbol{x}_j \in A$ such that $\|\boldsymbol{x}_j - \boldsymbol{x}_0\|_{\mathbb{R}^n} < \frac{1}{j}$ and such that $\mathsf{h}^*(\boldsymbol{F}(\boldsymbol{x}_0), \boldsymbol{F}(\boldsymbol{x}_j)) \geq \epsilon$. This means that, for the sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ so defined and for $V = N(\epsilon, \boldsymbol{F}(\boldsymbol{x}_0))$, we have $\boldsymbol{F}(\boldsymbol{x}_j) \not\subseteq V$. Thus $\boldsymbol{F}$ is not upper semicontinuous at $\boldsymbol{x}_0$.

(ii) Let $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence converging to $\boldsymbol{x}_0$, let $\boldsymbol{y}_0 \in \boldsymbol{F}(\boldsymbol{x}_0)$, and let V be an open subset of $\mathbb{R}^m$ for which $\boldsymbol{y}_0 \in V$. Let $\epsilon \in \mathbb{R}_{>0}$ be such that $\mathsf{B}^m(\epsilon, \boldsymbol{y}_0) \subseteq V$. Since $\boldsymbol{F}$ is lower hemicontinuous at x_0 let $\delta \in \mathbb{R}_{>0}$ be such that $\mathsf{h}_*(F(x_0), F(x)) < \epsilon$ for $x \in \mathsf{B}^n(\delta, x_0) \cap A$. Then there exists $N \in \mathbb{Z}_{>0}$ such that $x_j \in \mathsf{B}^n(\delta, x_0) \cap A$ for all $j \geq N$. It then follows that $F(x_0) \subseteq N(\epsilon, F(x_j))$ for $j \geq N$. In particular, for each $j \geq N$ there exists $y_j \in F(x_j)$ such that $y_j \in \mathsf{B}^m(\epsilon, y_j) \subseteq V$. Thus F is lower semicontinuous at x_0.

(iii) Let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence converging to x_0 and let $V \subseteq \mathbb{R}^m$ be an open set such that $F(x_0) \subseteq V$. By part (iii) of Lemma 1.10.28 we have $\mathsf{h}^*(F(x_0), \mathbb{R}^m \setminus V) < \infty$ and by part (vi) of Lemma 1.10.28 we have $\mathsf{h}^*(F(x_0), \mathbb{R}^m \setminus V) > 0$. Let us take $\epsilon = \mathsf{h}^*(F(x_0), \mathbb{R}^m \setminus V)$. Then, by upper hemicontinuity of F at x_0 there exists $\delta \in \mathbb{R}_{>0}$ such that $\mathsf{h}^*(F(x_0), F(x)) < \epsilon$ for every $x \in \mathsf{B}^n(\delta, x_0) \cap A$. Therefore, choosing N sufficiently large that $x_j \in \mathsf{B}^n(\delta, x_0) \cap A$ for $j \geq N$, we have $F(x_j) \subseteq N(\epsilon, F(x_0))$ for $j \geq N$. By the definition of ϵ this means that $F(x_j) \cap (\mathbb{R}^m \setminus V) = \emptyset$ and so $F(x_j) \subseteq V$ for $j \geq N$. Thus F is upper semicontinuous at x_0.

(iv) Suppose that F is not lower hemicontinuous at x_0. This means that there exists $\epsilon \in \mathbb{R}_{>0}$ such that, for every $\delta \in \mathbb{R}_{>0}$, $\mathsf{h}_*(F(x_0), F(x)) \geq \epsilon$ for some $x \in \mathsf{B}^n(\delta, x_0) \cap A$. Thus there exists a sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ in A such that, for every $j \in \mathbb{Z}_{>0}$, $\|x_j - x_0\|_{\mathbb{R}^n} < \frac{1}{j}$ and such that $F(x_0) \not\subseteq N(\epsilon, F(x_j))$. Since $F(x_0)$ is totally bounded there exists $y_1, \dots, y_k \in F(x_0)$ such that

$$F(x_0) \subseteq \cup_{l=1}^k \mathsf{B}^m(\tfrac{\epsilon}{2}, y_l). \tag{1.57}$$

We claim that there exists $l \in \{1, \dots, k\}$ such that, for every $N \in \mathbb{Z}_{>0}$, there exists $j \geq N$ such that $F(x_j) \cap \mathsf{B}^m(\frac{\epsilon}{2}, y_l) = \emptyset$. Suppose otherwise. That is, suppose that, for each $l \in \{1, \dots, k\}$, there exists $N \in \mathbb{Z}_{>0}$ such that $F(x_j) \cap \mathsf{B}^m(\frac{\epsilon}{2}, y_l) \neq \emptyset$. Let $y \in F(x_0)$. By (1.57) there exists $l(y) \in \{1, \dots, k\}$ such that $y \in \mathsf{B}^m(\frac{\epsilon}{2}, y_{l(y)})$. For $l \in \{1, \dots, k\}$ let $N_l \in \mathbb{Z}_{>0}$ be such that $F(x_j) \cap \mathsf{B}^m(\frac{\epsilon}{2}, y_l)$ for $j \geq N_l$. Let $N = \max\{N_1, \dots, N_k\}$. Then, for $j \geq N$ and for $y \in F(x_0)$ there exists $y(j, y) \in F(x_j) \cap \mathsf{B}^m(\frac{\epsilon}{2}, y_{l(y)})$. We then have

$$\|y - y(j, y)\|_{\mathbb{R}^m} \leq \|y - y_{l(y)}\|_{\mathbb{R}^m} + \|y_{l(y)} - y(j, y)\|_{\mathbb{R}^m} < \epsilon.$$

But this contradicts the fact that $F(x_0) \not\subseteq N(\epsilon, F(x_j))$ for every $j \in \mathbb{Z}_{>0}$. Thus we can conclude that there exists $l \in \{1, \dots, k\}$ such that, for every $N \in \mathbb{Z}_{>0}$, there exists $j \geq N$ such that $F(x_j) \cap \mathsf{B}^m(\frac{\epsilon}{2}, y_l) = \emptyset$. This means that F is not lower semicontinuous at x_0. ■

Now having at hand the definitions and the basic characterisations of hemicontinuity and semicontinuity, let us look at some examples and try to get a handle on what these notions mean.

1.10.36 Examples (Hemicontinuity and semicontinuity)

1. Let $A \subseteq \mathbb{R}^n$. A plain old map $f: A \to \mathbb{R}^m$ defines a set-valued map $F_f: A \twoheadrightarrow \mathbb{R}^m$ according to $F_f(x) = \{f(x)\}$.

   It is easy to show that F_f is upper semicontinuous at x_0 if and only if f is continuous at x_0 and that F_f is lower semicontinuous at x_0 if and only if f is continuous at x_0; see Exercise 1.10.3.

   Since $F_f(x)$ is compact for each $x \in A$ it follows from Theorem 1.10.35 that F_f is upper hemicontinuous if and only if it is upper semicontinuous if and only if it is lower hemicontinuous if and only if it is lower semicontinuous if and only if it is Hausdorff continuous if and only if it is continuous if and only if

f is continuous. That is to say, all possible notions of continuity (that we have defined) agree in this case. This is reassuring. Somewhat less reassuring is the fact that upper (resp. lower) semicontinuous functions do not correspond to upper (resp. lower) semicontinuous set-valued maps. This is one of the commonly accepted and confusing bits of this business. One just gets used to it.

2. Define $F\colon \mathbb{R} \twoheadrightarrow \mathbb{R}$ by $F(x) = (x-1, x+1)$. We claim that F is upper hemicontinuous but not upper semicontinuous.
Indeed, it is easy to verify that if $|x_1 - x_2| < \epsilon$ then
$$\mathsf{h}^*(F(x_1), F(x_2)) < \epsilon, \quad \mathsf{h}_*(F(x_1), F(x_2)) < \epsilon.$$
Thus F is actually Hausdorff continuous and so, in particular, upper hemicontinuous.
We next claim that F is not upper semicontinuous at 0. To see this, consider the open subset $V = (-1, 1)$ which has the property that $V \subseteq F(0)$. Consider the sequence $(x_j = \frac{1}{j})_{j\in\mathbb{Z}_{>0}}$ which converges to 0. Since $F(x_j) \not\subseteq V$ for every $j \in \mathbb{Z}_{>0}$ it follows that F is indeed not upper semicontinuous at 0.
Note that $F(0)$ is not compact.

3. Let us take $A = \{\frac{1}{k} \mid k \in \mathbb{Z}_{>0}\} \cup \{0\}$ and define $F\colon A \twoheadrightarrow \mathbb{R}$ by
$$F(x) = \begin{cases} \{0, 1, \dots, k\}, & x = \frac{1}{k},\ k \in \mathbb{Z}_{>0}, \\ \mathbb{Z}_{\ge 0}, & x = 0. \end{cases}$$
We claim that F is lower semicontinuous but not lower hemicontinuous.
To see that F is lower semicontinuous at $x_0 = \frac{1}{k}$ for $k \in \mathbb{Z}_{>0}$, note that any sequence $(x_j)_{j\in\mathbb{Z}_{>0}}$ in A converging to x_0 must be eventually constant; that is, there exists $N \in \mathbb{Z}_{>0}$ such that $x_j = \frac{1}{k}$ for all $j \ge N$. In this case we clearly have $F(x_j) \cap V = F(x_0) \cap V$ for any $j \ge N$ and for any open set $V \subseteq \mathbb{R}$. From this we easily deduce that F is lower semicontinuous at $x_0 = \frac{1}{k}$.
Next we show that F is lower semicontinuous at $x_0 = 0$. Indeed, let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in A converging to 0. If $V \subseteq \mathbb{R}$ is an open set for which $F(0) \cap V \neq \emptyset$ then this means exactly that $m \in V$ for some $m \in \mathbb{Z}_{\ge 0}$. In particular this means that $F(\frac{1}{k}) \cap V \neq \emptyset$ for all $k \ge m$. This means $F(x_j) \cap V \neq \emptyset$ for all j sufficiently large. This gives lower semicontinuity of F.
Finally, we claim that F is not lower hemicontinuous at 0. To see this one need only note that $\mathsf{h}_*(F(0), F(\frac{1}{k})) = \infty$ for all $k \in \mathbb{Z}_{>0}$. This precludes lower hemicontinuity of F at 0.
Note that $F(0)$ is not totally bounded.

The preceding two examples show that the converses of the first two parts of Theorem 1.10.35 are generally false.

4. Let $A \subseteq \mathbb{R}^n$ and let $f_-, f_+\colon A \to \mathbb{R}$ have the following properties:

(a) f_- is lower semicontinuous;

(b) f_+ is upper semicontinuous;

(c) $f_-(\boldsymbol{x}) \leq f_+(\boldsymbol{x})$ for all $\boldsymbol{x} \in A$.

Let us define $F\colon A \twoheadrightarrow \mathbb{R}$ by $F(\boldsymbol{x}) = [f_-(\boldsymbol{x}), f_+(\boldsymbol{x})]$. It is then an easy exercise to show that F is upper semicontinuous; see Exercise 1.10.4. In the case where $A = \mathbb{R}$ we depict the situation in Figure 1.25.

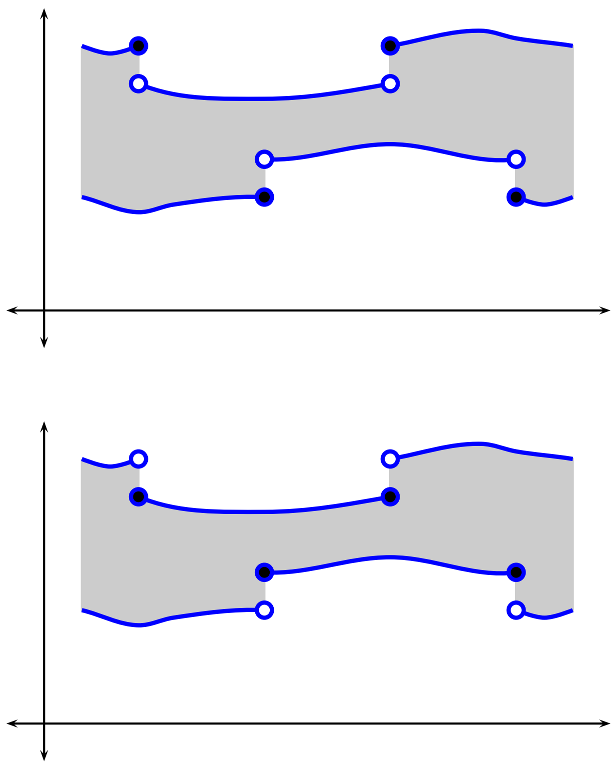

Figure 1.25 The graph of an upper continuous set-valued function (top) and a lower continuous set-valued function (bottom)

5. Now let $A \subseteq \mathbb{R}^n$ and let $f_-, f_+\colon A \to \mathbb{R}$ have the following properties:

(a) f_- is upper semicontinuous;

(b) f_+ is lower semicontinuous;

(c) $f_-(\boldsymbol{x}) \leq f_+(\boldsymbol{x})$ for all $\boldsymbol{x} \in A$.

Let us define $F\colon A \twoheadrightarrow \mathbb{R}$ by $F(\boldsymbol{x}) = [f_-(\boldsymbol{x}), f_+(\boldsymbol{x})]$. It is still an easy exercise to show that F is lower semicontinuous; see Exercise 1.10.4. We refer to Figure 1.25 for a depiction of this situation. •

Let us now consider some further properties one can assign to set-valued maps, and the relationships between these properties. First let us consider the important special case when the assigned sets are closed.

1.10.37 Definition (Closed set-valued maps) Let $A \subseteq \mathbb{R}^n$. A set-valued map $\boldsymbol{F}\colon A \twoheadrightarrow \mathbb{R}^m$ is

(i) ***closed*** at $\boldsymbol{x}_0 \in A$ if, for every sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in A converging to $\boldsymbol{x}_0$ and for every sequence $(\boldsymbol{y}_j)_{j\in\mathbb{Z}_{>0}}$ for which $\boldsymbol{y}_j \in \boldsymbol{F}(\boldsymbol{x}_j)$ and which converges to some $\boldsymbol{y}_0 \in \mathbb{R}^m$, it holds that $\boldsymbol{y}_0 \in \boldsymbol{F}(\boldsymbol{x}_0)$, and is

(ii) ***closed*** if it is closed at each point in A. •

Let us explore the relationship between closedness and our concepts of continuity.

1.10.38 Proposition (Closedness and upper semicontinuity) *If* $A \subseteq \mathbb{R}^n$, *if* $B \subseteq \mathbb{R}^m$, *and if* $\mathbf{F}\colon A \twoheadrightarrow B$ *is a set-valued map then the following statements hold:*

(i) $\mathbf{F}$ *is closed if and only if* $\mathrm{graph}(\mathbf{F})$ *is a relatively closed subset of* $A \times B$*;*

(ii) if $\mathbf{F}$ *is upper semicontinuous and if* $\mathbf{F}(\mathbf{x})$ *is relatively closed for every* $\mathbf{x} \in A$*, then* $\mathbf{F}$ *is closed;*

(iii) if B *is compact and if* $\mathbf{F}$ *is closed, then* $\mathbf{F}$ *is upper semicontinuous.*

Proof (i) Suppose that $\mathrm{graph}(\boldsymbol{F})$ is closed, let $\boldsymbol{x}_0 \in A$, let $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence converging to $\boldsymbol{x}_0$, and let $(\boldsymbol{y}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in B such that $\boldsymbol{y}_j \in \boldsymbol{F}(\boldsymbol{x}_j)$ and such that $\lim_{j\to\infty} \boldsymbol{y}_j = \boldsymbol{y}_0$. Thus $((\boldsymbol{x}_j, \boldsymbol{y}_j))_{j\in\mathbb{Z}_{>0}}$ is a sequence in $\mathrm{graph}(\boldsymbol{F})$ converging to $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in A \times B$. Therefore, closedness of $\mathrm{graph}(\boldsymbol{F})$ gives $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in \mathrm{graph}(\boldsymbol{F})$, or $\boldsymbol{y}_0 \in \boldsymbol{F}(\boldsymbol{x}_0)$, as desired.

Now suppose that $\boldsymbol{F}$ is closed at each $\boldsymbol{x} \in A$ and let $((\boldsymbol{x}_j, \boldsymbol{y}_j))_{j\in\mathbb{Z}_{>0}}$ be a convergent sequence in $\mathrm{graph}(\boldsymbol{F})$. Thus $\lim_{j\to\infty} \boldsymbol{x}_j = \boldsymbol{x}_0$ and $\lim_{j\to\infty} \boldsymbol{y}_j = \boldsymbol{y}_0$ and $\boldsymbol{y}_0 \in \boldsymbol{F}(\boldsymbol{x}_0)$. Since $\boldsymbol{y}_j \in \boldsymbol{F}(\boldsymbol{x}_j)$ it follows that $\boldsymbol{F}$ is closed at $\boldsymbol{x}_0$, and this argument applies to every $\boldsymbol{x}_0 \in A$.

(ii) We will show that, under the given hypotheses, the complement of $\mathrm{graph}(\boldsymbol{F})$ in $A \times B$ is relatively open. Let $(\boldsymbol{x}_0, \boldsymbol{y}_0) \in (A \times B) \setminus \mathrm{graph}(\boldsymbol{F})$. Since $\boldsymbol{F}(\boldsymbol{x}_0)$ is relatively closed, let V be a relative neighbourhood of $\boldsymbol{y}_0$ for which $\mathrm{cl}(V) \cap \boldsymbol{F}(\boldsymbol{x}_0) = \varnothing$. Let us define a relative neighbourhood of $\boldsymbol{F}(\boldsymbol{x}_0)$ by $V' = B \setminus \mathrm{cl}(V)$. By upper semicontinuity of $\boldsymbol{F}$ there exists a relative neighbourhood U of $\boldsymbol{x}_0$ such that $\boldsymbol{x} \in U$ implies that $\boldsymbol{F}(\boldsymbol{x}) \subseteq V'$. This means that $U \times V \cap \mathrm{graph}(\boldsymbol{F}) = \varnothing$ and so $U \times V$ is a relative neighbourhood of $(\boldsymbol{x}_0, \boldsymbol{y}_0)$ in $(A \times B) \setminus \mathrm{graph}(\boldsymbol{F})$. Thus $(A \times B) \setminus \mathrm{graph}(\boldsymbol{F})$ is open, as desired.

(iii) Suppose that B is compact and that $\boldsymbol{F}$ is not upper semicontinuous at some $\boldsymbol{x}_0 \in A$. Since $\boldsymbol{F}$ is not upper hemicontinuous at $\boldsymbol{x}_0$ (by Theorem 1.10.35) there exists a relative neighbourhood V of $\boldsymbol{F}(\boldsymbol{x}_0)$ such that for every relative neighbourhood U of $\boldsymbol{x}_0$ such that $\boldsymbol{F}(\boldsymbol{x}) \not\subseteq V$ for some $\boldsymbol{x} \in U$. Thus there is a sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in A converging to $\boldsymbol{x}_0$ and a sequence $(\boldsymbol{y}_j)_{j\in\mathbb{Z}_{>0}}$ in B for which $\boldsymbol{y}_j \in \boldsymbol{F}(\boldsymbol{x}_j)$ and for which $\boldsymbol{y}_j \notin V$ for each $j \in \mathbb{Z}_{>0}$. Compactness of B ensures that there is a convergent subsequence $(\boldsymbol{y}_{j_k})_{k\in\mathbb{Z}_{>0}}$ and the limit of this sequence cannot be in V by openness of V. But this means that $\boldsymbol{F}$ is not closed at $\boldsymbol{x}_0$. ■

Let us give examples to show that there is, in general, no correspondence between upper semicontinuity and closedness.

1.10.39 Examples (Closedness and upper semicontinuity)

1. The set-valued map $F\colon \mathbb{R} \twoheadrightarrow \mathbb{R}$ given by

$$F(x) = \begin{cases} \{\frac{1}{x}\}, & x \neq 0, \\ \{0\}, & x = 0 \end{cases}$$

is closed but not upper semicontinuous.

2. The set-valued map $F\colon \mathbb{R} \twoheadrightarrow \mathbb{R}$ given by $F(x) = (0, 1)$ is not closed but is upper semicontinuous. •

Next we consider some additional characterisations of set-valued maps when the sets assigned are compact. That this is of particular interest is already clear from Theorem 1.10.35.

1.10.40 Proposition (Compactness and upper semicontinuity) *Let* $A \subseteq \mathbb{R}^n$ *and let* $\mathbf{F}\colon A \twoheadrightarrow \mathbb{R}^m$ *be a set-valued map for which* $\mathbf{F}(\mathbf{x}_0)$ *is compact for some* $\mathbf{x}_0 \in A$*. Then the following statement are equivalent:*

(i) $\mathbf{F}$ *is upper semicontinuous at* $\mathbf{x}_0$*;*

(ii) for every sequence $(\mathbf{x}_j)_{j\in\mathbb{Z}_{>0}}$ *in* A *converging to* $\mathbf{x}_0$ *and for every sequence* $(\mathbf{y}_j)_{j\in\mathbb{Z}_{>0}}$ *in* $\mathbb{R}^m$ *for which* $\mathbf{y}_j \in \mathbf{F}(\mathbf{x}_j)$ *there exists a subsequence* $(\mathbf{y}_{j_k})_{k\in\mathbb{Z}_{>0}}$ *converging to a point in* $\mathbf{F}(\mathbf{x}_0)$*.*

Proof First suppose that $\boldsymbol{F}$ is upper semicontinuous, let $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in A converging to $\boldsymbol{x}_0 \in A$, and let $(\boldsymbol{y}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence for which $\boldsymbol{y}_j \in \boldsymbol{F}(\boldsymbol{x}_j)$. Let $\epsilon \in \mathbb{R}_{>0}$. By compactness of $\boldsymbol{F}(\boldsymbol{x}_0)$ the neighbourhood $N(\epsilon, \boldsymbol{F}(\boldsymbol{x}_0))$ of $\boldsymbol{F}(\boldsymbol{x}_0)$ is bounded. By upper semicontinuity of $\boldsymbol{F}$ at $\boldsymbol{x}_0$ there exists $N_\epsilon \in \mathbb{Z}_{>0}$ such that $\boldsymbol{y}_j \in N(\epsilon, \boldsymbol{F}(\boldsymbol{x}_0))$ for $j \geq N_\epsilon$. Now, for $k \in \mathbb{Z}_{>0}$ choose $j_k \geq N_{\frac{1}{k}}$ and suppose, moreover, that $j_{k+1} > j_k$ for each $k \in \mathbb{Z}_{>0}$. Note that

$$\lim_{k\to\infty} \operatorname{dist}(\boldsymbol{y}_{j_k}, \boldsymbol{F}(\boldsymbol{x}_0)) = 0. \tag{1.58}$$

Since the sequence $(\boldsymbol{y}_{j_k})_{k\in\mathbb{Z}_{>0}}$ is bounded (by virtue of boundedness of $\boldsymbol{F}(\boldsymbol{x}_0)$) it possess a convergent subsequence. By virtue of (1.58) it follows that the limit of this sequence is in $\boldsymbol{F}(\boldsymbol{x}_0)$.

Now suppose $\boldsymbol{F}$ is not upper semicontinuous at $\boldsymbol{x}_0$. Then there exists a neighbourhood V of $\boldsymbol{F}(\boldsymbol{x}_0)$ and a sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ such that $\boldsymbol{F}(\boldsymbol{x}_j) \not\subset V$ for each $j \in \mathbb{Z}_{>0}$. Let $(\boldsymbol{y}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence for which $\boldsymbol{y}_j \in \boldsymbol{F}(\boldsymbol{x}_j)$ and $\boldsymbol{y}_j \notin V$. Since $\mathbb{R}^m \setminus V$ is closed, and convergent subsequences of $(\boldsymbol{y}_j)_{j\in\mathbb{Z}_{>0}}$ must converge to a point in $\mathbb{R}^m \setminus V$. In particular, such subsequences cannot converge to a point in $\boldsymbol{F}(\boldsymbol{x}_0)$. ■

Having tried to develop some intuition for our notions of hemicontinuity and semicontinuity via examples, and having outlined some of the basic properties of hemicontinuity and semicontinuity, it is perhaps still far from clear why anyone would care about these ideas. There are, in fact, many reasons why these properties of set-valued maps are useful, and we discuss some of this in Section 1.10.10. However, let us here give a fairly simple theorem which illustrates where this might come up in practice.

1.10.41 Theorem (Berge's[10] Maximum Theorem) *Let* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$*, let* $\mathrm{P} \subseteq \mathbb{R}^{\mathrm{p}}$*, and let* $\mathbf{F}\colon \mathrm{A} \twoheadrightarrow \mathrm{P}$ *be lower semicontinuous. For* $\phi\colon \mathrm{A} \times \mathrm{P} \to \mathbb{R}$ *define* $\phi_{\min}, \phi_{\max}\colon \mathrm{A} \to \overline{\mathbb{R}}$ *by*

$$\phi_{\min}(\mathbf{x}) = \inf\{\phi(\mathbf{x}, \mathbf{y}) \mid \mathbf{y} \in \mathbf{F}(\mathbf{x})\},$$
$$\phi_{\max}(\mathbf{x}) = \sup\{\phi(\mathbf{x}, \mathbf{y}) \mid \mathbf{y} \in \mathbf{F}(\mathbf{x})\}.$$

Then the following statements hold:

(i) *if* ϕ *is upper semicontinuous then so is* $\phi_{\min}$*;*

(ii) *if* ϕ *is lower semicontinuous then so is* $\phi_{\max}$*.*

Moreover, if $\mathbf{F}(\mathbf{x})$ *is compact for every* $\mathbf{x} \in \mathrm{A}$ *and if we define* $\mathbf{F}_{\max}\colon \mathrm{A} \twoheadrightarrow \mathrm{P}$ *by*

$$\mathbf{F}_{\max}(\mathbf{x}) = \{\mathbf{y} \in \mathrm{P} \mid \phi(\mathbf{x}, \mathbf{y}) = \phi_{\max}(\mathbf{x})\},$$

then the following statement holds:

(iii) *if* ϕ *is continuous at* $\mathbf{x}_0$ *and if* $\mathbf{F}$ *is continuous at* $\mathbf{x}_0$*, then* $\mathbf{F}_{\max}$ *is upper semicontinuous at* $\mathbf{x}_0$ *and* $\mathbf{F}_{\max}(\mathbf{x}_0)$ *is compact.*

Proof We shall prove the first assertion; the second follows along similar lines.

Let $\alpha \in \mathbb{R}$ and let $x_0 \in \phi_{\min}^{-1}((-\infty, \alpha))$. Thus $\phi_{\min}(x_0) < \alpha$. There then exists $y_0 \in F(x_0)$ such that $\phi(x_0, y_0) < \alpha$. By Proposition 1.10.13 there exists relatively open sets $U \subseteq A$ and $V \subseteq P$ such that $(x_0, y_0) \in U \times V$ and such that $\phi(x, y) < \alpha$ for every $(x, y) \in U \times V$. Note that $y_0 \in F(x_0) \cap V$. By lower semicontinuity of F there exists a relative neighbourhood U' of x_0 in A such that $F(x) \cap V \neq \emptyset$ for $x \in U'$. We claim that $U \cap U' \subseteq \phi_{\min}^{-1}((-\infty, \alpha))$. Indeed, let $x \in U \cap U'$. Then $F(x) \cap V \neq \emptyset$. Let $y \in F(x) \cap V$. Then $(x, y) \in U \times V$ and so $\phi(x, y) < \alpha$. Since $\phi_{\min}(x) \leq \phi(x, y) < \alpha$ we can indeed conclude that $U \cap U' \subseteq \phi_{\min}^{-1}((-\infty, \alpha))$. This gives openness of $\phi_{\min}^{-1}((-\infty, \alpha))$ and so upper semicontinuity of $\phi_{\min}$ by Proposition 1.10.13.

For the final assertion of the theorem, first note that $F_{\max}(x)$ is not empty for each $x \in A$ by Proposition 1.10.19.

We claim that $F_{\max}$ is closed at x_0. Let $(x_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in A converging to x_0 and let $(y_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in P such that $y_j \in F_{\max}(x_j)$ and such that $\lim_{j\to\infty} y_j = y_0 \in P$. By Proposition 1.10.38 it follows that $y_0 \in F(x_0)$. Suppose that $y_0 \notin F_{\max}(x_0)$. Then there exists $z_0 \in F(x_0)$ such that $\phi(x_0, z_0) > \phi(x_0, y_0)$. By Exercise 1.10.2 let $(z_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence in P for which $z_j \in F(x_j)$ and for which $\lim_{j\to\infty} z_j = z_0$. Continuity of ϕ at x_0 then implies that, for sufficiently large j, $\phi(x_j, z_j) > \phi(x_j, y_j)$. But this means that for these same sufficiently large j we must have $y_j \notin F_{\max}(x_j)$. Therefore, we have $y_0 \in F_{\max}(x_0)$, so showing that $F_{\max}$ is closed at x_0.

Next we claim that the closedness of $F_{\max}$ at x_0 implies that $F_{\max}$ is upper semicontinuous at x_0. By Proposition 1.10.38 it suffices to show that there exists a compact set $K \subseteq \mathbb{R}^m$ such that $F_{\max}(x) \subseteq K$ for x in some neighbourhood of x_0. Since $F_{\max}(x) \subseteq F_(x)$ it suffices to show that there exists a compact subset K such that $F(x) \subseteq K$ for x in some neighbourhood of x_0. Let us take $K = \mathrm{cl}(N(1, F(x_0)))$ which is compact since $F(x_0)$ is compact (can you explain this?). Then, by continuity of F at x_0, there exists $\delta \in \mathbb{R}_{>0}$ such that $F(x) \subseteq K$ for $x \in \mathsf{B}^n(\delta, x_0) \cap A$, as desired. ∎

[10] The Berge here is Claude Berge (1926–2002), a French mathematician who made contributions to combinatorics and graph theory, and also to game theory.

The way to understand the Maximum Theorem is this. Points in the set A are the independent variables for the problem, points in the set P are parameters which, presumably, one has control over, $\boldsymbol{F}(\boldsymbol{x})$ gives the set of admissible parameters at $\boldsymbol{x}$, and ϕ is a cost function. One wishes to choose, for each $\boldsymbol{x} \in A$, a parameter $\boldsymbol{y} \in P$ that minimises (or maximises) ϕ. One would then like for the resulting function obtained by doing this for each $\boldsymbol{x}$ to have some reasonable properties, and one of the results of the Maximum Theorem is just what those properties are.

1.10.6 Extending continuous maps

Although it is difficult to imagine why for a newcomer to analysis, the question of, "When can I extend the definition of something from a subset to the whole set?" is an extremely important one. It comes up in all manner of existence proofs where it is easy to define "something" on a subset, but the extension to the entire set is not so easily done, at least not directly. In this section we investigate this question of extension in a simple setting. First we talk about extending functions, and then at the end of the section, we discuss briefly the rather more exotic topic of extending maps with more general codomains.

The basic building block for extending functions is the following result.

1.10.42 Theorem (Urysohn's[11] Lemma) *Let* $\mathrm{S} \subseteq \mathbb{R}^{\mathrm{n}}$ *and let* $\mathrm{a}, \mathrm{b} \in \mathbb{R}$ *with* $\mathrm{a} < \mathrm{b}$*. If* $\mathrm{A}, \mathrm{B} \subseteq \mathrm{S}$ *are disjoint relatively closed sets then there exists a continuous function* $\mathrm{f}\colon \mathrm{S} \to [\mathrm{a}, \mathrm{b}]$ *such that* $\mathrm{f}(\mathbf{x}) = \mathrm{a}$ *for all* $\mathbf{x} \in \mathrm{A}$ *and* $\mathrm{f}(\mathbf{x}) = \mathrm{b}$ *for all* $\mathbf{x} \in \mathrm{B}$*.*

Proof Since A and B are relatively closed and disjoint it follows from Proposition 1.10.25 (more precisely, from a slight modification of this result since A and B are here only relatively closed subsets of S) that $\operatorname{dist}(\boldsymbol{x}, B), \operatorname{dist}(A, \boldsymbol{y}) > 0$ for all $\boldsymbol{x} \in A$ and for all $\boldsymbol{y} \in B$. Therefore, the function $g\colon S \to [-1, 1]$ defined by

$$g(\boldsymbol{x}) = \frac{\operatorname{dist}(\boldsymbol{x}, A) - \operatorname{dist}(\boldsymbol{x}, B)}{\operatorname{dist}(\boldsymbol{x}, A) + \operatorname{dist}(\boldsymbol{x}, B)}$$

is continuous by Proposition 1.10.24, along with Propositions I-3.1.15 and 1.3.22. One directly checks that $g(\boldsymbol{x}) = -1$ for $\boldsymbol{x} \in A$ and that $g(\boldsymbol{x}) = 1$ for $\boldsymbol{x} \in B$. Taking

$$f(\boldsymbol{x}) = \tfrac{1}{2}(b - a)g(\boldsymbol{x}) + \tfrac{1}{2}(b + a)$$

gives the result. ∎

The fundamental result on extending continuous functions is the following important theorem.

[11] Pavel Samuilovich Urysohn (1898–1924) was a Czechoslovakian mathematician whose main contributions were to analysis and topology. He passed at a young age from drowning while swimming off the coast of Brittany.

1.10.43 Theorem (Tietze[12] Extension Theorem) *Let* $S \subseteq \mathbb{R}^n$ *and let* $a, b \in \mathbb{R}$ *satisfy* $a < b$. *If* $A \subseteq S$ *is relatively closed and if* $f: A \to [a, b]$ *is continuous, then there exists a continuous function* $\hat{f}: S \to [a, b]$ *such that* $\hat{f}|A = f$.

Proof Suppose that $f(A) \subseteq [a, b]$ for $a < b$. Then define

$$g(\boldsymbol{x}) = \frac{2}{b-a} f(\boldsymbol{x}) - \frac{b+a}{b-a}$$

so that $g(A) \subseteq [-1, 1]$. We will first show that there exists a continuous function $\hat{g}: S \to [-1, 1]$ such that $\hat{g}|A = g$. This will prove the theorem since we can take

$$\hat{f}(\boldsymbol{x}) = \tfrac{1}{2}(b-a)\hat{g}(\boldsymbol{x}) + \tfrac{1}{2}(b+a).$$

For $j \in \mathbb{Z}_{\geq 0}$ define $r_j = \frac{1}{2}\left(\frac{2}{3}\right)^j$ and note that

1. $r_1 = \frac{1}{3}$,
2. $r_{j+1} < r_j$ for $j \in \mathbb{Z}_{>0}$, and
3. $\sum_{j=1}^{\infty} r_j = 1$ (by Example I-2.4.2–1).

We shall define continuous functions $g_j: A \to [-3r_j, 3r_j]$. We do this inductively, first taking $g_1 = g$. Suppose now that we have defined $g_1, \dots, g_k$. Define

$$B_{-,k} = \{\boldsymbol{x} \in A \mid \ g_k(\boldsymbol{x}) \leq -r_k\}, \quad B_{+,k} = \{\boldsymbol{x} \in A \mid \ g_k(\boldsymbol{x}) \geq r_k\}.$$

Since f_k is continuous, the sets $B_{-,k}$ and $B_{+,k}$ are relatively closed in A, and so relatively closed in S since A is relatively closed in S. They are also certainly disjoint. Thus, by Urysohn's Lemma, there exists a continuous function $h_k: S \to [-r_k, r_k]$ having the property that $h_k(\boldsymbol{x}) = -r_k$ for $\boldsymbol{x} \in B_{-,k}$ and $h_k(\boldsymbol{x}) = r_k$ for $\boldsymbol{x} \in B_{+,k}$. Now define $g_{k+1} = g_k - h_k|A$ which is a continuous function on A by Propositions 1.3.22 and 1.3.24. Moreover, since $g_k(A) \subseteq [-3r_k, 3r_k]$ and since $h_k(A) \subseteq [-r_k, r_k]$ it follows that

$$g_{k+1}(A) \subseteq [-2r_k, 2r_k] = [-3r_{k+1}, 3r_{k+1}].$$

Now define $\hat{g}(\boldsymbol{x}) = \sum_{k=1}^{\infty} h_k(\boldsymbol{x})$ for $\boldsymbol{x} \in S$. Since $h_k(S) \in [-r_k, r_k]$ and since $\sum_{k=1}^{\infty} r_k$ converges, it follows from the Weierstrass M-test, Theorem 1.7.1, that the series defining $\hat{g}$ converges uniformly and so $\hat{g}$ is a continuous function. For $\boldsymbol{x} \in A$ we have

$$(h_1(\boldsymbol{x}) + \cdots + h_k(\boldsymbol{x}) = (g_1(\boldsymbol{x}) - g_2(\boldsymbol{x}) + \cdots + (g_k(\boldsymbol{x}) - g_{k+1}(\boldsymbol{x})) = g(\boldsymbol{x}) - g_{k+1}(\boldsymbol{x}).$$

Since $\lim_{k\to\infty} 3r_k = 0$ and since $f_k(A) \subseteq [-3r_k, 3r_k]$ it follows that

$$\hat{g}(\boldsymbol{x}) = g(\boldsymbol{x}) - \lim_{k\to\infty} g_{k+1}(\boldsymbol{x}) = g(\boldsymbol{x}), \qquad \boldsymbol{x} \in A.$$

Thus $\hat{g}|A = g$. Also, for any $\boldsymbol{x} \in S$,

$$|\hat{g}(\boldsymbol{x})| \leq \sum_{k=1}^{\infty} |h_k(\boldsymbol{x})| \leq \sum_{k=1}^{\infty} r_k = 1.$$

Thus $\hat{g}(S) \subseteq [-1, 1]$, as desired. ■

There are some fairly elementary consequences and extensions of the Tietze Extension Theorem that are worth recording.

[12]Heinrich Franz Friedrich Tietze (1880-1964) was Austrian and made contributions to the subject of topology

1.10.44 Corollary (Extensions of possibly unbounded functions) *Let* $\mathrm{S} \subseteq \mathbb{R}^{\mathrm{n}}$. *If* $\mathrm{A} \subseteq \mathrm{S}$ *is relatively closed and if* $\mathrm{f}\colon \mathrm{A} \to \mathbb{R}$ *is continuous, then there exists a continuous function* $\hat{\mathrm{f}}\colon \mathrm{S} \to \mathbb{R}$ *such that* $\hat{\mathrm{f}}|\mathrm{A} = \mathrm{f}$.

Proof Consider the map $\phi\colon \mathbb{R} \to (-1,1)$ given by $\phi(x) = \frac{2}{\pi}\tan^{-1}(x)$. By Proposition I-3.8.20 this map is a continuous bijection with a continuous inverse. Thus the function $\phi \circ f\colon A \to [-1,1]$ satisfies the hypotheses of the Tietze Extension Theorem. Therefore, there exists a continuous function $g\colon S \to [-1,1]$ agreeing with $\phi \circ f$ on A. Define

$$B = \{\boldsymbol{x} \in S \mid |g(\boldsymbol{x})| = 1\}.$$

Since the function $|g|$ is continuous it follows that B is closed. It is also evident that A and B are disjoint since $|\phi \circ f(\boldsymbol{x})| < 1$ for $\boldsymbol{x} \in A$. By Urysohn's Lemma let $h\colon S \to [0,1]$ have the property that $h(\boldsymbol{x}) = 1$ for $\boldsymbol{x} \in A$ and $h(\boldsymbol{x}) = 0$ for $\boldsymbol{x} \in B$. Then define $\hat{f}(\boldsymbol{x}) = \phi^{-1}(h(\boldsymbol{x})g(\boldsymbol{x}))$ to get the result. ■

1.10.45 Corollary (Extensions of certain bounded multivariable-valued maps) *Let* $\mathrm{S} \subseteq \mathbb{R}^{\mathrm{n}}$. *If* $\mathrm{A} \subseteq \mathrm{S}$ *is relatively closed and if* $\mathbf{f}\colon \mathrm{A} \to [0,1]^{\mathrm{m}}$ *is continuous, then there exists a continuous map* $\hat{\mathbf{f}}\colon \mathrm{S} \to [0,1]^{\mathrm{m}}$ *such that* $\hat{\mathbf{f}}|\mathrm{A} = \mathbf{f}$.

Proof This follows by applying the Tietze Extension Theorem to each of the components of $\boldsymbol{f}$. ■

1.10.46 Corollary (Extensions of certain unbounded multivariable-valued maps) *Let* $\mathrm{S} \subseteq \mathbb{R}^{\mathrm{n}}$. *If* $\mathrm{A} \subseteq \mathrm{S}$ *is relatively closed and if* $\mathbf{f}\colon \mathrm{A} \to \mathbb{R}^{\mathrm{m}}$ *is continuous, then there exists a continuous map* $\hat{\mathbf{f}}\colon \mathrm{S} \to \mathbb{R}^{\mathrm{m}}$ *such that* $\hat{\mathbf{f}}|\mathrm{A} = \mathbf{f}$.

Proof This follows by applying Corollary 1.10.44 to each of the components of $\boldsymbol{f}$. ■

The Tietze Extension Theorem can make a person inappropriately optimistic. The tendency might be to think that every map from a relatively closed subset of a set can be extended to the whole set. Let us give an example to illustrate why this is not so.

1.10.47 Example (A continuous map on a closed subset not admitting an extension) We let $S = \overline{\mathsf{B}}^2(1,\mathbf{0}) \subseteq \mathbb{R}^2$ be the closed ball of radius 1 about the origin so that $A = \mathbb{S}^1$, the unit circle in $\mathbb{R}^2$, is the boundary of S and so a closed subset of S. Then define $\boldsymbol{f}\colon A \to \mathbb{S}^1$ by $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$. Note that $\boldsymbol{f}(A) \subseteq [0,1]^2$ and so, by Corollary 1.10.45, it follows that there exists a continuous map $\hat{\boldsymbol{f}}\colon S \to [0,1]^2$ such that $\hat{\boldsymbol{f}}|A = \boldsymbol{f}$. However, let us instead ask a different question: "Can $\boldsymbol{f}$ be extended to a map from S *taking values in* $\mathbb{S}^1$?" Thus we restrict the codomain of the map as well as extend the domain. In this case, it is actually *not* possible to find such an extension. This is not quite trivial to prove, and indeed follows from the nontrivial Proposition 1.11.9 to the Brouwer Fixed Point Theorem.

However, let us see if we can understand the impossibility of an extension in this case. Suppose that there does exist a continuous map $\hat{\boldsymbol{f}}\colon \overline{\mathsf{B}}^2(1,\mathbf{0}) \to \mathbb{S}^1$ that restricts to $\boldsymbol{f}$, the identity map on $\mathbb{S}^1 \subseteq \overline{\mathsf{B}}^1(1,\mathbf{0})$. This means that one maps each point $x \in \text{int}(\overline{\mathsf{B}}^2(1, \mathbf{0}))$ to some point on the boundary. Think of points in $\overline{\mathsf{B}}^2(1, \mathbf{0})$ as being points in a circular sheet, the map $\hat{f}$ "rolls up" the sheet to the boundary. One can imagine that this is not possible without tearing the sheet. This tearing prohibits continuity. This is the intuition of Proposition 1.11.9. •

The preceding example, while seemingly simple, actually captures the essence of the continuous extension problem. Let us explore this. The "rolling up" in the example leads to the following definition.

1.10.48 Definition (Retract) Let $S \subseteq \mathbb{R}^n$. A subset $A \subseteq S$ is a ***retract*** of S if there exists a continuous map $\boldsymbol{r}\colon S \to A$ having the property that $\boldsymbol{r}|A = \text{id}_A$. The map $\boldsymbol{r}$ is a ***retraction*** of S onto A. •

We now have the following result which illustrates the importance of retracts to extending continuous functions.

1.10.49 Theorem (Continuous functions from retracts always possess extensions) *For $S \subseteq \mathbb{R}^n$ and for $A \subseteq S$ the following statements are equivalent:*

(i) A is a retract of S;

(ii) for every $m \in \mathbb{Z}_{>0}$, for every subset $B \subseteq \mathbb{R}^m$, and for every continuous map $\mathbf{f}\colon A \to B$ there exists a continuous map $\hat{\mathbf{f}}\colon S \to B$ such that $\hat{\mathbf{f}}|A = \mathbf{f}$.

Proof (i) $\Longrightarrow$ (ii) Let $B \subseteq \mathbb{R}^m$, let $\boldsymbol{f}\colon A \to B$ be continuous, and let $\boldsymbol{r}\colon S \to A$ be a retraction. Then $\hat{\boldsymbol{f}} = \boldsymbol{f} \circ \boldsymbol{r}$ is a continuous extension of $\boldsymbol{f}$ from A to S.

(ii) $\Longrightarrow$ (i) Consider the map $\boldsymbol{f}\colon A \to A$ given by $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{x}$. The hypotheses ensure that there exists $\hat{\boldsymbol{f}}\colon S \to A$ such that $\hat{\boldsymbol{f}}|A = \boldsymbol{f}$. This means exactly that $\hat{\boldsymbol{f}}$ is a retraction of S onto A. ■

Let us give some examples of retracts.

1.10.50 Examples (Retracts)

1. Let $S = \overline{\mathsf{B}}^n(1, \mathbf{0}) \setminus \{\mathbf{0}\}$ be the closed unit ball with the origin removed and let $A = \mathbb{S}^{n-1} = \text{bd}(\overline{\mathsf{B}}^n(1, \mathbf{0}))$. The map $\boldsymbol{r}\colon S \to A$ defined by $\boldsymbol{r}(\boldsymbol{x}) = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_{\mathbb{R}^n}}$ is a retraction of S onto A.
2. Let $S = \mathbb{R}^n \setminus \{\mathbf{0}\}$ and let $A = \mathbb{S}^{n-1}$. The map $\boldsymbol{r}\colon S \to A$ defined by $\boldsymbol{r}(\boldsymbol{x}) = \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_{\mathbb{R}^n}}$ is a retraction of S onto A.
3. Let $S = \overline{\mathsf{B}}^n(1, \mathbf{0})$ and let $A = \mathbb{S}^{n-1} = \text{bd}(S)$. As we claimed above in the case of $n = 2$, and as we shall prove in Proposition 1.11.9, A is not a retract of S.
4. Let $S = \overline{\mathsf{B}}^n(1, \mathbf{0})$ and let $A = \{\mathbf{0}\}$. Then $\boldsymbol{r}\colon S \to A$ be given by $\boldsymbol{r}(\boldsymbol{x}) = \mathbf{0}$ is a retraction. •

Let us close this section with a nontrivial result concerning continuous extensions of functions. It is very much not clear why a result like this is important, but it actually plays a crucial rôle in the proof of the extremely important Domain Invariance Theorem, which we state and prove as Theorem 1.3.44.

1.10.51 Theorem (Extending certain maps to spheres) *Let $\mathrm{K} \subseteq \mathbb{R}^{\mathrm{n}}$ be compact and let $\mathrm{A} \subseteq \mathrm{K}$ be relatively closed and such that $\mathrm{K} \setminus \mathrm{A}$ has measure zero. If $\mathbf{f}\colon \mathrm{A} \to \mathbb{S}^{\mathrm{n}-1}$ is continuous then there exists a continuous map $\hat{\mathbf{f}}\colon \mathrm{K} \to \mathbb{S}^{\mathrm{n}-1}$ such that $\hat{\mathbf{f}}|\mathrm{A} = \mathbf{f}$.*

Proof Let $\epsilon \in (0,1)$ so $f(A) \cap \mathsf{B}^n(\epsilon, \mathbf{0}) = \emptyset$. By the Weierstrass Approximation Theorem (more precisely, by Corollary 1.7.5 to the Weierstrass Approximation Theorem) let $f_1\colon \mathbb{R}^n \to \mathbb{R}^n$ be such that

$$\|f(x) - f_1(x)\|_{\mathbb{R}^n} < \tfrac{\epsilon}{8}, \qquad x \in A.$$

Let us prove a useful lemma.

1 **Lemma** *If $\mathrm{K} \subseteq \mathbb{R}^{\mathrm{n}}$ is compact, if $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ is a neighbourhood of K, if $\mathrm{Z} \subseteq \mathrm{K}$ has measure zero, and if $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{n}}$ is differentiable, then $\mathbf{f}(\mathrm{Z})$ has measure zero.*

Proof By Theorem 1.10.58 below of the proof of Theorem 1.11.5 it holds that there exists $M \in \mathbb{R}_{>0}$ for which

$$\|f(y) - f(x)\|_{\mathbb{R}^n} \le M\|y - x\|_{\mathbb{R}^n}, \qquad x, y \in K. \tag{1.59}$$

Now let $\epsilon \in \mathbb{R}_{>0}$ and let $(\mathsf{B}^n(r_j, x_j))_{j \in \mathbb{Z}_{>0}}$ be a cover of Z by balls whose volume sums to at most $\frac{\epsilon}{M^n}$. We suppose without loss of generality that the points x_j, $j \in \mathbb{Z}_{>0}$, are in U. Then

$$f(\mathsf{B}^n(r_j, x_j)) \subseteq \mathsf{B}^n(Mr_j, f(x_j))$$

by (1.59). The balls $(\mathsf{B}^n(Mr_j, f(x_j)))_{j \in \mathbb{Z}_{>0}}$ cover $f(Z)$. Moreover,

$$\sum_{j=1}^{\infty} \mathrm{vol}(\mathsf{B}^n(Mr_j, f(x_j)) = \sum_{j=1}^{\infty} M^n \mathrm{vol}(\mathsf{B}^n(r_j, \mathbf{0}) = M^n \sum_{j=1}^{\infty} \mathrm{vol}(\mathsf{B}^n(r_j, x_j) < \epsilon.$$

This gives the result. ▼

By the lemma $f_1(K \setminus A)$ cannot contain a neighbourhood of $\mathbf{0}$. Therefore, let $x_0 \in \mathsf{B}^n(\frac{\epsilon}{8}, \mathbf{0})$ be a point not in $f_1(K \setminus A)$. Then define $g\colon \mathbb{R}^n \to \mathbb{R}^n$ by $g(x) = f_1(x) - x_0$. Note that

$$\|f(x) - g(x)\|_{\mathbb{R}^n} \le \|f(x) - f_1(x)\|_{\mathbb{R}^n} + \|f_1(x) - g(x)\|_{\mathbb{R}^n} < \tfrac{\epsilon}{4}, \qquad x \in A, \tag{1.60}$$

and

$$\|f(x) - x_0\|_{\mathbb{R}^n} \ge |\|f(x)\|_{\mathbb{R}^n} - \|x_0\|_{\mathbb{R}^n}| \ge \tfrac{7\epsilon}{8}, \qquad x \in A,$$

using Exercise 1.1.3 and the fact that $\|f(x)\|_{\mathbb{R}^n} > \epsilon$ for every $x \in A$. Again using Exercise 1.1.3 we have

$$\|g(x)\|_{\mathbb{R}^n} \ge |\|f_1(x) - f(x)\|_{\mathbb{R}^n} - \|f(x) - x_0\|_{\mathbb{R}^n}| > \tfrac{3\epsilon}{4}, \qquad x \in A. \tag{1.61}$$

Define $r\colon K \to \mathbb{R}$ by $r(x) = \max\{\|g\|_{\mathbb{R}^n}(x), \frac{\epsilon}{2}\}$, and note that r is continuous (why?). From (1.61) we have

$$r(x) = \|g(x)\|_{\mathbb{R}^n}, \qquad x \in A. \tag{1.62}$$

If we define $\hat{g}\colon K \to \mathbb{R}^n$ by

$$\hat{g}(x) = r(x)\frac{g(x)}{\|g(x)\|_{\mathbb{R}^n}}$$

we then have

$$\hat{g}(x) = g(x), \qquad x \in A \tag{1.63}$$

and

$$\|\hat{g}\|_{\mathbb{R}^n}(x) \geq \tfrac{\epsilon}{2}, \qquad x \in K, \tag{1.64}$$

by (1.62) and by definition of r. Define $h\colon A \to \mathbb{R}^n$ by

$$h(x) = f(x) - \hat{g}(x).$$

By (1.60) and (1.63) we get

$$\|h(x)\|_{\mathbb{R}^n} < \tfrac{\epsilon}{4}, \qquad x \in A.$$

That is, $h(A) \subseteq \mathsf{B}^n(\frac{\epsilon}{4}, \mathbf{0})$. As per Exercise 1.3.15 there exists a continuous bijection $\phi\colon \mathsf{B}^n(\frac{\epsilon}{4}, \mathbf{0}) \to [0,1]^n$ with a continuous inverse. Thus $\phi \circ h\colon A \to [0,1]^n$ can be extended to a continuous map $\tilde{h}\colon K \to [0,1]^n$ in view of Theorem **??**(**??**). Then $\hat{h} = \phi^{-1} \circ \tilde{h}$ is a continuous map from K to $\mathsf{B}^n(\frac{\epsilon}{4}, \mathbf{0})$ which agrees with h on A.

Next define $\tilde{f}\colon K \to \mathbb{R}^n$ by $\tilde{f} = \hat{h} + \hat{g}$. Note that, for $x \in A$, we have

$$\tilde{f}(x) = \hat{h}(x) + \hat{g}(x) = h(x) + \hat{g}(x) = f(x).$$

Thus $\tilde{f}$ extends f. We claim that $\mathbf{0} \notin \operatorname{image}(\tilde{f})$. Indeed, since $\operatorname{image}(\hat{h}) \subseteq \mathsf{B}^n(\frac{\epsilon}{4}, \mathbf{0})$ and by (1.61), we have

$$\|\tilde{f}(x)\|_{\mathbb{R}^n} = \|\hat{g}(x) - (-\hat{h}(x))\|_{\mathbb{R}^n} \geq \left| \|\hat{g}(x)\|_{\mathbb{R}^n} - \|\hat{h}(x)\|_{\mathbb{R}^n} \right| \geq \tfrac{\epsilon}{2}.$$

Now define $\hat{f}\colon K \to \mathbb{S}^{n-1}$ by

$$\hat{f}(x) = \frac{\tilde{f}(x)}{\|\tilde{f}(x)\|_{\mathbb{R}^n}}$$

to give the theorem. ■

This theorem has many fascinating corollaries that really serve to illustrate that this business of extending continuous maps is rather more intricate than it seems. One of these is the following.

1.10.52 Corollary (Extensions of continuous maps between spheres) *For* $m, n \in \mathbb{Z}_{>0}$ *the following two statements are equivalent:*

(i) $m \leq n$;

(ii) for every relatively closed subset $A \subseteq \mathbb{S}^m$ *and every continuous map* $\mathbf{f}\colon A \to \mathbb{S}^n$ *there exists a continuous map* $\hat{\mathbf{f}}\colon \mathbb{S}^m \to \mathbb{S}^n$ *for which* $\hat{\mathbf{f}}|A = \mathbf{f}$.

Proof First suppose that $m \leq n$ and consider $\mathbb{S}^m$ to be the subset of $\mathbb{R}^{n+1}$ given by

$$\mathbb{S}^m = \{(x_1, \dots, x_{n+1}) \in \mathbb{R}^{n+1} \mid x_1^2 + \dots + x_{m+1}^2 = 1,\ x_{m+2} = \dots = x_{n+1} = 0\}.$$

Then $\mathbb{S}^m$ is compact and has measure zero. Let $A \subseteq \mathbb{S}^m$ and let $f\colon A \to \mathbb{S}^n$. By Theorem 1.10.51 there exists a continuous extension $\hat{f}\colon \mathbb{S}^m \to \mathbb{S}^n$.

Next, for $m > n$ we shall give a closed subset $A \subseteq \mathbb{S}^m$ and a continuous map $f\colon A \to \mathbb{S}^n$ that does not possess an extension to $\mathbb{S}^m$. We let

$$A = \{(x_1, \dots, x_{m+1}) \in \mathbb{S}^m \mid x_1^2 + \dots + x_{n+1}^2 = 1,\ x_{n+2} = \dots = x_{m+1} = 0\}$$

and define $f\colon A \to \mathbb{S}^n$ by

$$f(x_1, \dots, x_{n+1}, 0, \dots, 0) = (x_1, \dots, x_{n+1}).$$

If $\hat{f}\colon \mathbb{S}^m \to \mathbb{S}^n$ is an extension of f then

$$\hat{f}|\{(x_1, \dots, x_{m+1}) \in \mathbb{R}^{m+1} \mid x_1^2 + \dots + x_{n+1}^2 \leq 1,\ x_{n+2} = \dots = x_{m+1}\}$$

then defines a retraction of the closed unit ball in $\mathbb{R}^{m+1}$ onto its boundary, in contradiction to Proposition 1.11.9. ■

While we have successfully given elementary (different from "easy") proofs of some important results concerning extensions of continuous maps, this subject is really most naturally presented in the language of either dimension theory or algebraic topology. We refer to Section 1.10.10 for a discussion.

1.10.7 Partitions of unity

1.10.8 Lipschitz maps

The notion of a Lipschitz map arises naturally in many settings. As we shall see, a Lipschitz map can be thought of as being "between" continuous and differentiable.

Let us give the definition.

1.10.53 Definition ((Locally) Lipschitz map) Let $A \subseteq \mathbb{R}^n$. A map $f\colon A \to \mathbb{R}^m$

(i) is ***Lipschitz*** if there exists $L \in \mathbb{R}_{>0}$ such that

$$\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} \leq L\|x_1 - x_2\|_{\mathbb{R}^n}, \qquad x_1, x_2 \in A, \tag{1.65}$$

and

(ii) is ***locally Lipschitz*** if $f|K$ is Lipschitz for every relatively compact subset $K \subseteq A$.

For a Lipschitz map $f\colon A \to \mathbb{R}^m$,

(iii) a number L satisfying (1.65) is a ***Lipschitz constant*** for f and

(iv) $\|f\|_{\mathrm{Lip}} = \inf\{L \in \mathbb{R}_{>0} \mid \text{(1.65) holds}\}$ is the ***Lipschitz norm*** of f. •

The following result is useful.

1.10.54 Proposition (Property of the Lipschitz norm) *If* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *and if* $\mathbf{f}\colon \mathrm{A} \to \mathbb{R}^{\mathrm{m}}$ *is Lipschitz, then*

$$\|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\|_{\mathbb{R}^{\mathrm{m}}} \le \|\mathbf{f}\|_{\mathrm{Lip}} \|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbb{R}^{\mathrm{n}}}$$

for every $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{A}$.

Proof Suppose that there exists $x_1, x_2 \in A$ satisfying

$$\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} > \|f\|_{\mathrm{Lip}} \|x_1 - x_2\|_{\mathbb{R}^n}.$$

Then there exists $L > \|f\|_{\mathrm{Lip}}$ such that

$$\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} > L\|x_1 - x_2\|_{\mathbb{R}^n},$$

contradicting the definition of $\|f\|_{\mathrm{Lip}}$. ■

It is obvious that Lipschitz maps are also locally Lipschitz. In Exercise 1.10.5 the reader can see an example of a map that is locally Lipschitz but not Lipschitz.

First let us state that Lipschitz maps are often continuous.

1.10.55 Proposition ((Locally) Lipschitz maps are continuous) *For* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *and for* $\mathbf{f}\colon \mathrm{A} \to \mathbb{R}^{\mathrm{m}}$*, the following statements hold:*

(i) if $\mathbf{f}$ *is Lipschitz then* $\mathbf{f}$ *is continuous;*

(ii) if A *is locally compact and if* $\mathbf{f}$ *is locally Lipschitz, then* $\mathbf{f}$ *is continuous.*

Proof (i) Let $\epsilon \in \mathbb{R}_{>0}$ and note that, if $x_1, x_2 \in A$ satisfy $\|x_1 - x_2\|_{\mathbb{R}^n} < \frac{\epsilon}{\|f\|_{\mathrm{Lip}}}$, then

$$\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} \le \|f\|_{\mathrm{Lip}} \|x_1 - x_2\|_{\mathbb{R}^n} < \epsilon.$$

(ii) Let $x \in A$. Let $U \subseteq A$ be a relative neighbourhood of x such that $\mathrm{cl}_A(U)$ is relatively compact, this being possible since A is locally compact. Since f is locally Lipschitz, there exists $L \in \mathbb{R}_{>0}$ such that

$$\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} \le L\|x_1 - x_2\|_{\mathbb{R}^n}$$

for all $x_1, x_2 \in \mathrm{cl}_A(U)$. From part (i) we conclude that $f|\mathrm{cl}_A(U)$ is continuous, and so f is continuous at x. ■

Let us show that continuous maps are not necessarily Lipschitz.

1.10.56 Example (Continuous but not Lipschitz map) The function $f\colon \mathbb{R} \to \mathbb{R}$ defined by $f(x) = \sqrt{|x|}$ is continuous but not Lipschitz. To see that f is not Lipschitz, note that

$$\lim_{x \downarrow 0} \frac{f(x) - f(0)}{x - 0} = \lim_{x \downarrow 0} \frac{\sqrt{x}}{x} = \infty$$

by L'Hôpital's Rule. This prohibits f from being Lipschitz in any neighbourhood of 0. •

The following characterisation of locally Lipschitz maps is useful.

1.10.57 Proposition (Characterisation of locally Lipschitz maps) *Let* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *be locally compact. For a map* $\mathbf{f}\colon \mathrm{A} \to \mathbb{R}^{\mathrm{m}}$ *the following statements are equivalent:*

(i) $\mathbf{f}$ *is locally Lipschitz;*

(ii) for every $\mathbf{x} \in \mathrm{A}$ *there exists* $\mathrm{r} \in \mathbb{R}_{>0}$ *and* $\mathrm{L} \in \mathbb{R}_{>0}$ *such that*

$$\|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\|_{\mathbb{R}^{\mathrm{m}}} \le \mathrm{L}\|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbb{R}^{\mathrm{n}}}$$

for every $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{A} \cap \mathsf{B}(\mathrm{r}, \mathbf{x})$.

Proof (i) $\Longrightarrow$ (ii) Let $x \in A$ and let $U \subseteq A$ be a relative neighbourhood of x such that $\mathrm{cl}_A(U)$ is relatively compact in A, this being possible since A is locally compact. There exists $r \in \mathbb{R}_{>0}$ such that $\mathsf{B}^n(r, x) \cap A \subseteq U$. Since $\mathrm{cl}_A(U)$ is compact and f is locally Lipschitz, it follows that there exists $L \in \mathbb{R}_{>0}$ such that

$$\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} \le L\|x_1 - x_2\|_{\mathbb{R}^n}$$

for all $x_1, x_2 \in \mathrm{cl}_A(U)$. In particular, this inequality holds for $x_1, x_2 \in A \cap \mathsf{B}(r, x)$.

(ii) $\Longrightarrow$ (i) Let $K \subseteq A$ be relatively compact. For each $x \in K$, let $r_x, L_x \in \mathbb{R}_{>0}$ be such that

$$\|f(x_1) - x_2\|_{\mathbb{R}^m} \le L_x\|x_1 - x_2\|_{\mathbb{R}^n}$$

for every $x_1, x_2 \in \mathsf{B}(r_x, x) \cap A$. Let us abbreviate $U_x = \mathsf{B}(r_x, x) \cap A$. Note that $(U_x)_{x \in A}$ is an open cover of K. Thus there exists $y_1, \dots, y_k \in K$ such that $K \subseteq \cup_{j=1}^k U_{y_j}$. Since K is compact and f is continuous, f is bounded. Let $M \in \mathbb{R}_{>0}$ be such that

$$\|f(x)\|_{\mathbb{R}^m} \le M, \qquad x \in K.$$

Let

$$L = \max\{L_{y_1}, \dots, L_{y_k}, \tfrac{2M}{\epsilon}\}.$$

By Theorem 1.2.38 there exists $\epsilon \in \mathbb{R}_{>0}$ such that if $x_1, x_2 \in K$ satisfy $\|x_1 - x_2\|_{\mathbb{R}^n} \le \epsilon$, then $x_1, x_2 \in U_{y_j}$ for some $j \in \{1, \dots, k\}$. Now let $x_1, x_2 \in K$. If $\|x_1 - x_2\|_{\mathbb{R}^n} < \epsilon$ then, as we just indicated, $x_1, x_2 \in U_{y_j}$ for some $j \in \{1, \dots, k\}$. Therefore,

$$\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} \le L_{y_j}\|x_1 - x_2\|_{\mathbb{R}^n} \le L\|x_1 - x_2\|_{\mathbb{R}^n}.$$

If $\|x_1 - x_2\|_{\mathbb{R}^n} \ge \epsilon$ then

$$\|f(x_1) - f(x_2)\|_{\mathbb{R}^m} \le \|f(x_1)\|_{\mathbb{R}^m} + \|f(x_2)\|_{\mathbb{R}^m} \le \frac{2M}{\epsilon}\epsilon \le L\|x_1 - x_2\|_{\mathbb{R}^n},$$

giving the result. ■

We have established the relationship between continuous and Lipschitz maps. Let us now consider the relationship between differentiable and Lipschitz maps. Let us first show that differentiable maps are locally Lipschitz.

1.10.58 Theorem (Differentiable maps are locally Lipschitz) *If* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *and if* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{k}}$ *is of class* C^1*, then* $\mathbf{f}$ *is locally Lipschitz.*

Proof Let $K \subseteq U$ be compact. Let $B \subseteq U$ be an open ball and let $\boldsymbol{x}, \boldsymbol{y} \in B$. By the Mean Value Theorem,

$$\|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^{n+1}} \leq M_B \|\boldsymbol{y} - \boldsymbol{x}\|_{\mathbb{R}^{n+1}}, \qquad \boldsymbol{x}, \boldsymbol{y} \in B,$$

where

$$M_B = \sup\{\|\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^{n+1}} \mid \boldsymbol{x} \in B\}.$$

Now, since K is compact, we can cover it with a finite number of balls $B_1, \dots, B_N$, each contained in U. Let us denote

$$C = (K \times K) - \cup_{j=1}^{N} B_j \times B_j$$

and note that C is compact. Moreover, the function $d\colon \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}_{\geq 0}$ defined by $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^{n+1}}$ is strictly positive when restricted to C. Therefore, there exists $m \in \mathbb{R}_{>0}$ such that $d(\boldsymbol{x}, \boldsymbol{y}) \geq m$ for all $(\boldsymbol{x}, \boldsymbol{y}) \in C$. Let

$$M_0 = \sup\{\|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^{n+1}} \mid \boldsymbol{x}, \boldsymbol{y} \in K\},$$

noting that this number is finite since $\boldsymbol{f}$ is continuous and K is compact. Now define

$$L = \max\{\tfrac{M_0}{m}, M_{B_1}, \dots, M_{B_N}\}.$$

Now let $\boldsymbol{x}, \boldsymbol{y} \in K$. If $\boldsymbol{x}, \boldsymbol{y} \in B_j$ for some $j \in \{1, \dots, N\}$ then

$$\|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^{n+1}} \leq M_{B_j} \|\boldsymbol{y} - \boldsymbol{x}\|_{\mathbb{R}^{n+1}} \leq L \|\boldsymbol{y} - \boldsymbol{x}\|_{\mathbb{R}^{n+1}}.$$

If $\boldsymbol{x}$ and $\boldsymbol{y}$ are not together contained in any of the balls $B_1, \dots, B_N$ then $(\boldsymbol{x}, \boldsymbol{y}) \in C$. Thus

$$\|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^{n+1}} \leq M_0 = \tfrac{M_0}{m} m \leq L \|\boldsymbol{y} - \boldsymbol{x}\|_{\mathbb{R}^{n+1}}.$$

Thus we have

$$\|\boldsymbol{f}(\boldsymbol{y}) - \boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^{n+1}} \leq L \|\boldsymbol{y} - \boldsymbol{x}\|_{\mathbb{R}^{n+1}}, \qquad \boldsymbol{x}, \boldsymbol{y} \in K,$$

as desired. ■

There are Lipschitz maps that are not differentiable.

1.10.59 Example (Lipschitz but not differentiable map) The function $x \mapsto |x|$ is Lipschitz, but is not differentiable at $x = 0$. •

What is true is that Lipschitz maps are "nearly" differentiable, in a sense made precise in the following important theorem.

1.10.60 Theorem (Rademacher's[13] Theorem) *Let* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *be open. If* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$ *is locally Lipschitz, then the following statements hold:*

(i) $\mathbf{f}$ *is differentiable almost everywhere;*

(ii) if $\mathbf{f}$ *is differentiable at* $\mathbf{x} \in \mathrm{U}$ *and if* $\mathrm{L}_{\mathbf{x}} \in \mathbb{R}_{>0}$ *and a neighbourhood* $\mathrm{V} \subseteq \mathrm{U}$ *of* $\mathbf{x}$ *are such that*

$$\|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\|_{\mathbb{R}^{\mathrm{m}}} \leq \mathrm{L}_{\mathbf{x}}\|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbb{R}^{\mathrm{n}}}$$

for all $\mathbf{x}_1, \mathbf{x}_2 \in \mathrm{V}$*, then*

$$\|\mathbf{D}\mathrm{f}(\mathbf{x})\|_{\mathbb{R}^{\mathrm{n}},\mathbb{R}^{\mathrm{m}}} \leq \mathrm{L}_{\mathbf{x}}.$$

Proof In the proof we shall make free use of advanced ideas that we have not yet covered. Forward references will be provided. To begin with, all integrals in the proof are Lebesgue integrals as discussed in Section III-2.10.

Let us first prove the theorem when $n = m = 1$.

1 Lemma *Let* $\mathrm{U} \subseteq \mathbb{R}$ *be open. If* $\mathrm{f}\colon \mathrm{U} \to \mathbb{R}$ *is locally Lipschitz, then the following statements hold:*

(i) f *is differentiable almost everywhere;*

(ii) if f *is differentiable at* $\mathrm{x} \in \mathrm{U}$ *and if* $\mathrm{L}_{\mathrm{x}} \in \mathbb{R}_{>0}$ *and a neighbourhood* $\mathrm{V} \subseteq \mathrm{U}$ *of* x *are such that*

$$|\mathrm{f}(\mathrm{x}_1) - \mathrm{f}(\mathrm{x}_2)| \leq \mathrm{L}_{\mathrm{x}}\|\mathrm{x}_1 - \mathrm{x}_2\|_{\mathbb{R}^{\mathrm{n}}}$$

for all $\mathrm{x}_1, \mathrm{x}_2 \in \mathrm{V}$*, then* $|\mathrm{f}'(\mathrm{x})| \leq \mathrm{L}_{\mathrm{x}}$*;*

(iii) if $\phi\colon \mathrm{U} \to \mathbb{R}$ *is infinitely differentiable with compact support, then*

$$\int_{\mathrm{U}} \mathrm{f}'(\mathrm{x})\phi(\mathrm{x}) = -\int_{\mathrm{U}} \mathrm{f}(\mathrm{x})\phi'(\mathrm{x})\,\mathrm{dx}.$$

Proof We first claim that if $I \subseteq \mathbb{R}$ is a compact interval and if $f\colon I \to \mathbb{R}$ is Lipschitz, then f is absolutely continuous (see Definition III-2.9.23). Indeed, let $\epsilon \in \mathbb{R}_{>0}$ and let $\delta = \frac{\epsilon}{L}$ where L is a Lipschitz constant for f. Let $((a_j, b_j))_{j\in\{1,\dots,k\}}$ be a finite family of disjoint intervals such that

$$\sum_{j=1}^{k} |b_j - a_j| < \delta.$$

Then

$$\sum_{j=1}^{l} |f(b_j) - f(a_j)| \leq \sum_{j=1}^{k} L|b_j - a_j| < \epsilon,$$

showing that f is indeed absolutely continuous. Thus f is differentiable almost everywhere by Theorem III-2.9.33.

Now, if $U \subseteq \mathbb{R}$ is open and if $f\colon U \to \mathbb{R}$ is locally Lipschitz, then, by Proposition 1.10.57, for each $x \in U$ there exists an open interval $I_x \subseteq\subseteq U$ containing x such that $f|I_x$ is Lipschitz with Lipschitz constant denoted by $L_x \in \mathbb{R}_{>0}$. Let $K_x \subseteq I_x$ be a

[13]Hans Rademacher (1892–1869) was a German mathematician whose primary mathematical contributions were to number theory, and who also made contributions to analysis.

compact subinterval containing x in its interior. From the preceding paragraph, since $f|K_x$ is Lipschitz, $f|K_x$ is absolutely continuous. Thus f is differentiable on the open set $U_x \triangleq \mathrm{int}(K_x)$ except possibly at points in a subset $Z_x \subseteq U_x$ of measure zero. Note that $(U_x)_{x\in U}$ is an open cover of U and so, by Lemma I-2.5.25, there exists $x_j \in U$, $j \in \mathbb{Z}_{>0}$, such that $U = \cup_{j\in\mathbb{Z}_{>0}} U_j$. It follows that f is differentiable except at points in the set $Z = \cup_{j\in\mathbb{Z}_{>0}} Z_j$. By Exercise I-2.5.11 it follows that Z has measure zero. Thus f is differentiable almost everywhere.

Next suppose that f is differentiable at $x_0 \in U$ and let $V \subseteq U$ be a neighbourhood of x_0 such that $f|V$ is Lipschitz with Lipschitz constant $L \in \mathbb{R}_{>0}$. Then, since

$$f'(x_0) = \lim_{x\to x_0} \frac{f(x_0) - f(x)}{x - x_0}$$

and since

$$-L \le \frac{f(x_0) - f(x)}{x - x_0} \le L,$$

it follows that $|f'(x_0)| \le L$, as claimed.

For the final assertion of the lemma, since U is open, it is a countable disjoint union of open intervals by Proposition I-2.5.6. Let $I \subseteq U$ be one such interval and let $K \subseteq I$ be a compact interval for which $\mathrm{supp}(\phi|I) \subseteq K$. By Proposition III-2.9.36, since ϕ must vanish at the endpoints of K,

$$\int_I f'(x)\phi(x)\,\mathrm{d}x = -\int_I f(x)\phi'(x)\,\mathrm{d}x.$$

Since ϕ has compact support, there are only finitely many such intervals $I \subseteq U$ such that $\mathrm{supp}(\phi) \cap I \neq \emptyset$. Summing over these finitely many intervals gives

$$\int_U f'(x)\phi(x)\,\mathrm{d}x = -\int_U f(x)\phi'(x)\,\mathrm{d}x,$$

as desired. ▼

Until we state otherwise in the proof, we shall suppose that

$$U = \{\boldsymbol{x}_0 + \boldsymbol{x} \mid x_j \in (-\epsilon, \epsilon),\ j \in \{1, \dots, n\}\},$$

i.e., that U is a cube of radius $\epsilon \in \mathbb{R}_{>0}$ and centre $\boldsymbol{x}_0$. We shall also assume that $f\colon U \to \mathbb{R}$ is Lipschitz with Lipschitz constant L.

Now let $U \subseteq \mathbb{R}^n$ and f satisfy these assumptions. As in Section 1.4.4, for $j \in \{1, \dots, n\}$ we denote the jth partial derivative of f at $\boldsymbol{x}$ by $\boldsymbol{D}_j f(\boldsymbol{x})$ when this derivative exists. If all partial derivatives of f exist at $\boldsymbol{x}$ then we denote

$$\hat{\boldsymbol{D}} f(\boldsymbol{x}) = \begin{bmatrix} \boldsymbol{D}_1 f(\boldsymbol{x}) & \dots & \boldsymbol{D}_n f(\boldsymbol{x}) \end{bmatrix} \in \mathrm{L}(\mathbb{R}^n; \mathbb{R}).$$

Note that we make no assumptions about the continuity of the partial derivatives, so we cannot necessarily conclude that f is differentiable if $\hat{\boldsymbol{D}} f(\boldsymbol{x})$ exists, cf. Example 1.4.13.

With the above notation, we prove another lemma.

2 Lemma *If* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *is an open cube of radius* ϵ *and if* $\mathrm{f}\colon \mathrm{U} \to \mathbb{R}$ *is Lipschitz with Lipschitz constant* L*, then the following statements hold:*

(i) $\hat{\mathbf{D}}\mathrm{f}(\mathbf{x})$ *exists for almost every* $\mathbf{x} \in \mathrm{U}$*;*

(ii) *if* $\hat{\mathbf{D}}\mathrm{f}(\mathbf{x})$ *exists, then* $|\mathbf{D}_{\mathrm{j}}\mathrm{f}(\mathbf{x})| \leq \mathrm{L}$, $\mathrm{j} \in \{1, \dots, \mathrm{n}\}$*;*

(iii) *if* $\phi\colon \mathrm{U} \to \mathbb{R}$ *is infinitely differentiable with compact support, then*

$$\int_{\mathrm{U}} \mathbf{D}_{\mathrm{j}}\mathrm{f}(\mathbf{x})\phi(\mathbf{x})\,\mathrm{d}\mathbf{x} = -\int_{\mathrm{U}} \mathrm{f}(\mathbf{x})\mathbf{D}_{\mathrm{j}}\phi(\mathbf{x})\,\mathrm{d}\mathbf{x}$$

for each $\mathrm{j} \in \{1, \dots, \mathrm{n}\}$.

Proof (i) First, fix $j \in \{1, \dots, n\}$ and for $\boldsymbol{x} = (x_1, \dots, x_n) \in U$ let us denote $\hat{x}_1 \in \mathbb{R}$ and $\hat{\boldsymbol{x}}_2 \in \mathbb{R}^{n-1}$ by

$$\hat{x}_1 = x_j, \quad \hat{\boldsymbol{x}}_2 = (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n).$$

We also define

$$\hat{U} = \{(\hat{x}_1, \hat{\boldsymbol{x}}_2) \in \mathbb{R}^n \mid \boldsymbol{x} \in U\}.$$

Since $\hat{U}$ is itself a cube of radius ϵ, we can write $\hat{U} = \hat{U}_1 \times \hat{U}_2$ where $\hat{U}_1 \subseteq \mathbb{R}$ is an interval of length 2ϵ and $\hat{U}_2$ is a cube or radius ϵ in $\mathbb{R}^{n-1}$. Let us also define $\hat{f}\colon \hat{U} \to \mathbb{R}$ by

$$\hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2) = f(\boldsymbol{x}).$$

Let us show that $\boldsymbol{D}_j f(\boldsymbol{x})$ exists for almost every $\boldsymbol{x} \in U$. This is clearly equivalent to showing that $\boldsymbol{D}_1 \hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2)$ exists for almost every $\boldsymbol{x} \in U$. Let us fix $\hat{\boldsymbol{x}}_2 \in \hat{U}_2$ and denote

$$Z_{j,\hat{\boldsymbol{x}}_2} = \{\hat{x}_1 \in \hat{U}_1 \mid \boldsymbol{D}_1 \hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2) \text{ exists}\}.$$

By Lemma 1, $Z_{j,\hat{\boldsymbol{x}}_2}$ has measure zero. We then define

$$Z_j = \{(\hat{x}_1, \hat{\boldsymbol{x}}_2) \in \hat{U} \mid \hat{x}_1 \in Z_{j,\hat{\boldsymbol{x}}_2}\}.$$

We claim that Z_j is measurable. Since $\hat{U}$ is a cube with radius ϵ we can write

$$\hat{U} = (\hat{x}_{01} - \epsilon, \hat{x}_{01} + \epsilon) \times \cdots \times (\hat{x}_{0n} - \epsilon, \hat{x}_{0n} + \epsilon)$$

for some $\hat{x}_{0j} \in \mathbb{R}$, $j \in \mathbb{R}$. For $j \in \mathbb{Z}_{>0}$ and $(\hat{x}_1, \hat{\boldsymbol{x}}_2) \in \hat{U}$ define

$$\rho(\hat{x}_1) = \epsilon - |\hat{x}_1 - \hat{x}_{01}|.$$

Note that

$$(\hat{x}_1 \pm h\rho(\hat{x}_1), \hat{\boldsymbol{x}}_2) \in \hat{U}$$

for every $\hat{\boldsymbol{x}}_2 \in \hat{U}_2$ and $h \in (-1, 1) \setminus \{0\}$. Note that $\boldsymbol{D}_1 \hat{f}(\hat{x}_1, \boldsymbol{x}_2)$ exists if and only if the limit

$$\lim_{h \to 0} \frac{\hat{f}(\hat{x}_1 + h\rho(\hat{x}_1), \hat{\boldsymbol{x}}_2) - \hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2)}{h\rho(\hat{x}_1)}$$

exists. Define

$$g_h(\hat{x}_1, \hat{\boldsymbol{x}}_2) = \frac{\hat{f}(\hat{x}_1 + h\rho(\hat{x}_1), \hat{\boldsymbol{x}}_2) - \hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2)}{h\rho(\hat{x}_1)}$$

and note that g_h is continuous for each $h \in (-1,1) \setminus \{0\}$. Thus, by Proposition 1.10.17, if we define $h_k(\hat{x}_1, \hat{\boldsymbol{x}}_2)$ by

$$h_k(\hat{x}_1, \hat{\boldsymbol{x}}_2) = \sup\{g_h(\hat{x}_1, \hat{\boldsymbol{x}}_2) \mid h \in (-\tfrac{1}{k}, \tfrac{1}{k}) \setminus \{0\}\},$$

then h_k is lower semicontinuous, and so $\mathscr{B}(\mathbb{R})$-measurable by Proposition 1.10.14. By Proposition III-2.6.18 it follows that if

$$d^+\hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2) = \limsup_{h \to 0} g_h(\hat{x}_1, \hat{\boldsymbol{x}}_2) = \inf\{h_k(\hat{x}_1, \hat{\boldsymbol{x}}_2) \mid k \in \mathbb{Z}_{>0}\},$$

(using Proposition I-2.3.15) then the function $d^+\hat{f}$ is measurable. In like manner, one shows that if

$$d^-\hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2) = \liminf_{h \to 0} g_h(\hat{x}_1, \hat{\boldsymbol{x}}_2,$$

then $d^-\hat{f}$ is measurable. Since the set of points where $\boldsymbol{D}_1\hat{f}$ does not exist is given by

$$\{(\hat{x}_1, \hat{\boldsymbol{x}}_2) \in \hat{U} \mid d^+\hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2) - d^-\hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2) > 0\},$$

we conclude that Z_j is measurable, by definition of measurability of functions.

Now, knowing that Z_j is measurable, we can use Fubini's Theorem in the form of Theorem III-2.8.2 to conclude that Z_j has measure zero. Since $\hat{\boldsymbol{D}}f(\boldsymbol{x})$ does not exist precisely at points in $Z = \cup_{j=1}^n Z_j$, this part of the theorem follows since Z has measure zero by virtue of the properties of measure.

(ii) Let $j \in \{1, \dots, n\}$ and retain the notation from the proof of the first part of the theorem. We have

$$|\boldsymbol{D}_j f(\boldsymbol{x})| = |\boldsymbol{D}_1\hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2)| = \left|\lim_{h \to 0} \frac{\hat{f}(\hat{x}_1 + h\rho(\hat{x}_1), \hat{\boldsymbol{x}}_2) - \hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2)}{h\rho(\hat{x}_1, \hat{\boldsymbol{x}}_2)}\right| \le L,$$

since L is a Lipschitz constant.

(iii) Here we define

$$\hat{\phi}(\hat{x}_1, \hat{\boldsymbol{x}}_2) = \phi(\boldsymbol{x})$$

and use Fubini's Theorem, the change of variable theorem, and Proposition III-2.9.36:

$$\begin{aligned}
\int_U \boldsymbol{D}jf(\boldsymbol{x})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \int_{\hat{U}_2}\Big(\int_{\hat{U}_1} \boldsymbol{D}_1\hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2)\hat{\phi}(\hat{x}_1, \hat{\boldsymbol{x}}_2)\,\mathrm{d}\hat{x}_1\Big)\,\mathrm{d}\hat{\boldsymbol{x}}_2 \\
&= -\int_{\hat{U}_2}\Big(\int_{\hat{U}_1} \hat{f}(\hat{x}_1, \hat{\boldsymbol{x}}_2)\boldsymbol{D}_1\hat{\phi}(\hat{x}_1, \hat{\boldsymbol{x}}_2)\,\mathrm{d}\hat{x}_1\Big)\,\mathrm{d}\hat{\boldsymbol{x}}_2 \\
&= \int_U f(\boldsymbol{x})\boldsymbol{D}j\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x},
\end{aligned}$$

as desired. ▼

Next we recall the notion of directional derivative from Section 1.4.3.

3 Lemma *If* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *is an open cube of radius* ϵ*, if* $\mathrm{f}\colon \mathrm{U} \to \mathbb{R}$ *is Lipschitz with Lipschitz constant* L*, and if* $\mathbf{v} \in \mathbb{R}^{\mathrm{n}}$*, then the directional derivative* $\mathbf{D}\mathrm{f}(\mathbf{x};\mathbf{v})$ *exists for almost every* $\mathbf{x} \in \mathrm{U}$ *and, for almost every* $\mathbf{x} \in \mathrm{U}$*,*

$$\mathbf{D}\mathrm{f}(\mathbf{x};\mathbf{v}) = \hat{\mathbf{D}}\mathrm{f}(\mathbf{x}) \cdot \mathbf{v}.$$

Proof Let $\{f_1, \dots, f_n\}$ be a basis for $\mathbb{R}^n$ such that $f_1 = v$. Then define a map $\phi\colon \mathbb{R}^n \to \mathbb{R}^n$ by asking that $\phi(x) = y$ with y uniquely defined by the requirement that

$$x = y_1 f_1 + \cdots + y_n f_n;$$

thus y is the vector of components of x in the basis $\{f_1, \dots, f_n\}$. Then denote $\hat{f} = f \circ \phi$. Note that

$$Df(x;v) = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} f(x + sv) = \frac{\mathrm{d}}{\mathrm{d}s}\bigg|_{s=0} f((y_1 + s)f_1 + \cdots + y_n f_n) = D_1\hat{f}(\phi(x)).$$

Thus the directional derivative $Df(x;v)$ exists if and only if $D_{\hat{f}}(\phi(x))$ exists. We now show two things.

1. ϕ^{-1} *maps sets of measure zero to sets of measure zero:* This follows from the change of variable theorem, , noting that ϕ^{-1} is an invertible linear map. what
2. $\hat{f}$ *is Lipschitz:* To see this we note that for $y_1 =, y_2 \in \phi(U)$,

$$\begin{aligned} |\hat{f}(y_1) - \hat{f}(y_2)| &= |f \circ \phi^{-1}(y_1) - f \circ \phi^{-1}(y_2)| \\ &\leq L\|\phi^{-1}(y_1 - y_2)\|_{\mathbb{R}^n} \leq L\|\phi^{-1}\|_{\mathbb{R}^n,\mathbb{R}}\|y_1 - y_2\|_{\mathbb{R}^n}, \end{aligned}$$

showing that $\hat{f}$ is Lipschitz with Lipschitz constant $L\|\phi^{-1}\|_{\mathbb{R}^n,\mathbb{R}}$.

Now it suffices to show that $D_1\hat{f}(y)$ exists for almost every $y \in \phi(U)$. We can do this using Lemma 2. The only problem is that $\phi(U)$ is not an open cube. However, if $y \in \phi(U)$ then there exists an open cube $V_y \subseteq \phi(U)$ containing y. One can use Lemma 2 to conclude that $D_1\hat{f}$ exists almost everywhere in V_y. Now, $(V_y)_{y\in\phi(U)}$ is an open cover of $\phi(U)$ and so, by Lemma 1.2.33, there exists $y_j \in \phi(U)$, $j \in \mathbb{Z}_{>0}$, such that $\phi(U) = \cup_{j\in\mathbb{Z}_{>0}} V_{y_j}$. If $Z_j \subseteq V_{y_j}$ denotes the set of points where $D_1\hat{f}$ does not exist in V_{y_j}, then the set of points in $\phi(U)$ where $D_1\hat{f}$ does not exist is $Z = \cup_{j\in\mathbb{Z}_{>0}} Z_j$. This set has measure zero, being a countable union of sets of measure zero (this is a result of countable-subadditivity of measure). Thus we can finally conclude that $Df(x;v)$ exists for almost every $x \in U$.

Now we show that $Df(x;v) = \hat{D}f(x) \cdot v$ for almost every $x \in U$. First of all, note that for almost every $x \in U$, $\hat{D}f(x)$ and $Df(x;v)$ exist. Therefore, if $v = \mathbf{0}$ the desired equality holds trivially. Thus we let $v \neq 0$. Following the proof of Lemma 2 we can write

$$U = (x_{01} - \epsilon, x_{01} + \epsilon) \times \cdots \times (x_{0n} - \epsilon, x_{0n} + \epsilon)$$

for $x_{01}, \dots, x_{0n} \in \mathbb{R}$. Then, for $x \in U$, define

$$\rho_j(x) = \begin{cases} \frac{\epsilon - |x_j - x_{0j}|}{|v_j|}, & v_j \neq 0, \\ 0, & v_j = 0. \end{cases}$$

One can verify that

$$(x_1 + \rho_1(\boldsymbol{x})v_1, \dots, x_n + \rho_n(\boldsymbol{x})v_n) \in U$$

for every $\boldsymbol{x} \in U$. Now define

$$\rho(\boldsymbol{x}) = \min\{\rho_j(\boldsymbol{x}) \mid \rho_j(\boldsymbol{x}) \neq 0\},$$

noting that $\rho(\boldsymbol{x}) \in \mathbb{R}_{>0}$ for every $\boldsymbol{x} \in U$. One can then see that $\boldsymbol{x} + \rho(\boldsymbol{x})\boldsymbol{v} \in U$ for every $\boldsymbol{x} \in U$. Moreover, if $\boldsymbol{D}f(\boldsymbol{x};\boldsymbol{v})$ exists then

$$\boldsymbol{D}f(\boldsymbol{x};\boldsymbol{v}) = \lim_{s\to 0} \frac{f(\boldsymbol{x} + s\rho(\boldsymbol{x})\boldsymbol{v}) - f(\boldsymbol{x})}{s\rho(\boldsymbol{x})}.$$

Note that

$$\Big|\frac{f(\boldsymbol{x} + s\rho(\boldsymbol{x})\boldsymbol{v}) - f(\boldsymbol{x})}{s\rho(\boldsymbol{x})}\phi(\boldsymbol{x})\Big| \leq L\|\boldsymbol{v}\|_{\mathbb{R}^n}\|\phi\|_\infty,$$

where $\|\cdot\|_\infty$ denotes the ∞-norm as described in . Thus the family of functions what?

$$\boldsymbol{x} \mapsto \frac{f(\boldsymbol{x} + s\rho(\boldsymbol{x})\boldsymbol{v}) - f(\boldsymbol{x})}{s\rho(\boldsymbol{x})}\phi(\boldsymbol{x}), \qquad s \in [-1,1],$$

defined on U is uniformly bounded. Now we use the Dominated Convergence Theorem, Theorem III-2.7.28, the change of variable theorem, , Lemma 1, and the fact that ϕ has compact support to compute what

rho-1rho needs to be worked out

$$\begin{aligned}
\int_U \boldsymbol{D}f(\boldsymbol{x};\boldsymbol{v})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} &= \int_U \lim_{s\to 0}\frac{f(\boldsymbol{x} + s\rho(\boldsymbol{x})\boldsymbol{v}) - f(\boldsymbol{x})}{s\rho(\boldsymbol{x})}\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&= \lim_{s\to 0}\frac{1}{s}\Big(\int_U \rho^{-1}(\boldsymbol{x})f(\boldsymbol{x} + s\boldsymbol{v})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_U \rho^{-1}(\boldsymbol{x})f(\boldsymbol{x})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\Big)\\
&= \lim_{s\to 0}\frac{1}{s}\Big(\int_U \rho^{-1}(\boldsymbol{x})f(\boldsymbol{x} + s\boldsymbol{v})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \int_U \rho^{-1}(\boldsymbol{x})f(\boldsymbol{x})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\Big)\\
&= \lim_{s\to 0}\frac{1}{s}\Big(\int_U \rho^{-1}(\boldsymbol{y} - s\boldsymbol{v})f(\boldsymbol{y})\phi(\boldsymbol{y} - s\boldsymbol{v})\,\mathrm{d}\boldsymbol{y} - \int_U \rho^{-1}(\boldsymbol{x})f(\boldsymbol{x})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\Big)\\
&= \int_U \lim_{s\to 0}\frac{\rho^{-1}(\boldsymbol{x} - s\boldsymbol{v})\rho(\boldsymbol{x})\phi(\boldsymbol{x} - s\boldsymbol{v}) - \phi(\boldsymbol{x})}{s\rho(\boldsymbol{x})}f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&= \int_U \boldsymbol{D}\phi(\boldsymbol{x};-\boldsymbol{v})f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_U (\boldsymbol{D}\phi(\boldsymbol{x})\cdot(-\boldsymbol{v}))f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&= -\sum_{j=1}^n v_j\int_U \boldsymbol{D}_j\phi(\boldsymbol{x})f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \sum_{j=1}^n v_j\int_U \boldsymbol{D}_jf(\boldsymbol{x})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\\
&= \int_U (\hat{\boldsymbol{D}}f(\boldsymbol{x})\cdot\boldsymbol{v})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}.
\end{aligned}$$

Since the equality

$$\int_U \boldsymbol{D}f(\boldsymbol{x};\boldsymbol{v})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_U (\hat{\boldsymbol{D}}f(\boldsymbol{x})\cdot\boldsymbol{v})\phi(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$$

holds for every infinitely differentiable function ϕ with compact support, it follows from that what?

$$\boldsymbol{D}f(\boldsymbol{x};\boldsymbol{v}) = \hat{\boldsymbol{D}}f(\boldsymbol{x})\cdot\boldsymbol{v}$$

for almost every $\boldsymbol{x}\in U$, as desired. ▼

Now we can prove the theorem for $m=1$ in the case when U is a cube.

4 Lemma *If* $U \subseteq \mathbb{R}^n$ *is an open cube of radius* ϵ *and if* $f\colon U\to\mathbb{R}$ *is Lipschitz with Lipschitz constant* L*, then*

(i) f *is differentiable almost everywhere and*

(ii) for each $j\in\{1,\dots,n\}$, $|\mathbf{D}_j f(\mathbf{x})|\leq L$ *at points* $\mathbf{x}$ *where* f *is differentiable.*

Proof Since U is a cube of radius ϵ we write

$$U = (x_{01}-\epsilon, x_{01}+\epsilon)\times\cdots\times(x_{0n}-\epsilon, x_{0n}+\epsilon)$$

for $x_{01},\dots,x_{0n}\in\mathbb{R}$. For $j\in\{1,\dots,n\}$ and $\boldsymbol{x}\in U$ define

$$\rho_j(\boldsymbol{x}) = \epsilon - |x_j - x_{0j}|.$$

Then

$$x_j + s\rho_j(\boldsymbol{x}) \in (x_{0j}-\epsilon, x_{0j}+\epsilon)$$

for each $\boldsymbol{x}\in U$, $j\in\{1,\dots,n\}$, and $s\in[-1,1]$. Also define

$$\rho(\boldsymbol{x}) = \inf\{\rho_j(\boldsymbol{x}) \mid j\in\{1,\dots,n\}\}.$$

Note that $\rho(\boldsymbol{x})\in\mathbb{R}_{>0}$ and that

$$\overline{\mathsf{B}}(\rho(\boldsymbol{x}),\boldsymbol{x})\subseteq U.$$

Let $A_0\subseteq U$ be the set of points $\boldsymbol{x}\in U$ for which $\hat{\boldsymbol{D}}f(\boldsymbol{x})$ exists. Let

$$\mathbb{S}^{n-1} = \{\boldsymbol{v}\in\mathbb{R}^n \mid \|\boldsymbol{v}\|_{\mathbb{R}^n} = 1\}.$$

Define $F\colon A_0\times\mathbb{S}^{n-1}\times([-1,1]\setminus\{0\})\to\mathbb{R}$ by

$$F(\boldsymbol{x},\boldsymbol{u},s) = \frac{f(\boldsymbol{x}+s\rho(\boldsymbol{x})\boldsymbol{u}) - f(\boldsymbol{x})}{s\rho(\boldsymbol{x})} - \hat{\boldsymbol{D}}f(\boldsymbol{x})\cdot\boldsymbol{u}.$$

Note that if $\boldsymbol{u}_1,\boldsymbol{u}_2\in\mathbb{S}^{n-1}$ then

$$\begin{aligned}|F(\boldsymbol{x},\boldsymbol{u}_1,s)-F(\boldsymbol{x},\boldsymbol{u}_2,t)| &\leq \left|\frac{f(\boldsymbol{x}+s\rho(\boldsymbol{x})\boldsymbol{u}_1) - f(\boldsymbol{x}+s\rho(\boldsymbol{x})\boldsymbol{u}_2)}{s\rho(\boldsymbol{x})}\right| + |\hat{\boldsymbol{D}}f(\boldsymbol{x})\cdot(\boldsymbol{u}_1-\boldsymbol{u}_2)| \\ &\leq (L + \|\hat{\boldsymbol{D}}f(\boldsymbol{x})\|_{\mathbb{R}^n,\mathbb{R}})\|\boldsymbol{u}_1-\boldsymbol{u}_2\|_{\mathbb{R}^n} \\ &\leq L(1+\sqrt{n})\|\boldsymbol{u}_1-\boldsymbol{u}_2\|_{\mathbb{R}^n},\end{aligned}$$

after noting from Theorem 1.1.14 and Lemma 2 that

$$\|\hat{\boldsymbol{D}}f(\boldsymbol{x})\|_{\mathbb{R}^n,\mathbb{R}} = \Big(\sum_{j=1}^n |\boldsymbol{D}_j f(\boldsymbol{x})|^2\Big)^{1/2} \leq L\sqrt{n}.$$

what?

Now, by , let $(\boldsymbol{u}_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathbb{S}^{n-1}$ such that

$$\mathrm{cl}(\{\boldsymbol{u}_j \mid j \in \mathbb{Z}_{>0}\}) = \mathbb{S}^{n-1}.$$

For $j \in \mathbb{Z}_{>0}$ define

$$A_j = \{\boldsymbol{x} \in A_0 \mid \boldsymbol{D}f(\boldsymbol{x};\boldsymbol{u}_j) \text{ exists and } \boldsymbol{D}f(\boldsymbol{x};\boldsymbol{u}_j) = \hat{\boldsymbol{D}}f(\boldsymbol{x})\cdot\boldsymbol{u}_j\},$$

noting from Lemma 3 that $U \setminus A_j$ has measure zero. Thus, if we define $A = \cap_{j\in\mathbb{Z}_{>0}} A_j$, then $U \setminus A$ has measure zero since the countable union of sets of measure zero has measure zero.

Let $\boldsymbol{x} \in A$ and let $\epsilon \in \mathbb{R}_{>0}$. Abbreviate $\epsilon' = \frac{\epsilon}{2L(1+\|f\|_{\mathbb{R}^n,\mathbb{R}})}$. Note that $(\mathsf{B}(\epsilon',\boldsymbol{u}_j))_{j\in\mathbb{Z}_{>0}}$ is an open cover of $\mathbb{S}^{n-1}$. Thus, by the Heine–Borel Theorem, let $\boldsymbol{u}_{j_1},\dots,\boldsymbol{u}_{j_k}$ be such that $\mathbb{S}^{n-1} \subseteq \cup_{l=1}^k \mathsf{B}(\epsilon',\boldsymbol{u}_{j_l})$. Since $\boldsymbol{D}f(\boldsymbol{x};\boldsymbol{u}_{j_l})$ exists and is equal to $\hat{\boldsymbol{D}}f(\boldsymbol{x})\cdot\boldsymbol{u}_{j_l}$ for each $l \in \{1,\dots,k\}$, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $|s| \in (0,\delta]$, then $|F(\boldsymbol{x},\boldsymbol{u}_{j_l},s)| \le \frac{\epsilon}{2}$ for each $l \in \{1,\dots,k\}$. Now, if $\boldsymbol{u} \in \mathbb{S}^{n-1}$, there exists $l \in \{1,\dots,k\}$ such that

$$L(1+\sqrt{n})\|\boldsymbol{u}-\boldsymbol{u}_{j_l}\|_{\mathbb{R}^n} < \frac{\epsilon}{2}.$$

Therefore, for $|s| \in (0,\delta)$,

$$\begin{aligned}|F(\boldsymbol{x},\boldsymbol{u},s)| &\le |F(\boldsymbol{x},\boldsymbol{u},s) - F(\boldsymbol{x},\boldsymbol{u}_{j_l})| + |F(\boldsymbol{x},\boldsymbol{u}_{j_l},s)|\\ &\le \frac{\epsilon}{2} + L(1+\sqrt{n})\|\boldsymbol{u}-\boldsymbol{u}_{j_l}\|_{\mathbb{R}^n} < \epsilon.\end{aligned}$$

From this we conclude that f is differentiable at $\boldsymbol{x}$ and that $\boldsymbol{D}f(\boldsymbol{x}) = \hat{\boldsymbol{D}}f(\boldsymbol{x})$. Since this holds for every $\boldsymbol{x} \in A$ and since $U \setminus A$ has measure zero, the first part of the lemma holds.

The second assertion of the lemma follows from follows from Lemma 2 since, if f is differentiable at $\boldsymbol{x}$ then the partial derivatives $\boldsymbol{D}_j f(\boldsymbol{x})$ exist. ▼

Next we can prove the theorem for arbitrary m, but still assuming that U is a cube.

5 Lemma *If* $\mathrm{U} \subseteq \mathbb{R}^{\mathrm{n}}$ *is an open cube of radius* ϵ *and if* $\mathbf{f}\colon \mathrm{U} \to \mathbb{R}^{\mathrm{m}}$ *is Lipschitz with Lipschitz constant* L*, then*

(i) $\mathbf{f}$ *is differentiable almost everywhere and*

(ii) $\|\mathbf{Df}(\mathbf{x})\|_{\mathbb{R}^{\mathrm{n}},\mathbb{R}^{\mathrm{m}}} \le \mathrm{L}$ *at points* $\mathbf{x}$ *where* $\mathbf{f}$ *is differentiable.*

Proof For $a \in \{1,\dots,m\}$ let $V_a \subseteq U$ be the set of points $\boldsymbol{x} \in U$ such that f_a is differentiable at $\boldsymbol{x}$. By Lemma 4, $U \setminus V_a$ has measure zero. If $V = \cup_{a=1}^m V_a$ then $U \setminus V$ has measure zero. By 1.4.17 we conclude that $\boldsymbol{f}$ is differentiable at all points in V.

Let $\boldsymbol{u} \in \mathbb{R}^n$ be such that $\|\boldsymbol{u}\|_{\mathbb{R}^n} = 1$ and suppose that $\boldsymbol{f}$ is differentiable at $\boldsymbol{x}$. Then, by definition of the derivative,

$$s\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x})\cdot\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{x}+s\boldsymbol{u}) - \boldsymbol{f}(\boldsymbol{x}) + o(s).$$

Therefore,

$$\|\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x})\cdot\boldsymbol{u}\|_{\mathbb{R}^m} \le \frac{\|\boldsymbol{f}(\boldsymbol{x}+s\boldsymbol{u}) - \boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^m}}{s\|\boldsymbol{u}\|_{\mathbb{R}^n}} + \frac{|o(s)|}{|s|} \le L + \frac{|o(s)|}{|s|}.$$

Since $\lim_{s\to 0} \frac{|o(s)|}{|s|} = 0$, we have

$$\|\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x}) \cdot \boldsymbol{u}\|_{\mathbb{R}^m} \leq L$$

for every $\boldsymbol{u} \in \mathbb{R}^n$ such that $\|\boldsymbol{u}\|_{\mathbb{R}^n} = 1$. Referring to the first few lines of the proof of Theorem 1.1.14, this gives $\|\boldsymbol{D}\boldsymbol{f}(\boldsymbol{x})\|_{\mathbb{R}^n,\mathbb{R}^m} \leq L$, as desired. ▼

Finally, we complete the proof of the theorem in the case that U is a general open set. Let $\boldsymbol{x} \in U$ and let $U_{\boldsymbol{x}} \subseteq U$ be an open cube for which $\boldsymbol{x} \in U_{\boldsymbol{x}}$. By Lemma 5, $\boldsymbol{f}$ is differentiable almost everywhere in $U_{\boldsymbol{x}}$. Note that $(U_{\boldsymbol{x}})_{\boldsymbol{x}\in U}$ is an open cover for U. By Lemma 1.2.33 there exists $\boldsymbol{x}_j \in U$, $j \in \mathbb{Z}_{>0}$, such that $U = \cup_{j\in\mathbb{Z}_{>0}} U_{\boldsymbol{x}_j}$. Let $Z_j \subseteq U_{\boldsymbol{x}_j}$ be the set of points in $U_{\boldsymbol{x}_j}$ at which $\boldsymbol{f}$ is differentiable. Then the set of points in U for which $\boldsymbol{f}$ is not differentiable is $Z = \cup_{j\in\mathbb{Z}_{>0}} Z_j$. Since Z has measure zero, the first conclusion of the theorem applies.

Next, suppose that $\boldsymbol{f}$ is differentiable at $\boldsymbol{x} \in U$, and let $V \subseteq U$ be a neighbourhood of $\boldsymbol{x}$ and $L_{\boldsymbol{x}} \in \mathbb{R}_{>0}$ be such that $\boldsymbol{f}|V$ has Lipschitz constant $L_{\boldsymbol{x}}$. Let $U_{\boldsymbol{x}} \subseteq V$ be an open cube containing $\boldsymbol{x}$. By Lemma 5 we have $|\boldsymbol{D}_j f_a(\boldsymbol{x})| \leq L_{\boldsymbol{x}}$. ■

1.10.9 Sard's Theorem

1.10.10 Notes

The Hausdorff metric was presented by Felix Hausdorff in 1914 and is described in his book [Hausdorff 1937]. The basic properties of the Hausdorff metric are described in many places; we refer to [Searcóid 2007] as a specific instance. A place where Hausdorff distance is important is in the theory of so-called "fractal" sets. For a development of this subject we refer to the book of Edgar [2007].

The uses of "hemicontinuity" and "semicontinuity" we adopt for set-valued maps is that of Klein and Thompson [1984].

The Maximum Theorem we state as Theorem 1.10.41 is stated in some similar form by Berge [1959].

The very important Theorem 1.10.51 is really a result in either dimension theory or algebraic topology. In each case, once one understands the basic of these subjects, Theorem 1.10.51 comes out as a somewhat natural conclusion. However, such matters are a little beyond even our scope here, and we refer to the classical text of Hurewicz and Wallman [1941] for a presentation of dimension theory and to, for example, Munkres [1984] for some background in algebraic topology. The elementary proof of Theorem 1.10.51 we give is due to [Kulpa and Tursański 1988].

Theorem 1.10.60 was first presented by Rademacher [1919]. The proof we give follows [Ziemer 1989].

Exercises

1.10.1 Let $A = B = \mathbb{R} \subseteq \mathbb{R}$ and for each of the functions $f\colon A \to B$ given below, determine whether f is continuous, open, closed:

(a) $f(x) = x^2$;

(b) $f(x) = \sin(x)$;

(c) $f(x) = e^x$.

1.10.2 Let $A \subseteq \mathbb{R}^n$ and let $\boldsymbol{F}: A \twoheadrightarrow \mathbb{R}^m$ be a set-valued map. Show that $\boldsymbol{F}$ is lower semicontinuous at $\boldsymbol{x}_0$ if and only if, for every sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$ in A converging to $\boldsymbol{x}_0$ and for every $\boldsymbol{y}_0 \in \boldsymbol{F}(\boldsymbol{x}_0)$, there exists a sequence $(\boldsymbol{y}_j)_{j\in\mathbb{Z}_{>0}}$ in $\mathbb{R}^m$ such that $\boldsymbol{y}_j \in \boldsymbol{F}(\boldsymbol{x}_j)$ and such that $\lim_{j\to\infty} \boldsymbol{y}_j = \boldsymbol{y}_0$.

1.10.3 Let $A \subseteq \mathbb{R}^n$ and for $\boldsymbol{f}: A \to \mathbb{R}^m$ define $\boldsymbol{F}_{\boldsymbol{f}}: A \twoheadrightarrow \mathbb{R}^m$ by $\boldsymbol{F}_{\boldsymbol{f}}(\boldsymbol{x}) = \{\boldsymbol{f}(\boldsymbol{x})\}$. Show that for $\boldsymbol{x}_0 \in A$ the following three statements are equivalent:

1. $\boldsymbol{f}$ is continuous at $\boldsymbol{x}_0$;
2. $\boldsymbol{F}_{\boldsymbol{f}}$ is upper semicontinuous at $\boldsymbol{x}_0$;
3. $\boldsymbol{F}_{\boldsymbol{f}}$ is lower semicontinuous at $\boldsymbol{x}_0$.

1.10.4 Let $A \subseteq \mathbb{R}^n$, let $f_-, f_+: A \to \mathbb{R}$ satisfy $f_-(\boldsymbol{x}) \leq f_+(\boldsymbol{x})$ for each $\boldsymbol{x} \in A$. Define $F: A \twoheadrightarrow \mathbb{R}$ by $F(\boldsymbol{x}) = [f_-(\boldsymbol{x}), f_+(\boldsymbol{x})]$.

(a) Show that if f_- is lower semicontinuous and f_+ is upper semicontinuous at $\boldsymbol{x}_0$ then F is upper semicontinuous at $\boldsymbol{x}_0$.

(b) Show that if f_- is upper semicontinuous and f_+ is lower semicontinuous at $\boldsymbol{x}_0$ then F is upper semicontinuous at $\boldsymbol{x}_0$.

1.10.5 Show that the map $f: \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x^2$ is locally Lipschitz but not Lipschitz.

1.10.6 Let $I, J \subseteq \mathbb{R}$ be a intervals, and let $\phi: I \to J$ be locally Lipschitz and let $f: J \to \mathbb{R}$ be locally absolutely continuous. Show that $f \circ \phi$ is locally absolutely continuous.

Section 1.11

Fixed point theorems

In this section we present a few well-known fixed point theorems as a means of illustrating what can be achieved using the tools we have developed thus far. Such fixed point theorems are often very important in applications. For example, (1) the Contraction Mapping Theorem is used in the proof of the existence and uniqueness theorem for solutions of ordinary differential equations, (2) the Brouwer Fixed Point Theorem is used crucially in the proof of the Pontryagin Maximum Principle in optimal control theory, and (3) Nash used both the Brouwer and Kakutani Fixed Point Theorems in his proof of the existence of (now) so-called "Nash equilibria" in non-cooperative game theory. We shall say a few things about game theory in Section 1.12.

Do I need to read this section? The Contraction Mapping Theorem has already been used in the proof of the change of variables formula (in the proof of Lemma 2), so it is fair game. However, the other results in this section can be overlooked until needed, or until the reader has developed sufficient interest. •

1.11.1 The Contraction Mapping Theorem in $\mathbb{R}^n$

Before we get started stating our fixed point theorems, perhaps we should say what a fixed point is.

1.11.1 Definition (Fixed point) If S is a set and if $f: S \to S$ is a map, a ***fixed point*** for f is a point $x_0 \in S$ such that $f(x_0) = x_0$. •

Let us give some examples of fixed points.

1.11.2 Examples

1. Of course, every element of S is a fixed point for the identity map id_S.
2. If V is a vector space over a field F and if $\mathsf{L} \in \mathrm{End}_{\mathsf{F}}(\mathsf{V})$, a fixed point for L is, by definition, simply an element of the eigenspace for the eigenvalue 1_{F}. Thus L has nonzero fixed points if and only if L has 1_{F} as an eigenvalue.
3. Another way of understanding the definition of a fixed point is as follows. Let

$$\Delta_A = \{(x, x) \mid x \in S\} \subseteq S \times S$$

be the diagonal in $S \times S$. A fixed point is that a point $x_0 \in S$ such that $(x_0, f(x_0)) \in \Delta_S$. That is, fixed point occur where $\mathrm{graph}(f)$ intersects Δ_S. •

We will be interested here in fixed points of maps from a subsets of $\mathbb{R}^n$ to themselves. Perhaps the simplest such result is the following.

1.11.3 Theorem (Contraction Mapping Theorem in $\mathbb{R}^n$) *If* $A \subseteq \mathbb{R}^n$ *is closed and if* $\mathbf{f}\colon A \to A$ *has the property that there exists* $\lambda \in [0, 1)$ *such that*

$$\|\mathbf{f}(\mathbf{x}_1) - \mathbf{f}(\mathbf{x}_2)\|_{\mathbb{R}^n} \leq \lambda \|\mathbf{x}_1 - \mathbf{x}_2\|_{\mathbb{R}^n}, \qquad \mathbf{x}_1, \mathbf{x}_2 \in A,$$

then there exists a unique fixed point for $\mathbf{f}$.

Proof Let $y_0 \in A$ 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

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

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

$$\|y_k - y_l\|_{\mathbb{R}^n} \leq \|y_k - y_{k+1}\|_{\mathbb{R}^n} + \cdots + \|y_{l-1} - y_l\|_{\mathbb{R}^n} \leq (\lambda^k + \cdots + \lambda^{l-1}) \|y_0 - y_1\|_{\mathbb{R}^n}.$$

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

$$\|y_0 - y_1\|_{\mathbb{R}^n} \sum_{j=k}^{l-1} \lambda^j < \epsilon, \qquad k, l \geq N,\ l > k.$$

Then, for $k, l \geq N$ with $l > k$ we have $\|y_k - y_l\|_{\mathbb{R}^n} < \epsilon$, giving the sequence $(y_j)_{j\in\mathbb{Z}_{>0}}$ as a Cauchy sequence, as desired. By Theorem 1.2.5 there exists $x_0 \in \mathbb{R}^n$ such that $(y_j)_{j\in\mathbb{Z}_{>0}}$ converges to x_0. Moreover, by Proposition 1.2.26 we have $x_0 \in \mathrm{cl}(A) = A$. We claim that $f(x_0) = x_0$. For $\epsilon \in \mathbb{R}_{>0}$ let $j \in \mathbb{Z}_{>0}$ be sufficiently large that $\|x_0 - y_j\|_{\mathbb{R}^n} < \frac{\epsilon}{2(1+\lambda)}$ and such that $\lambda^j \|y_0 - y_1\|_{\mathbb{R}^n} < \frac{\epsilon}{2}$. Then

$$\begin{aligned}
\|x_0 - f(x_0)\|_{\mathbb{R}^n} &\leq \|x_0 - y_j\|_{\mathbb{R}^n} + \|y_j - f(y_j)\|_{\mathbb{R}^n} + \|f(y_j) - f(x_0)\|_{\mathbb{R}^n} \\
&\leq (1 + \lambda)\|x_0 - y_j\|_{\mathbb{R}^n} + \lambda^j \|y_0 - y_1\|_{\mathbb{R}^n} < \epsilon.
\end{aligned}$$

Thus $\|x_0 - f(x_0)\|_{\mathbb{R}^n} = 0$ 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

$$\|x_0 - \tilde{x}_0\|_{\mathbb{R}^n} = \|f(x_0) - f(\tilde{x}_0)\|_{\mathbb{R}^n} \leq \lambda \|x_0 - \tilde{x}_0\|_{\mathbb{R}^n} < \|x_0 - \tilde{x}_0\|_{\mathbb{R}^n}.$$

Therefore, $\|x_0 - \tilde{x}_0\|_{\mathbb{R}^n} = 0$ and so $x_0 = \tilde{x}_0$. ■

The Contraction Mapping Theorem holds in a more general setup than we give here, and we state a version of this as Theorem III-1.1.23.

1.11.2 The Hairy Ball Theorem

In the next section we shall state and prove the Brouwer Fixed Point Theorem. Our proof is different than many proofs which have an essential topological flavour. Our proof instead has an analytical flavour, and relies on an initial proof of the independently interesting Hairy Ball Theorem. We recall that

$$\mathbb{S}^n = \{x \in \mathbb{R}^{n+1} \mid \|x\|_{\mathbb{R}^{n+1}} = 1\}$$

denotes the n-dimensional sphere. We will be interested in vector fields on $\mathbb{S}^n$. A vector field on $\mathbb{S}^n$ assigns to $x \in \mathbb{S}^n$ a vector in $\mathbb{R}^{n+1}$ that is tangent to the sphere at x. Note that the set of vectors tangent to $\mathbb{S}^n$ at x is the set of vectors in $\mathbb{R}^{n+1}$ orthogonal to x using the standard inner product. Thus we shall think of a vector field on $\mathbb{S}^n$ as being a map $f\colon \mathbb{S}^n \to \mathbb{R}^{n+1}$ such that $\langle f(x), x\rangle_{\mathbb{R}^{n+1}} = 0$ for all $x \in \mathbb{S}^n$.

The question we are interested in is this: Is there a vector field on $\mathbb{S}^n$ that is nowhere zero?[14] First let us consider the case when n is odd.

1.11.4 Proposition (Odd-dimension spheres possess nowhere vanishing vector fields) *If $n \in \mathbb{Z}_{>0}$ is odd then there exists an infinitely differentiable vector field $\mathbf{f}\colon \mathbb{S}^n \to \mathbb{R}^{n+1}$ such that $\mathbf{f}(\mathbf{x}) \neq \mathbf{0}$ for every $\mathbf{x} \in \mathbb{S}^n$.*

Proof Define f by

$$f(x_1, \dots, x_{n+1}) = (x_2, -x_1, x_4, -x_3, \dots, x_n, -x_{n+1}),$$

and check that $\langle f(x), x\rangle_{\mathbb{R}^{n+1}} = 0$ and $\|f(x)\|_{\mathbb{R}^{n+1}}$. ■

For even n we have the following result.

1.11.5 Theorem (Hairy Ball Theorem) *Let $n \in \mathbb{Z}_{>0}$ be even. If $\mathbf{f}\colon \mathbb{S}^n \to \mathbb{R}^{n+1}$ is a continuous vector field on $\mathbb{S}^n$ then there exists $\mathbf{x}_0 \in \mathbb{S}^n$ such that $\mathbf{f}(\mathbf{x}_0) = \mathbf{0}$.*

Proof We first prove the result supposing that f is not only continuous but of class C^1.

We use a series of lemmata to prove the theorem in this case.

1 Lemma *Let $A \subseteq \mathbb{R}^{n+1}$ be compact, let U be a neighbourhood of A, let $\mathbf{g}\colon U \to \mathbb{R}^{n+1}$ be of class C^1, and for $s \in \mathbb{R}$ define $\mathbf{h}_s\colon A \to \mathbb{R}^{n+1}$ by $\mathbf{h}_s(\mathbf{x}) = \mathbf{x} + s\mathbf{g}(\mathbf{x})$. Then there exists $\epsilon \in \mathbb{R}_{>0}$ such that*

(i) for each $s \in [-\epsilon, \epsilon]$, $\mathbf{h}_s$ is injective and

(ii) the function $s \mapsto \mathrm{vol}(\mathbf{h}_s(A))$ is a polynomial.

Proof By Theorem 1.10.58 let $M \in \mathbb{R}_{>0}$ be such that

$$\|g(y) - g(x)\|_{\mathbb{R}^{n+1}} \leq M\|y - x\|_{\mathbb{R}^{n+1}}, \qquad x, y \in A,$$

[14] The main reason this question is interesting is somewhat beyond the scope of our presentation here. If there is no vector field on $\mathbb{S}^n$ that is nowhere zero this means that the tangent bundle to $\mathbb{S}^n$ is not trivialisable. This is an essentially interesting fact about the topology of spheres.

and let $\epsilon \in (0, M^{-1})$. Then for $|s| < \epsilon$ we claim that $\boldsymbol{h}_s$ is injective. Indeed, if $\boldsymbol{h}_s(\boldsymbol{x}) = \boldsymbol{h}_s(\boldsymbol{y})$ then

$$\begin{aligned} & \boldsymbol{x} - \boldsymbol{y} = s(\boldsymbol{g}(\boldsymbol{y}) - \boldsymbol{g}(\boldsymbol{x})) \\ \implies \quad & \|\boldsymbol{y} - \boldsymbol{x}\|_{\mathbb{R}^{n+1}} = |s| \|\boldsymbol{g}(\boldsymbol{y}) - \boldsymbol{g}(\boldsymbol{x})\|_{\mathbb{R}^{n+1}} \le |s| M \|\boldsymbol{x} - \boldsymbol{y}\|_{\mathbb{R}^{n+1}}. \end{aligned}$$

Since $|s|M < 1$ this implies that $\boldsymbol{x} = \boldsymbol{y}$. This gives the first assertion in the lemma.

For the second assertion we observe that $\det \boldsymbol{D}\boldsymbol{h}_s(\boldsymbol{x})$ is a polynomial function of s, it being the determinant of a matrix whose entries are linear in s. Thus we can write

$$\det \boldsymbol{D}\boldsymbol{h}_s(\boldsymbol{x}) = 1 + a_1(\boldsymbol{x})s + \cdots + a_{n+1}(\boldsymbol{x})s^{n+1}$$

for continuous functions $a_1, \ldots, a_{n+1}$. Therefore, using the change of variables formula,

$$\operatorname{vol}(\boldsymbol{h}_s(A)) = \int_{\boldsymbol{h}_s(A)} \mathrm{d}x_1 \cdots \mathrm{d}x_{n+1} = \int_A \det \boldsymbol{D}\boldsymbol{h}_s(\boldsymbol{x})\, \mathrm{d}x_1 \cdots \mathrm{d}x_{n+1},$$

which is clearly a polynomial function in s. ▼

2 Lemma *Let* $\mathbf{f}\colon \mathbb{S}^n \to \mathbb{R}^{n+1}$ *have the following properties:*

(i) $\langle \mathbf{f}(\mathbf{x}), \mathbf{x} \rangle_{\mathbb{R}^{n+1}} = 0$ *for each* $\mathbf{x} \in \mathbb{S}^n$*;*

(ii) $\|\mathbf{f}(\mathbf{x})\|_{\mathbb{R}^{n+1}} = 1$ *for each* $\mathbf{x} \in \mathbb{S}^n$*.*

Let U *be a neighbourhood of* $\mathbb{S}^n$ *with* $\bar{\mathbf{f}}\colon \mathrm{U} \to \mathbb{R}^{n+1}$ *a continuously differentiable extension of* $\mathbf{f}$*, and for* $\mathrm{s} \in \mathbb{R}$ *define* $\mathbf{h}_\mathrm{s}\colon \mathrm{U} \to \mathbb{R}^{n+1}$ *by* $\mathbf{h}_\mathrm{s}(\mathbf{x}) = \mathbf{x} + \mathrm{s}\bar{\mathbf{f}}(\mathbf{x})$*. Then, for* $|\mathrm{s}|$ *sufficiently small,* $\bar{\mathbf{f}}$ *maps* $\mathbb{S}^n$ *onto the sphere*

$$\mathbb{S}^n(\sqrt{1 + \mathrm{s}^1}) = \left\{\mathbf{x} \in \mathbb{R}^{n+1} \;\middle|\; \|\mathbf{x}\|_{\mathbb{R}^{n+1}} = \sqrt{1 + \mathrm{s}^2}\right\}$$

of radius $\sqrt{1 + \mathrm{s}^2}$*.*

Proof First note that $\boldsymbol{h}_s(\mathbb{S}^n) \subseteq \mathbb{S}^n(\sqrt{1 + s^2})$ for any $s \in \mathbb{R}$ by direct computation. As we saw in the proof of Lemma 1, for s sufficiently small $\boldsymbol{D}\boldsymbol{h}_s(\boldsymbol{x})$ is nonsingular for each $\boldsymbol{x} \in \mathbb{S}^n$. By the Inverse Function Theorem this means that, for s sufficiently small, $\boldsymbol{h}_s$ is a local diffeomorphism about every point in $\mathbb{S}^n$. Thus $\boldsymbol{h}_s|\mathbb{S}^n$ maps every sufficiently small open set to an open set, provided that s is sufficiently small. This in turn means that $\boldsymbol{h}_s|\mathbb{S}^n$ is an open mapping for s sufficiently small. In particular, $\boldsymbol{h}_s(\mathbb{S}^n)$ is an open subset of $\mathbb{S}^n(\sqrt{1 + s^2})$ for s sufficiently small. However, $\boldsymbol{h}_s(\mathbb{S}^n)$ is also compact, the image of compact sets under continuous maps being compact. The only subset of $\mathbb{S}^n(\sqrt{1 + s^2})$ that is open and closed is $\mathbb{S}^n(\sqrt{1 + s^2})$ since $\mathbb{S}^n(\sqrt{1 + s^2})$ is connected. ▼

Now suppose that $f\colon \mathbb{S}^n \to \mathbb{R}^{n+1}$ is such that

1. f is of class C^1,
2. $\langle f(\boldsymbol{x}), \boldsymbol{x} \rangle_{\mathbb{R}^{n+1}} = 0$ for $\boldsymbol{x} \in \mathbb{S}^n$, and
3. $f(\boldsymbol{x}) \neq 0$ for every $\boldsymbol{x} \in \mathbb{S}^n$.

We may assume without loss of generality (by dividing f by the function $x \mapsto \|f(x)\|_{\mathbb{R}^{n+1}}$) that $\|f(x)\|_{\mathbb{R}^{n+1}} = 1$ for each $x \in \mathbb{S}^n$. For $a, b \in \mathbb{R}_{>0}$ satisfying $a < 1 < b$ define

$$A = \{x \in \mathbb{R}^{n+1} \mid a \le \|x\|_{\mathbb{R}^{n+1}} \le b\},$$

and note that A is compact. For the function f as in the theorem statement (but now of class C^1) extend f to A by $f(rx) = rf(x)$, $x \in \mathbb{S}^n$. Then, with $h_s(x) = x + sf(x)$ for $s \in \mathbb{R}$, $h_s(rx) = rh_s(x)$ for $x \in \mathbb{S}^n$. Therefore, h_s maps the sphere of radius $r \in [a, b]$ into the sphere of radius $r\sqrt{1+s^2}$. By Lemma 2, for s sufficiently small h_s maps the sphere of radius r onto the sphere of radius $r\sqrt{1+s^2}$ for each $r \in [a, b]$. Therefore,

$$\text{vol}(h_s(A)) = (\sqrt{1+s^2})^{n+1}\text{vol}(A)$$

for s sufficiently small. For n even this is not a polynomial in s, and this contradicts Lemma 1. This proves the theorem for f of class C^1.

Finally, we prove the theorem for f continuous. Thus we let $f\colon \mathbb{S}^n \to \mathbb{R}^{n+1}$ have the following properties:

1. f is continuous;
2. $\langle f(x), x\rangle_{\mathbb{R}^{n+1}} = 0$ for each $x \in \mathbb{S}^n$;
3. $\|f(x)\|_{\mathbb{R}^{n+1}} \neq 0$ for $x \in \mathbb{S}^n$.

Let

$$m = \inf\{\|f(x)\|_{\mathbb{R}^{n+1}} \mid x \in \mathbb{S}^n\}$$

and let $p\colon \mathbb{S}^n \to \mathbb{R}^{n+1}$ be a polynomial function such that

$$\sup\{\|p(x) - f(x)\|_{\mathbb{R}^{n+1}} \mid x \in \mathbb{S}^n\} < \tfrac{m}{2},$$

this being possible by the Weierstrass Approximation Theorem (precisely, by Corollary 1.7.5). Now define a continuously differentiable function $g\colon \mathbb{S}^n \to \mathbb{R}^{n+1}$ by

$$g(x) = p(x) - \langle p(x), x\rangle_{\mathbb{R}^{n+1}} x,$$

and note that $\langle g(x), x\rangle_{\mathbb{R}^{n+1}} = 0$ by direct computation. We have

$$\begin{aligned} & \langle p(x) - f(x), x\rangle_{\mathbb{R}^{n+1}} = \langle p(x), x\rangle_{\mathbb{R}^{n+1}} \\ \implies \quad & |\langle p(x), x\rangle_{\mathbb{R}^{n+1}}| = |\langle p(x) - f(x), x\rangle_{\mathbb{R}^{n+1}}| \le \|p(x) - f(x)\|_{\mathbb{R}^{n+1}} < \tfrac{m}{2} \end{aligned}$$

for each $x \in \mathbb{S}^n$. This gives

$$\|g(x) - p(x)\|_{\mathbb{R}^{n+1}} = |\langle p(x), x\rangle_{\mathbb{R}^{n+1}}| < \tfrac{m}{2},$$

and so

$$\begin{aligned} |\|g(x)\|_{\mathbb{R}^{n+1}} - \|f(x)\|_{\mathbb{R}^{n+1}}| &\le \|g(x) - f(x)\|_{\mathbb{R}^{n+1}} \\ &\le \|g(x) - p(x)\|_{\mathbb{R}^{n+1}} + \|p(x) - f(x)\|_{\mathbb{R}^{n+1}} < m \end{aligned}$$

for all $x \in \mathbb{S}^n$. This implies that $\|g(x)\|_{\mathbb{R}^{n+1}} > 0$ for all $x \in \mathbb{S}^n$, which is in contradiction to what we proved in the first part of the proof since g is continuously differentiable. ■

We encourage the reader to attempt to ingest the Hairy Ball Theorem by thinking of the case when $n = 2$, since the name "Hairy Ball Theorem" is derived in this context. The idea is that a vector field on $\mathbb{S}^2$ can be thought of as assigning a "strand of hair" tangent to $\mathbb{S}^2$ at each point. The theorem says that, if there is a strand of hair of nonzero length at every point, there is no way combing the hair to lie flat at each point on $\mathbb{S}^2$.

1.11.3 The Brouwer[15] Fixed Point Theorem

Now we state and prove (using the Hairy Ball Theorem) the Brouwer Fixed Point Theorem. We define

$$\mathbb{D}^n = \{x \in \mathbb{R}^n \mid \|x\|_{\mathbb{R}^n} \leq 1\}$$

to be the closed ball of unit radius, or the so-called unit disk. This is redundant notation that is convenient in this case.

1.11.6 Theorem (Brouwer Fixed Point Theorem) *If* $\mathbf{f}\colon \mathbb{D}^{\mathrm{n}} \to \mathbb{D}^{\mathrm{n}}$ *is continuous then there exists* $\mathbf{x}_0 \in \mathbb{D}^{\mathrm{n}}$ *such that* $\mathbf{f}(\mathbf{x}_0) = \mathbf{x}_0$*.*

Proof First let us suppose that n is even. Suppose that $f(x) \neq x$ for every $x \in \mathbb{D}^n$. Then define $g\colon \mathbb{D}^n \to \mathbb{R}^n$ by

$$g(x) = x - f(x)\frac{1 - \langle x, x\rangle_{\mathbb{R}^n}}{1 - \langle f(x), x\rangle_{\mathbb{R}^n}}.$$

Note that $g(x) = x$ for $\langle x, x\rangle_{\mathbb{R}^n} = 1$, and so g points "outward" on $\mathbb{S}^{n-1} = \mathrm{bd}(\mathbb{D}^n)$. Since

$$|\langle f(x), x\rangle_{\mathbb{R}^n}| < |f(x)| \leq 1$$

it follows that g is continuous. We also claim that g is nowhere zero. If $\{f(x), x\}$ is linearly independent then $g(x)$ is clearly nonzero. If $\{f(x), x\}$ is linearly dependent then $\langle x, x\rangle_{\mathbb{R}^n} f(x) = \langle f(x), x\rangle_{\mathbb{R}^n} x$ and so

$$g(x) = \frac{x - f(x)}{1 - \langle f(x), x\rangle_{\mathbb{R}^n}} \neq 0.$$

Now consider $\mathbb{S}^n \subseteq \mathbb{R}^{n+1}$ and denote by

$$\mathbb{S}^n_- = \{x \in \mathbb{S}^n \mid x_{n+1} \leq 0\}, \quad \mathbb{S}^n_+ = \{x \in \mathbb{S}^n \mid x_{n+1} \geq 0\}$$

the southern and northern hemispheres, respectively. We also denote by $E = \mathbb{S}^n_- \cap \mathbb{S}^n_+$ the equator. Now define a map ϕ_- from $\mathbb{D}^n$ to $\mathbb{S}^n_-$ by

$$\phi_-(x) = \frac{(2x_1, \dots, 2x_n, 1 - \langle x, x\rangle_{\mathbb{R}^n})}{1 + \langle x, x\rangle_{\mathbb{R}^n}}.$$

[15] Luitzen Egbertus Jan Brouwer (1881–1966) was a Dutch mathematician who made significant contributions to topology and mathematical logic. Brouwer was notable for advocating a "constructionist" approach to mathematics. He rejected mathematics as a purely formal occupation, thinking of it rather as being an intuitive exercise. It is interesting to note that often Brouwer did not lecture on his mathematical research (mainly in topology) that was not actually done in accordance with his own mathematical philosophy. He was an interesting character.

(One may verify that this map is stereographic projection from the north pole, thinking of $\mathbb{D}^n$ as being the disk whose boundary is E.) Now define a vector field $\boldsymbol{h}$ on $\mathbb{S}^n_-$ by

$$\boldsymbol{h}(z) = \boldsymbol{D}\boldsymbol{\phi}_-(\boldsymbol{\phi}_-^{-1}(z)) \cdot \boldsymbol{g}(\boldsymbol{\phi}_-^{-1}(z)).$$

This is a nowhere zero vector field on $\mathbb{S}^n_-$. Moreover, a direct computation shows that for $\boldsymbol{z} \in E$ we have $\boldsymbol{h}(z) = (0, \dots, 0, 1)$. Define a map $\boldsymbol{\phi}_+$ from $\mathbb{D}^n$ to $\mathbb{S}^n_+$ by

$$\boldsymbol{\phi}_+(\boldsymbol{x}) = \frac{(2x_1, \dots, 2x_n, -1 + \langle \boldsymbol{x}, \boldsymbol{x}\rangle_{\mathbb{R}^n})}{1 + \langle \boldsymbol{x}, \boldsymbol{x}\rangle_{\mathbb{R}^n}}.$$

Then define a vector field $\boldsymbol{h}$ on $\mathbb{S}^n_+$ by

$$\boldsymbol{h}(z) = -\boldsymbol{D}\boldsymbol{\phi}_+(\boldsymbol{\phi}_+^{-1}(z)) \cdot \boldsymbol{g}(\boldsymbol{\phi}_+^{-1}(z)).$$

This vector field does not vanish on $\mathbb{S}^n_+$ and a computation gives $\boldsymbol{h}(z) = (0, \dots, 0, 1)$ for $\boldsymbol{z} \in E$, so $\boldsymbol{h}$ is consistently defined. Moreover, $\boldsymbol{h}$ is continuous and nowhere zero. This contradicts the Hairy Ball Theorem since we are assuming that n is even.

If n is odd, suppose again that $\boldsymbol{f}(\boldsymbol{x}) \neq \boldsymbol{x}$ for every $\boldsymbol{x} \in \mathbb{D}^n$. Then define $\boldsymbol{F}\colon \mathbb{D}^{n+1} \to \mathbb{D}^{n+1}$ by

$$\boldsymbol{F}(z_1, \dots, z_{n+1}) = (\boldsymbol{f}(z_1, \dots, z_n), 0),$$

and note that $\boldsymbol{F}$ is continuous and has the property that $\boldsymbol{F}(\boldsymbol{z}) \neq \boldsymbol{z}$ for every $\boldsymbol{z} \in \mathbb{D}^{n+1}$. But we have just showed that this is a contradiction since $n + 1$ is even. ■

For $n = 1$ this gives something familiar, and maybe helps us understand the Brouwer Fixed Point Theorem.

1.11.7 Corollary (A version of the Intermediate Value Theorem) *If* $\mathrm{f}\colon [-1, 1] \to [-1, 1]$ *is continuous then* f *has a fixed point.*

1.11.8 Corollary (Brouwer Fixed Point Theorem for sets homeomorphic to a disk) *Let* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *be a set homeomorphic to* $\mathbb{D}^{\mathrm{n}}$. *If* $\mathbf{f}\colon \mathrm{A} \to \mathrm{A}$ *is continuous then there exists* $\mathbf{x}_0 \in \mathrm{A}$ *such that* $\mathbf{f}(\mathbf{x}_0) = \mathbf{x}_0$.

Proof Let $\boldsymbol{\phi}\colon A \to \mathbb{D}^n$ be a homeomorphism and define $\boldsymbol{g}\colon \mathbb{D}^n \to \mathbb{D}^n$ by $\boldsymbol{g} = \boldsymbol{\phi} \circ \boldsymbol{f} \circ \boldsymbol{\phi}^{-1}$, i.e., such that the diagram

$$\begin{array}{ccc} A & \xrightarrow{f} & A \\ {\scriptstyle \phi}\downarrow & & \downarrow{\scriptstyle \phi} \\ \mathbb{D}^n & \xrightarrow[g]{} & \mathbb{D}^n \end{array}$$

commutes. Then $\boldsymbol{g}$ is continuous and so has a fixed point $\boldsymbol{y}_0$ by the Brouwer Fixed Point Theorem. If we take $\boldsymbol{x}_0 = \boldsymbol{\phi}^{-1} * \boldsymbol{y}_0$ then

$$\boldsymbol{f}(\boldsymbol{x}_0) = \boldsymbol{f} \circ \boldsymbol{\phi}^{-1}(\boldsymbol{y}_0) = \boldsymbol{\phi}^{-1} \circ \boldsymbol{g}(\boldsymbol{y}_0) = \boldsymbol{x}_0,$$

and so $\boldsymbol{x}_0$ is a fixed point for $\boldsymbol{f}$. ■

A common application of the preceding corollary occurs when A is a convex set, cf. Theorem 1.9.14.

The following result also gives some insight into the content of the Brouwer Foxed Point Theorem. We refer to Definition 1.10.48 and the ensuing discussion for a little context for the next result.

1.11.9 Proposition *The following statements are equivalent:*

(i) there exists a continuous map $\mathbf{f}\colon \mathbb{D}^n \to \mathbb{D}^n$ *that has no fixed point;*

(ii) there exists a retraction of $\mathbb{D}^n$ *onto* $\mathbb{S}^{n-1}$*.*

In particular, there is no retraction of $\mathbb{D}^n$ *onto* $\mathbb{S}^{n-1}$*.*

Proof First suppose that every continuous map $f\colon \mathbb{D}^n \to \mathbb{D}^n$ has a fixed point and that there exists a retraction $r\colon \mathbb{D}^n \to \mathbb{S}^{n-1}$. Let $g\colon \mathbb{S}^{n-1} \to \mathbb{S}^{n-1}$ be continuous and by Theorem 1.10.49 let $f\colon \mathbb{D}^n \to \mathbb{D}^n$ be a continuous extension of g. Let x_0 be a fixed point of f and note that $r \circ r(x_0) = r(x_0$ since $r|\mathbb{S}^{n-1} = \mathrm{id}_{\mathbb{S}^{n-1}}$. Therefore,

$$f \circ r(r(x_0)) = f \circ r(x_0),$$

giving $r(x_0)$ as a fixed point for $f|\mathbb{S}^{n-1} = g$. This implies that every continuous map $g\colon \mathbb{S}^{n-1} \to \mathbb{S}^{n-1}$ has a fixed point. However, this is not true since the continuous map $x \mapsto -x$ from $\mathbb{S}^{n-1}$ to $\mathbb{S}^{n-1}$ obviously has no fixed points. Therefore, if every continuous map $f\colon \mathbb{D}^n \to \mathbb{D}^n$ has a fixed point, there can be no retraction of $\mathbb{D}^n$ onto $\mathbb{S}^{n-1}$.

Next suppose that there exists a continuous map $f\colon \mathbb{D}^n \to \mathbb{D}^n$ having no fixed points. As in (1.14), the ray emanating from $f(x)$ through x intersects $\mathbb{S}^{n-1}$ at the point

$$r(x) \triangleq f(x) + \frac{\|x - f(x)\|^2_{\mathbb{R}^n} - \|x\|^2_{\mathbb{R}^n} + \|x - f(x)\|_{\mathbb{R}^n}}{\|x - f(x)\|^2_{\mathbb{R}^n}}(x - f(x)).$$

By construction we have $r(x) = x$ for $x \in \mathbb{S}^{n-1}$ and so r is a retraction of $\mathbb{D}^n$ onto $\mathbb{S}^{n-1}$. ■

1.11.4 The Kakutani Fixed Point Theorem

The Kakutani Fixed Point Theorem is a far-reaching generalisation of the Brouwer Fixed Point Theorem. Its statement and proof rely on notions of set-valued maps discussed in Section 1.10.5.

1.11.10 Theorem (Kakutani Fixed Point Theorem) *Let* $\mathrm{C} \subseteq \mathbb{R}^n$ *be compact and convex. If* $\mathbf{F}\colon \mathrm{C} \twoheadrightarrow \mathrm{C}$ *is upper semicontinuous and has the property that* $\mathbf{F}(\mathbf{x})$ *is compact and convex for each* $\mathbf{x} \in \mathrm{C}$*, then there exists* $\mathbf{x}_0 \in \mathrm{C}$ *such that* $\mathbf{x}_0 \in \mathbf{F}(\mathbf{x}_0)$*.*

Proof We can assume without loss of generality, essentially by Example 1.3.38–2, assume that C is n-dimensional.

We first prove the theorem in the case when $C = \Delta$, an n-simplex. For $j \in \mathbb{Z}_{>0}$ we define a map $f_j\colon \Delta \to \Delta$ as follows. Consider $\mathscr{S}^j_{\mathrm{bc}}(\Delta)$, the jth barycentric subdivision of Δ. There are N_j simplices of dimension zero in $\mathscr{S}^j_{\mathrm{bc}}(\Delta)$ which we denote by $u_{j1}, \dots, u_{jN_j}$. For each $k \in \{1, \dots, N_j\}$ let $v_{jk} \in F(u_{jk})$. We then require that $f_j(u_{jk}) = v_{jk}$. To define

$\boldsymbol{f}_j$ at an arbitrary point in Δ note that every point $\boldsymbol{x} \in \Delta$ lies in some uniquely defined simplex $\Delta' \in \mathscr{S}^j_{\mathrm{bc}}(\Delta)$. Let us denote

$$\Delta' = \Delta(\boldsymbol{u}_{jk_0(x)}, \boldsymbol{u}_{jk_1(x)}, \dots, \boldsymbol{u}_{jk_{m(x)}}).$$

We then have

$$\boldsymbol{x} = \sum_{l=0}^{m(\boldsymbol{x})} \lambda_l(\boldsymbol{x}) \boldsymbol{u}_{jk_l(x)}$$

for some uniquely defined strictly positive numbers $\lambda_0(\boldsymbol{x}), \lambda_1(\boldsymbol{x}), \dots, \lambda_{m(x)}(\boldsymbol{x})$ summing to one. Now define

$$\boldsymbol{f}_j(\boldsymbol{x}) = \sum_{l=0}^{m(\boldsymbol{x})} \lambda_l(\boldsymbol{x}) \boldsymbol{v}_{jk_l(x)}.$$

To see that $\boldsymbol{f}_j$ is continuous, note that we can write

$$\boldsymbol{x} = \sum_{k=0}^{N_j} \lambda_k(\boldsymbol{x}) \boldsymbol{u}_{jk},$$

where we now allow that some of the coefficients $\lambda_k(\boldsymbol{x})$ are zero. It is then easy to see that the functions $\lambda_k(\boldsymbol{x})$, $k \in \{1, \dots, N_j\}$, are continuous, essentially because barycentric coordinates define a homeomorphism of a simplex with the standard simplex. Since $\boldsymbol{f}_j$ is continuous, by the Brouwer Fixed Point Theorem there exists a fixed point $\boldsymbol{x}_j$ of $\boldsymbol{f}_j$. Since Δ is compact, by the Bolzano–Weierstrass Theorem there exists a convergent subsequence $(\boldsymbol{x}_{j_k})_{k\in\mathbb{Z}_{>0}}$ of the sequence $(\boldsymbol{x}_j)_{j\in\mathbb{Z}_{>0}}$. We claim that, if $\boldsymbol{x}_0 = \lim_{k\to\infty} \boldsymbol{x}_{j_k}$, then $\boldsymbol{x}_0 \in \boldsymbol{F}(\boldsymbol{x}_0)$.

Suppose that $\boldsymbol{x}_j \in \Delta_j$ for some (possibly not uniquely defined, but that's okay) n-dimensional simplex $\Delta_j \in \mathscr{S}^j_{\mathrm{bc}}(\Delta)$. Write

$$\Delta_j = \Delta(\boldsymbol{\mu}_{j0}, \boldsymbol{\mu}_{j1}, \dots, \boldsymbol{\mu}_{jn}).$$

By (1.51) it follows that $\lim_{k\to\infty} \boldsymbol{\mu}_{j_k l} = \boldsymbol{x}_0$ for each $l \in \{0, 1, \dots, n\}$. Let $\boldsymbol{\nu}_{jl} = \boldsymbol{f}_j(\boldsymbol{\mu}_{jl})$ for $j \in \mathbb{Z}_{>0}$ and $l \in \{0, 1, \dots, n\}$, noting that the definition of $\boldsymbol{f}_j$ ensures that $\boldsymbol{\nu}_{jl} \in \boldsymbol{F}(\boldsymbol{\mu}_{jl})$. Moreover, if we write

$$\boldsymbol{x} = \sum_{l=0}^{n} \lambda_{jl} \boldsymbol{\mu}_{jl}$$

for suitable nonnegative coefficients $\lambda_{j0}, \lambda_{j1}, \dots, \lambda_{jn}$ summing to 1, then the definition of $\boldsymbol{f}_j$ gives

$$\boldsymbol{x}_j = \boldsymbol{f}_j(\boldsymbol{x}_j) = \sum_{l=0}^{n} \lambda_{jl} \boldsymbol{\nu}_{jl}.$$

By the Bolzano–Weierstrass Theorem there exists a sequence $(j'_r)_{r\in\mathbb{Z}_{>0}}$ in $\mathbb{Z}_{>0}$ with the following properties:

1. $j'_r < j'_{r+1}$ for each $r \in \mathbb{Z}_{>0}$;
2. for each $r \in \mathbb{Z}_{>0}$ there exists $k \in \mathbb{Z}_{>0}$ such that $j'_r = j_k$, i.e., $\lim_{r\to\infty} \boldsymbol{x}_{j'_r} = \boldsymbol{x}_0$;

3. the sequences $(\boldsymbol{v}_{j'_r l})_{r\in\mathbb{Z}_{>0}}$, $l \in \{0, 1, \dots, n\}$, converge;
4. the sequences $(\lambda_{j'_r l})_{r\in\mathbb{Z}_{>0}}$, $l \in \{0, 1, \dots, n\}$, converge.

Let us denote

$$\boldsymbol{v}_{0l} = \lim_{r\to\infty} \boldsymbol{v}_{j'_r l}, \quad \lambda_{0l} = \lim_{r\to\infty} \lambda_{j'_r l},$$

and note that

$$\sum_{l=0}^{n} \lambda_{0l} = \sum_{l=0}^{n} \lim_{r\to\infty} \lambda_{j'_r l} = \lim_{r\to\infty} \sum_{l=0}^{n} \lambda_{j'_r l} = 1,$$
$$\lambda_{0l} \geq 0, \qquad l \in \{0, 1, \dots, n\}$$

and

$$\boldsymbol{x}_0 = \lim_{r\to\infty} \boldsymbol{x}_{j'_r} = \lim_{r\to\infty} \sum_{l=0}^{n} \lambda_{j'_r l} \boldsymbol{v}_{j'_r l} = \sum_{l=0}^{n} \lim_{r\to\infty} \lambda_{j'_r l} \boldsymbol{v}_{j'_r l} = \sum_{l=0}^{n} \lambda_{0l} \boldsymbol{v}_{0l}.$$

Now fix $l \in \{0, 1, \dots, n\}$. Note that

$$\lim_{r\to\infty} \boldsymbol{\mu}_{j'_r l} = \boldsymbol{x}_0, \quad \lim_{r\to\infty} \boldsymbol{v}_{j'_r l} = \boldsymbol{v}_{0l}, \quad \boldsymbol{v}_{j'_r l} \in \boldsymbol{F}(\boldsymbol{v}_{j'_r l}), \qquad r \in \mathbb{Z}_{>0}.$$

By Proposition 1.10.38 the upper semicontinuity of $\boldsymbol{F}$ ensures that $\boldsymbol{v}_{0l} \in \boldsymbol{F}(\boldsymbol{x}_0)$. Convexity of $\boldsymbol{F}(\boldsymbol{x}_0)$ then ensures that

$$\boldsymbol{x} = \sum_{l=0}^{n} \lambda_{0l} \boldsymbol{v}_{0l} \in \boldsymbol{F}(\boldsymbol{x}_0),$$

as desired.

Now we prove the theorem for a general compact and convex set C. We let Δ be an n-simplex for which $C \subseteq \Delta$. For simplicity of the ensuing constructions, let us also suppose that $\mathrm{bd}(C) \cap \mathrm{bd}(\Delta) = \emptyset$. Fix $\boldsymbol{z}_0 \in \mathrm{int}(C) \subseteq \mathrm{int}(\Delta)$. As we saw during the course of the proof of Theorem 1.9.14, if we let $\boldsymbol{u} \in \mathbb{S}^{n-1}$ then the ray $\rho_{\boldsymbol{x}_0,\boldsymbol{u}} \colon \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ defined by $\rho_{\boldsymbol{x},\boldsymbol{u}}(s) = \boldsymbol{x}_0 + s\boldsymbol{u}$ will intersect $\mathrm{bd}(C)$ and $\mathrm{bd}(\Delta)$ for unique $s_C(\boldsymbol{u}), s_\Delta(\boldsymbol{u}) \in \mathbb{R}_{>0}$. Since $C \subseteq \Delta$ we have $s_C(\boldsymbol{u}) < s_\Delta(\boldsymbol{u})$. Moreover, as also follows from the constructions of the proof of Theorem 1.9.14, for every $\boldsymbol{x} \in \Delta \setminus \{\boldsymbol{x}_0\}$, there exists a unique $\boldsymbol{u}(\boldsymbol{x}) \in \mathbb{S}^{n-1}$ and $s(\boldsymbol{x}) \in \mathbb{R}_{>0}$ such that $\boldsymbol{x} = \rho_{\boldsymbol{x}_0,\boldsymbol{u}(\boldsymbol{x})}(s(\boldsymbol{x}))$. Clearly $\boldsymbol{x} \in C \setminus \{\boldsymbol{x}_0\}$ if and only if $s(\boldsymbol{x}) \leq s_C(\boldsymbol{u}(\boldsymbol{x}))$ and $\boldsymbol{x} \in \Delta \setminus \{\boldsymbol{x}_0\}$ if and only if $s(\boldsymbol{x}) \leq s_\Delta(\boldsymbol{u}(\boldsymbol{x}))$. Now define $\boldsymbol{r} \colon \Delta \to C$ by

$$\boldsymbol{r}(\boldsymbol{x}) = \begin{cases} \boldsymbol{x}, & \boldsymbol{x} \in C, \\ \rho_{\boldsymbol{x}_0,\boldsymbol{u}(\boldsymbol{x})}(s_C(\boldsymbol{u}(\boldsymbol{x}))), & \boldsymbol{x} \in \Delta \setminus C. \end{cases}$$

Continuity of $\boldsymbol{r}$ follows from continuity of the various maps constructed in the proof of Theorem 1.9.14. The idea is that $\boldsymbol{r}$ retracts $\Delta \setminus C$ onto $\mathrm{bd}(C)$. Now define $\boldsymbol{F}' \colon \Delta \twoheadrightarrow C \subseteq \Delta$ by $\boldsymbol{F}'(\boldsymbol{x}) = \boldsymbol{F}(\boldsymbol{r}(\boldsymbol{x}))$. It is a dull exercise using continuity of the retraction $\boldsymbol{r}$ and Proposition 1.10.38 to show that $\boldsymbol{F}'$ is upper semicontinuous. Therefore, by the first part of the proof, there exists $\boldsymbol{x}_0 \in \Delta$ such that $\boldsymbol{x}_0 \in \boldsymbol{F}'(\boldsymbol{x}_0)$. Since $\boldsymbol{F}'(\boldsymbol{x}_0) \subseteq C$ it must then follow that $\boldsymbol{x}_0 \in C$. Therefore, $\boldsymbol{F}'(\boldsymbol{x}_0) = \boldsymbol{F}(\boldsymbol{x}_0)$ and so $\boldsymbol{x}_0 \in \boldsymbol{F}(\boldsymbol{x}_0)$, as desired. ■

We shall give a significant application to game theory of the Kakutani Fixed Point Theorem in Theorem 1.12.29.

1.11.5 Notes

Most proofs of the Brouwer Fixed Point Theorem are topological in nature. Our proof, relying only on multivariable calculus, closely follows Milnor [1978]. The Kakutani Fixed Point Theorem was proved, oddly enough, by Kakutani [1941].

Section 1.12

Game theory: An application of multivariable analysis

In this chapter, particularly in Sections 1.10 and 1.11, we have presented some fairly sophisticated and specialised ideas in real analysis without much context for how they are important or useful. While later volumes in this series will deal in great detail with applications of ideas that have their origin in this chapter. However, perhaps it is interesting to take a little self-contained detour to provide a non-trivial and (hopefully) interesting application of some of the techniques presented in this chapter. We have chosen to tell a little story behind game theory in order to achieve this end. This is *extremely* far from a complete treatment of game theory. It is merely a very selective set of ideas intended to serve as an illustration of real analysis.

Do I need to read this section? Read this section if and only if you think it might be interesting, or it might be useful to read it to give some context for some ideas in this chapter. •

1.12.1 Two-player games

Imagine a game with two players, call them A and B; let us restrict ourselves to two players for simplicity, since it is enough to illustrate the main ideas. The game consists of each player making a single decision. A decision means that each player makes a choice from a list of possible moves. Let us be precise about this by defining the notion of an outcome, by which we mean that which happens upon playing a game once.

1.12.1 Definition (Decision set, outcome) Let D_A and D_B be sets, called ***decision sets***. An element of D_A (resp. D_B) is a ***decision*** for player A (resp. player B). An ***outcome*** is an ordered pair $(a, b) \in D_A \times D_B$. •

We shall be interested mainly in finite decision sets which we denote by $D_A = \{a_1, \dots, a_m\}$ and $D_B = \{b_1, \dots, b_n\}$. In order to make it possible to declare winners and losers, we need to assign a value to the outcome of a game.

1.12.2 Definition (Value, two-player game) If D_A and D_B are decision sets, a pair of maps $v_A, v_B\colon D_A \times D_B \to \mathbb{R}$ are called the **A-*value function*** and the **B-*value function***, respectively. A ***two-player game*** is a quadruple (D_A, D_B, v_A, v_B) where D_A and D_B are decision sets and v_A and v_B are A- and B-value functions, respectively. A two-player game (D_A, D_B, v_A, v_B) is a ***zero-sum game*** if $v_B(a, b) = -v_A(a, b)$ for each outcome (a, b). •

The idea of a zero sum game is that the benefit to player A from outcome (a, b) is the same as the cost to player B of the same outcome.

Let us look at some simple examples.

1.12.3 Example (Prisoner's Dilemma) The Prisoner's Dilemma is a well-known simple example in game theory. The situation is this. We have two "players," let us call them A =Albert and B =Betty. Albert and Betty have been arrested for robbing a bank and are now being held in separate jail cells (i.e., they are unable to communicate). The District Attorney makes each of them the following offer.

1. You may confess or remain silent.
2. If you confess and your accomplice confesses, you both get prison sentences of C years.
3. If you confess and your accomplice remains silent, then you get set free (i.e., you get zero years in prison) and your accomplice gets D years in prison where $D > C$.
4. If you remain silent and your accomplice remains silent, then both get sentences of S where $S < C$.

(The letters C, D, and S stand for "confess," "deal," and "silence.") For Albert and Betty the decision set is thus $D_A = D_B = \{\text{confess}, \text{silent}\}$. Typically, for two-player games with finite decision sets, such as the Prisoner's Dilemma, one represents the A- and B-value functions via a matrix. For simplicity let us suppose that the value function is derived directly from the number of years, C and S, each might spend in prison. Note that these numbers should be thought of as having negative value, since a larger number for C and S has a lower value. Thus the A- and B-value functions are represented as follows:

Betty		
Albert	confess	silent
confess	$-C$	0
silent	$-D$	$-S$

Betty		
Albert	confess	silent
confess	$-C$	$-D$
silent	0	$-S$

The left table represents the A-value function and the right table represents the B-value function. Note that this is not a zero sum game.

It is possible to arrive at some sort of deductions about how Albert and Betty might reasonably behave. First of all, remaining silent has the least penalty (apart from being set free) and so must be thought of as being desirable. However, by remaining silent you run the risk of also incurring the maximum penalty if your partner confesses. It is not unreasonable, therefore, for both partners to confess in order to protect themselves from the maximum penalty. As we go along, we shall see that this is the outcome predicted by the decision making model we suggest. Note, however, that this solution is not a good one for either partner. Indeed, if they had agreed to remain silent, they would have incurred a smaller penalty. This is one of the contradictory facets that makes the Prisoner's Dilemma so interesting. •

1.12.4 Example (De Montmort's[16] gift) A father named Anthony (player A) gives his son named Bill (player B) an opportunity for a gift. The situation is this. Anthony grabs a number N of marbles.

1. If Bill guesses that N is even and N is in fact even, Anthony will give Bill $2.
2. If Bill guesses that N is odd and N is in fact even, Bill will have to give Anthony $1.
3. If Bill guesses that N is even and N is in fact odd, Bill will have to give Anthony $1.
4. If Bill guesses that N is odd and N is in fact odd, Anthony will give Bill $1.

Thus each player has the two-element decision set given by $\{\text{even}, \text{odd}\}$ based on (1) the number of marbles chosen in the case of Anthony and (2) the guess as to whether the number of marbles is even or odd in the case of Bill. The tables representing the values to each player are these:

	Bill	
Anthony	even	odd
even	−2	1
odd	1	−1

	Bill	
Anthony	even	odd
even	2	−1
odd	−1	1

The left table is the A-value table and the right table is th B-value table. This game is zero-sum.

Let us explore how each of Anthony and Bill might approach the game. If Bill adopts the "conservative" strategy of always choosing odd to minimise his losses. If Anthony is aware of his son's conservative nature, he will then choose an even number of marbles and get $1 from Bill. However, if Bill knows that Anthony thinks he is conservative, he will then use this against his father and choose even and so get his maximum gain of $2. But then if Anthony knows that Bill knows that Anthony believes that Bill will choose conservatively, Anthony will then go with an odd number of marbles. But then if Bill knows that Anthony knows that Bill knows that Anthony believes Bill will be conservative, be will then decide on guessing even. The point is that this second guessing goes in an endless loop, and neither will be in a position to make a "good" choice. •

Now one can imagine, with the data given thus far, that our two players A and B play a game, and that after playing the game, one is the winner and loser. Without any further information, there is not much one can really do in this framework. To give the affair some depth, there has to be some additional structure, and much of game theory revolves around what this structure is, and how it is to be modelled mathematically. The objective of this structure is that it should provide the players with some sort of basis for making decisions. One way to do this is to

[16]Pierre R/'emond de Montmort (1678–1719) was a French mathematician who made contributions to probability theory. The game to which we refer here was communicated in a letter to Nicoli Bernoulli in 1713.

add probabilistic effects into the setup. It is convenient to introduce some notation. For $n \in \mathbb{Z}_{>0}$ define

$$\Pi_n = \Big\{(p_1, \dots, p_n) \in \mathbb{R}^n \;\Big|\; p_1 \in \mathbb{R}_{>0},\ \sum_{j=1}^{n} p_j = 1\Big\}.$$

Note that Π_n is an $(n-1)$-dimensional convex set.

1.12.5 Definition (Mixed strategy) For a finite decision set $D_A = \{a_1, \dots, a_m\}$, a ***mixed strategy*** is a vector $\boldsymbol{p} = (p_1, \dots, p_m) \in \Pi_m$. A mixed strategy $\boldsymbol{p} \in \Pi_m$ is ***pure*** if $p_j = 1$ for some $j \in \{1, \dots, m\}$. For a mixed strategy p, p_j, $j \in \{1, \dots, m\}$, is the ***probability*** of player A making decision a_j. If $D_A = \{a_1, \dots, a_m\}$ and $D_B = \{b_1, \dots, b_n\}$ are two decision sets, a ***bistrategy*** is a pair $(\boldsymbol{p}, \boldsymbol{q}) \in \Pi_m \times \Pi_n$. •

It is worth thinking, for a moment, why we require that a mixed strategy be an element of the standard simplex. First of all, because each number p_j, $j \in \{1, \dots, m\}$, is a probability, we must have $p_j \in [0, 1]$. Moreover, since a move must be made, and once a move is made this means that none of the other moves have been made, we must have $p_1 + \cdots + p_m = 1$. It is certainly a debatable point whether the notion of a mixed strategy is reasonable in all or any instances. The supposition is that, in playing a game, a player's strategy amounts to a choice of probabilities she assigns to each decision. We will, however, sidestep this matter and just suppose that strategies are of the mixed variety. Note that a pure strategy is an extreme case, and is tantamount to a decision, which we distinguish from a strategy in general.

Initially, when talking about assessing the outcome of a game, we talked only about applying the value functions to outcomes of the game. Now, however, we have opened things up a little and instead wish to evaluate strategies.

1.12.6 Definition (Evaluation function) Let D_A and D_B be finite decision sets with m and n elements, respectively. A pair of maps $f_A, f_B \colon \Pi_m \times \Pi_n \to \mathbb{R}$ are called the **A**-***evaluation function*** and the **B**-***evaluation function***, respectively. A ***two-player game with mixed strategies*** if a quadruple (D_A, D_B, f_A, f_B) where D_A and D_B are finite decision sets and f_A and f_B are A- and B-evaluation functions, respectively. A two-player game with mixed strategies (D_A, D_B, f_A, f_B) is a ***zero-sum game with strategies*** if $f_B(\boldsymbol{p}, \boldsymbol{q}) = -f_A(\boldsymbol{p}, \boldsymbol{q})$ for all bistrategies $(\boldsymbol{p}, \boldsymbol{q}) \in \Pi_m \times \Pi_n$. •

As one might imagine, the notions of value functions for outcomes and evaluation functions for strategies are very often tied to one another. Indeed, there is a standard manner for relating these.

1.12.7 Definition (Canonical evaluation function) If $D_A = (a_1, \dots, a_m)$ and $D_B = (b_1, \dots, b_n)$ be decision sets with (D_A, D_B, v_A, v_B) an associated two-player game. The ***canonical evaluation functions*** for this game are given by

$$f_A(\boldsymbol{p}, \boldsymbol{q}) = \sum_{j=1}^{m} \sum_{k=1}^{n} v_A(a_j, b_k) p_j q_k, \quad f_B(\boldsymbol{p}, \boldsymbol{q}) = \sum_{j=1}^{m} \sum_{k=1}^{n} v_B(a_j, b_k) p_j q_k,$$

respectively. •

These definitions of evaluations for strategies are perfectly reasonable since they represented the probability-weighted sum of the values of the outcomes. In most applications it is the canonical evaluation functions one considers, however, for generality it is interesting to consider the possibility of evaluation functions that are not necessarily related at all to the value functions.

Let us consider our examples in terms of mixed strategies.

1.12.8 Example (Prisoner's Dilemma (cont'd)) A mixed strategy for Albert will be a pair (p_1, p_2) and a mixed strategy for Betty will be a pair (q_1, q_2). Since $p_1 + p_2 = 1$ and $q_1 + q_2 = 1$ we shall simply write $p_1 = p$, $p_2 = 1 - p$, $q_1 = q$, and $q_2 = 1 - q$. Thus p represents the probability that Albert will confess (and so $1 - p$ represents the probability that Albert will remain silent) and q represents the probability that Betty will confess (and so $1 - q$ represents the probability that Betty will remain silent). In this case the canonical evaluation functions are computed to be

$$\begin{aligned} f_A((p, 1-p), (q, 1-q)) &= (D - C - S)pq + Sp - (D - S)q - S, \\ f_B((p, 1-p), (q, 1-q)) &= (D - C - S)pq - (D - S)p + Sq - S. \end{aligned} \tag{1.66}$$

In Figure 1.26 we show the surface plots of these evaluation functions. •

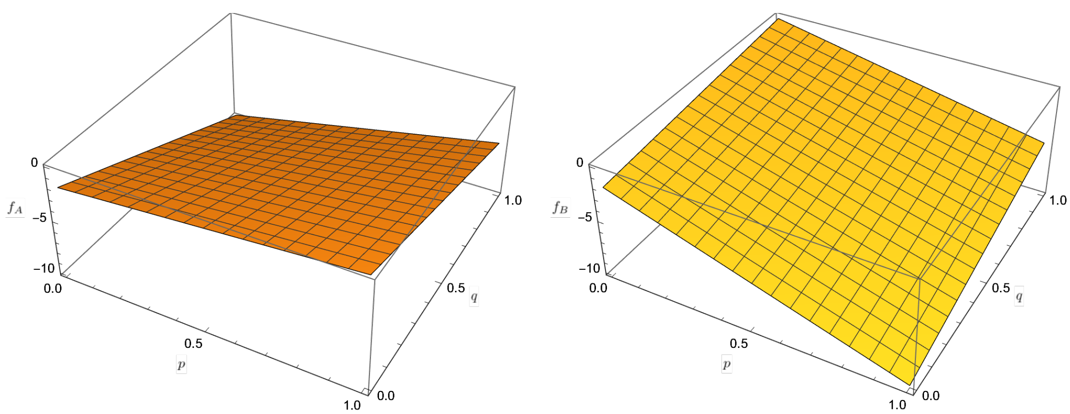

Figure 1.26 Canonical evaluation functions for the Prisoner's Dilemma for $C = 5$, $D = 10$, and $S = 2$, Albert's on the left and Betty's on the right

1.12.9 Example (De Montmort's gift (cont'd)) A mixed strategy for Anthony will be denoted by $(p, 1 - p)$ and a mixed strategy for Bill will be denoted by $(q, 1 - q)$. Thus p is the probability that Anthony will pick an even number of marbles (and so $1 - p$ is the probability that Anthony will pick an odd number of marbles) and q is the probability that Bill will guess an even number of marbles (and so $1 - q$

is the probability that Bill will guess an even number of marbles). The canonical evaluation functions are then readily computed to be

$$f_A((p, 1-p), (q, 1-q)) = -5pq + 2p + 2q - 1,$$
$$f_B((p, 1-p), (q, 1-q)) = 5pq - 2p - 2q + 1.$$

Note that $f_B = -f_A$ since this is a zero sum game. We shall explore this in more detail in Section 1.12.3. In Figure 1.27 we show the surface plots of the canonical

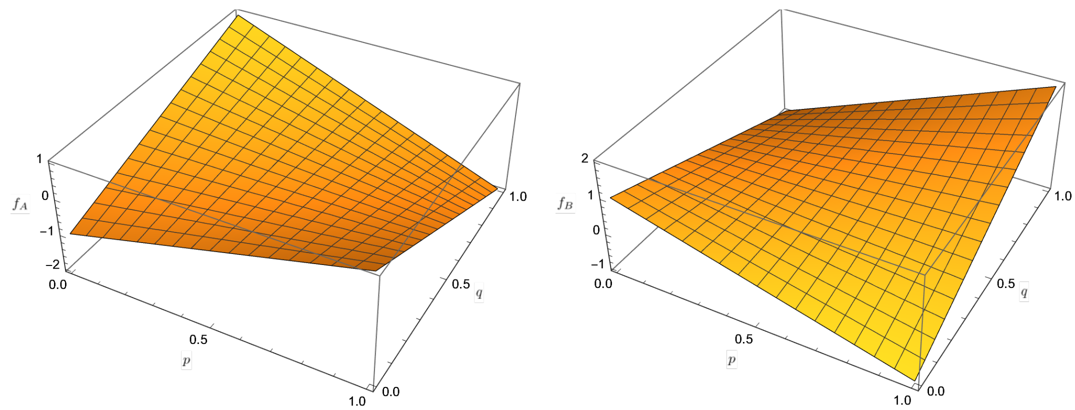

Figure 1.27 Canonical evaluation functions for de Montmort's gift, Anthony's (the father) on the left and Bill's (the son) on the right

evaluation functions. •

1.12.2 Decision theory

Let us summarise where we are now. We have two players A and B who each have finite decision sets D_A and D_B representing their "moves" in the game. For each pair of moves $(a_j, b_k) \in D_A \times D_B$ there is a value $v_A(a_j, b_k)$ that player A attaches to the outcome and a value $v_B(a_j, b_k)$ that player B attaches to that outcome. We moreover have decided that each player will play the game by choosing a mixed strategy which assigns a probability to each of their possible moves. And, in a way that may or may not be related to the value functions v_A and v_B, the players have evaluation functions f_A and f_B which tell them how good a pair of mixed strategies are for them. Now each player has to decide on the best strategy. There are various ways of approaching this, and we will discuss some of them.

To do this, we first generalise our discussion of game theory up to this point to emphasise the salient (to our objectives here) features.

1.12.10 Definition (Generalised two-player game) A ***generalised two-player game*** is a quadruple (P_A, P_B, f_A, f_B) where $P_A \subseteq \mathbb{R}^m$ and $P_B \subseteq \mathbb{R}^n$ and $f_A, f_B \colon P_A, P_B \to \mathbb{R}$. The sets P_A and P_B are the **A-*strategy set*** and the **B-*strategy set***, respectively.

The functions f_A and f_B are the **A-*evaluation function*** and **B-*evaluation function***, respectively. A ***strategy*** for player A (resp. play B) is an element $\boldsymbol{p} \in P_A$ (resp. $\boldsymbol{q} \in P_B$). A ***bistrategy*** is a pair $(\boldsymbol{p}, \boldsymbol{q}) \in P_A \times P_B$. •

In this generalisation we have removed the decision sets D_A and D_B from the setup, instead focusing on the strategies. As we saw in the preceding section, the setup with the decision sets is reducible to the setup of a generalised two-player game after one introduces the idea of a mixed strategy and uses the canonical evaluation functions. For our purposes this link between the decision sets D_A and D_B and the sets Π_m and Π_n is not as important as just the structure of the evaluation functions. It is useful to remark that in moving to this generalised setup, a possible confusion has been introduced. A game has now been reduced to a strategy. One should be careful to understand that the sets P_A and P_B represent strategies for playing the game, and are not the set of actual decisions available to the player. Therefore, a selection of a bistrategy $(\boldsymbol{p}, \boldsymbol{q})$ does not lead to a certain outcome of the game. This facet of strategy sees representation in the way people play actual games. It is not the case that in identical circumstances the same person will make the same decisions at all times about the move to make. It is debatable, however, whether the use of a mixed strategy to represent this variability is accurate. But this is rather outside the scope of mathematics, *per se*.

In any event, our focus now is on choosing strategies.

1.12.11 Definition (Decision rule) Let $P_A \subseteq \mathbb{R}^m$ and $P_B \subseteq \mathbb{R}^n$ be A- and B-strategy sets. A pair of set-valued maps $\boldsymbol{d}_A \colon P_B \twoheadrightarrow P_A$ and $\boldsymbol{d}_B \colon P_A \twoheadrightarrow P_B$ are called an **A-*decision rule*** and a **B-*decision rule***, respectively. •

The idea of a decision rule is the following. If player A knows that player B will be using strategy $\boldsymbol{q}_0 \in P_B$, then the A-decision rule $\boldsymbol{d}_A \colon P_B \to P_A$ indicates that player A will choose their strategies from $\boldsymbol{d}_A(\boldsymbol{q}_0)$. Similar interpretations, of course, hold for a B-decision rule.

Given a generalised two-player game (P_A, P_B, f_A, f_B), there is a natural way to choose a decision rule.

1.12.12 Definition (Canonical decision rule) For a generalised two-player game (P_A, P_B, f_A, f_B), the ***canonical decision rules*** are the A- and B-decision rules $\boldsymbol{c}_A \colon P_B \twoheadrightarrow P_A$ and $\boldsymbol{c}_B \colon P_A \twoheadrightarrow P_B$ given by

$$\begin{aligned}\boldsymbol{c}_A(\boldsymbol{q}_0) &= \{\boldsymbol{p}_0 \in P_A \mid f_A(\boldsymbol{p}_0, \boldsymbol{q}_0) = \sup\{f_A(\boldsymbol{p}, \boldsymbol{q}_0) \mid \boldsymbol{p} \in P_A\}\},\\ \boldsymbol{c}_B(\boldsymbol{p}_0) &= \{\boldsymbol{q}_0 \in P_B \mid f_B(\boldsymbol{p}_0, \boldsymbol{q}_0) = \sup\{f_B(\boldsymbol{p}_0, \boldsymbol{q}) \mid \boldsymbol{q} \in P_B\}\},\end{aligned}$$

respectively. •

The idea is that, knowing that player B will use strategy $\boldsymbol{q}_0$, player A canonically chooses a strategy that will maximise the A-evaluation function. The set of such strategies forms the canonical decision rule $\boldsymbol{c}_A$ for player A. The same interpretation, with the players swapped, holds for $\boldsymbol{c}_B$.

Let us determine the canonical decision rules for the simple games we introduced in the preceding section.

1.12.13 Example (Prisoner's Dilemma (cont'd)) We fix $\boldsymbol{q}_0 = (q_0, 1 - q_0)$ so that, to determine the canonical decision rule for Albert at $\boldsymbol{q}_0$ we must find the set of p_0's that maximise the function

$$[0,1] \ni p \mapsto (D - C - S)pq_0 + Sp - (D - S)q_0 - S.$$

Maxima can arise in two ways. They can occur in the interior of $[0,1]$ (where one can use the derivative tests of Theorem I-3.2.16) or they occur on the boundary of $[0,1]$. To check for the instances in the interior we compute the derivative of the function we need to maximise to be $(D - C)q_0 + S(1 - q_0)$. That is, the derivative is constant; this is obvious as the function is linear in p. Therefore, we need only compute the boundary values to see where the maximum occurs. However, given that $D > C$, $S > 0$, and $q_0 \in [0,1]$, it follows that the derivative is always positive, and so the value at $p = 1$ will always exceed the value at $p = 0$. Thus we have

$$\boldsymbol{c}_A(\boldsymbol{q}_0) = \{(1,0)\}.$$

Similarly, for Betty's canonical decision rule we have, with $\boldsymbol{p}_0 = (p_0, 1 - p_0)$,

$$\boldsymbol{c}_B(\boldsymbol{p}_0) = \{(1,0)\}.$$

Thus, for the Prisoner's Dilemma, the only strategy that is feasible is that of both players always confessing. Note, however, that both Albert and Betty would both be better served by always remaining silent. This points out a limitation of this methodology of modelling human behaviour. •

1.12.14 Example (De Montmort's gift (cont'd)) For the game between the father Anthony and the son Bill, let us fix a strategy $\boldsymbol{q}_0 = (q_0, 1-q_0)$ for Bill. The function for Anthony evaluating his possible strategies is

$$[0,1] \ni p \mapsto -5pq_0 + 2p + 2q_0 - 1.$$

The derivative of this function is $-5q_0 + 2$. The sign of this derivative varies depending on the value of q_0. Thus to evaluate the maximum value we need to compare the value at $p = 0$ (which is $2q_0 - 1$) with the value at $p = 1$ (which is $-3q_0 + 1$). If the former value exceeds the latter, then Bill will choose the strategy $(0,1)$ and otherwise will choose the strategy $(1,0)$. There is the special case when the values are equal, and in this case Anthony can choose from all possible strategies for the canonical decision rule. In summary,

$$\boldsymbol{c}_A(\boldsymbol{q}_0) = \begin{cases} \{(0,1)\}, & 5q_0 - 2 > 0, \\ \{(1,0)\}, & 5q_0 - 2 < 0, \\ \Pi_2, & 5q_0 - 2 = 0 \end{cases}$$

and, similarly if $\boldsymbol{p}_0 = (p_0, 1 - p_0)$,

$$\boldsymbol{c}_B(\boldsymbol{p}_0) = \begin{cases} \{(0,1)\}, & 5p_0 - 2 < 0, \\ \{(1,0)\}, & 5p_0 - 2 > 0, \\ \Pi_2, & 5p_0 - 2 = 0. \end{cases}$$

Let us see if this makes sense by looking at some extreme case. If $q_0 = 0$ (i.e., if Bill selects the strategy of always guessing odd) then the canonical strategy for Anthony will be to always take an even number of marbles. At the other extreme, if $q_0 = 1$ (i.e., if Bill selects the strategy of always guessing even) then the canonical strategy for Anthony will be to always take an odd number of marbles. At the special value of $q_0 = \frac{2}{5}$ Anthony will allow all possible strategy. This is reasonable. We leave it for the reader to verify the reasonableness of the canonical decision rule for player Bill. •

Given that players A and B have adopted decision rules $\boldsymbol{d}_A$ and $\boldsymbol{d}_B$, there are then distinguished bistrategies.

1.12.15 Definition (Consistent bistrategy) Let P_A and P_B be strategy sets with $\boldsymbol{d}_A \colon P_B \twoheadrightarrow P_A$ and $\boldsymbol{d}_B \colon P_A \twoheadrightarrow P_B$ being A- and B-decision rules. A bistrategy $(\boldsymbol{p}_0, \boldsymbol{q}_0)$ is **$(\mathbf{d_A}, \mathbf{d_B})$-*consistent*** if $\boldsymbol{p}_0 \in \boldsymbol{d}_A(\boldsymbol{q}_0)$ and if $\boldsymbol{q}_0 \in \boldsymbol{d}_B(\boldsymbol{p}_0)$. •

The idea of a consistent bistrategy $(\boldsymbol{q}_0, \boldsymbol{p}_0)$ is that, if all player A knows is that player B will use strategy $\boldsymbol{q}_0$ and if all player B knows is that player A will use strategy $\boldsymbol{p}_0$, then both players will be content with using strategies $\boldsymbol{p}_0$ and $\boldsymbol{q}_0$. That is to say, with the information they have, there will be no incentive to change their strategies.

An important concept extracted from the above discussion is the following.

1.12.16 Definition (Non-cooperative equilibrium or Nash equilibrium) Let (P_A, P_B, f_A, f_B) be a generalised two-played game. A ***non-cooperative equilibrium*** or a ***Nash equilibrium*** for the game is a consistent bistrategy for the canonical decision rule. •

The potential problem with a non-cooperative equilibrium $(\boldsymbol{p}_0, \boldsymbol{q}_0)$ is that there may be strategies $\boldsymbol{p}$ and $\boldsymbol{q}$ for which

$$f_A(\boldsymbol{p}, \boldsymbol{q}) > f_A(\boldsymbol{p}_0, \boldsymbol{q}_0), \quad f_B(\boldsymbol{p}, \boldsymbol{q}) > f_B(\boldsymbol{p}_0, \boldsymbol{q}_0).$$

This leads to the following notion.

1.12.17 Definition (Pareto optimal) A ***Pareto optimal*** bistrategy for a generalised two-player game (P_A, P_B, f_A, f_B) is a bistrategy $(\boldsymbol{p}_0, \boldsymbol{q}_0)$ such that, for every bistrategy $(\boldsymbol{p}, \boldsymbol{q})$ it holds that $f_A(\boldsymbol{p}, \boldsymbol{q}) \leq f_A(\boldsymbol{p}_0, \boldsymbol{q}_0)$ or $f_B(\boldsymbol{p}, \boldsymbol{q}) \leq f_B(\boldsymbol{p}_0, \boldsymbol{q}_0)$. •

The idea of a Pareto optimal bistrategy is that is may be possible that one player or the other may be able to do better with a different strategy, but it is not possible for both players to simultaneously do better. If one chooses a non-cooperative equilibrium $(\boldsymbol{p}_0, \boldsymbol{q}_0)$ that is not Pareto optimal, and if Pareto optimal bistrategies exist, both players would want to choose these other strategies which do better for them than the non-cooperative equilibrium. To do this, however, they will have to exchange information, in the sense that each will have to know that the other will act so as to mutually maximise the value of the outcome of the game. Generally speaking, it will not be possible for both players to simultaneously maximise their gains. That is to say, generally there may be no Pareto optimal bistrategies.

There are some obvious sorts of Pareto optimal bistrategies that are interesting in that they are extreme cases.

1.12.18 Definition (Single-player Pareto optimal) For a generalised two-player game (P_A, P_B, f_A, f_B), a ***Pareto optimal*** bistrategy for player A is a bistrategy $(\boldsymbol{p}_0, \boldsymbol{q}_0)$ for which

$$f_A(\boldsymbol{p}_0, \boldsymbol{q}_0) = \sup\{f_A(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{p} \in P_A,\ \boldsymbol{q} \in P_B\}.$$

The value

$$v_A^{\max} = \sup\{f_A(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{p} \in P_A,\ \boldsymbol{q} \in P_B\}$$

is the ***maximum gain*** for player A. •

By definition, a single-player Pareto optimal bistrategy is necessarily Pareto optimal. Of course, one can similarly define the notion of a Pareto optimal bistrategy for player B and the maximum gain $v_B^{\max}$ for player B. Note that player B will only be content with a Pareto optimal bistrategy for player A if player B's objective is to maximise the gain of player A. Thus, in this case, player B is acting solely in the interests of player A. This is obviously an extreme situation.

Let us consider some of these concepts for our examples.

1.12.19 Example (Prisoner's Dilemma (cont'd)) Since there is only one strategy played by each player in the canonical decision rule for the Prisoner's Dilemma, it is easy to characterise all the phenomenon we have discussed so far. For example, there is only one consistent bistrategy for the canonical decision rule, and thus the only non-cooperative or Nash equilibrium is the bistrategy $((1,0),(1,0))$ where both Albert and Betty confess all the time.

We also claim that the only Pareto optimal bistrategy is $((1,0),(1,0))$. To see this, one can examine the evaluation functions f_A and f_B in (1.66) as functions of p and q. For $(p,q) \in [0,1] \times [0,1] \setminus \{(1,1)\}$ there then exists $(r,s) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}$ such that $(p+r, q+s) \in [0,1] \times [0,1]$. The derivative analysis of Example 1.12.13 we can conclude that for such as (r,s) we have

$$\begin{aligned} f_A((p+r, 1-p-r), (q+s, 1-q-s)) &> f_A((p, 1-p), (q, 1-q)), \\ f_B((p+r, 1-p-r), (q+s, 1-q-s)) &> f_B((p, 1-p), (q, 1-q)). \end{aligned}$$

This prohibits $((p, 1-p), (q, 1-q))$ from being a Pareto optimal bistrategy unless $p = q = 1$, just as claimed. As a consequence of this we have

$$v_A^{\max} = f_A((1,0),(1,0)) = -C, \quad v_B^{\max} = f_B((1,0),(1,0)) = -C.$$

This makes sense, of course, since under the Pareto optimal bistrategy both Albert and Betty will confess, and so will go to prison for C years. •

1.12.20 Example (De Montmort's gift (cont'd)) For the Anthony/Bill gift problem things are more interesting. We claim that the only Nash equilibrium is $((\frac{2}{5}, \frac{3}{5}), (\frac{2}{5}, \frac{3}{5}))$. To verify this we investigate various cases for a general bistrategy $((p_0, 1-p_0), (q_0, 1-q_0))$.

1. $5q_0 - 2 > 0$ and $5p_0 - 2 > 0$: In this case consistency and $5q_0 - 2 > 0$ implies that $p_0 = 0$ and this contradicts $5p_0 - 2 > 0$. This case, therefore, cannot happen.
2. $5q_0 - 2 > 0$ and $5p_0 - 2 < 0$: In this case consistency and $5p_0 - 2 < 0$ implies that $q_0 = 0$ and this contradicts $5q_0 - 2 > 0$. This case, therefore, cannot happen.
3. $5q_0 - 2 > 0$ and $5p_0 - 2 = 0$: In this case consistency and $p_0 = \frac{2}{5}$ implies that $5q_0 - 2 = 0$. This case, therefore, cannot happen.
4. $5q_0 - 2 < 0$ and $5p_0 - 2 > 0$: In this case consistency and $5p_0 - 2 > 0$ implies that $q_0 = 1$ and this contradicts $5q_0 - 2 < 0$. This case, therefore, cannot happen.
5. $5q_0 - 2 < 0$ and $5p_0 - 2 < 0$: In this case consistency and $5q_0 - 2 < 0$ implies that $p_0 = 1$ and this contradicts $5p_0 - 2 < 0$. This case, therefore, cannot happen.
6. $5q_0 - 2 < 0$ and $5p_0 - 2 = 0$: In this case consistency and $p_0 = \frac{2}{5}$ implies that $5q_0 - 2 = 0$. This case, therefore, cannot happen.
7. $5q_0 - 2 = 0$ and $5p_0 - 2 > 0$: In this case consistency and $q_0 = \frac{2}{5}$ implies that $5p_0 - 2 - 0$. This case, therefore, cannot happen.
8. $5q_0 - 2 = 0$ and $5p_0 - 2 < 0$: In this case consistency and $q_0 = \frac{2}{5}$ implies that $5p_0 - 2 - 0$. This case, therefore, cannot happen.
9. $5q_0 - 2 = 0$ and $5p_0 - 2 = 0$: In this case we see that no contradictions arise.

We claim that every bistrategy is Pareto optimal. Indeed, consider a bistrategies $(\boldsymbol{p}_0, \boldsymbol{q}_0) \in \Pi_2 \times \Pi_2$. If $(\boldsymbol{p}_0, \boldsymbol{q}_0)$ is not Pareto optimal then there exists a bistrategy $(\boldsymbol{p}, \boldsymbol{q})$ for which

$$f_A(\boldsymbol{p}, \boldsymbol{q}) > f_A(\boldsymbol{p}_0, \boldsymbol{q}_0), \quad f_B(\boldsymbol{p}, \boldsymbol{q}) > f_B(\boldsymbol{p}_0, \boldsymbol{q}_0).$$

However, since $f_B = -f_A$ this is obviously impossible.

Since every neighbourhood of a point $(\boldsymbol{p}_0, \boldsymbol{q}_0) \in \operatorname{int}(\Pi_2 \times \Pi_2)$ must contain points $(\boldsymbol{p}_1, \boldsymbol{q}_1)$ and $(\boldsymbol{p}_2, \boldsymbol{q}_2)$ for which

$$f_A(\boldsymbol{p}_1, \boldsymbol{q}_1) > f_A(\boldsymbol{p}_0, \boldsymbol{q}_0), \quad f_B(\boldsymbol{p}_2, \boldsymbol{q}_2) > f_B(\boldsymbol{p}_0, \boldsymbol{q}_0),$$

we conclude that the maximum values of f_A and f_B must occur on the boundary of $\Pi_2 \times \Pi_2$. Thus, to determine $v_A^{\max}$ we need only take the maximum of the following four numbers:

$$\sup\{2q-1 \mid q \in [0,1]\}, \quad \sup\{-3q+1 \mid q \in [0,1]\},$$
$$\sup\{2p-1 \mid p \in [0,1]\}, \quad \sup\{-3p+1 \mid p \in [0,1]\}.$$

This gives $v_A^{\max} = 1$. Similarly, to determine $v_B^{\max}$ we take the maximum of the following four numbers:

$$\sup\{-2q+1 \mid q \in [0,1]\}, \quad \sup\{3q-1 \mid q \in [0,1]\},$$
$$\sup\{-2p+1 \mid p \in [0,1]\}, \quad \sup\{3p-1 \mid p \in [0,1]\}.$$

This gives $v_B^{\max} = 2$.

That every strategy is Pareto optimal is something rather special about this game, and some consequences of this are explored in Section 1.12.3. Moreover, we will also see in this general development that our prediction above about the Nash equilibria can be derived from a general theorem for certain types of games. •

Next we consider another extreme situation. In this setup, let us suppose that (1) player B acts in such a way as to minimise the gains of player A with no thought to her own gain and/or (2) player A is convinced that player B is going to act this way. In such a situation, one might expect player A to choose a strategy that maximises the minimum gain, no matter what player B does. To investigate this strategy we make the following definition.

1.12.21 Definition (Least gain function) For a generalised two-player game (P_A, P_B, f_A, f_B), the ***least gain function*** for player A is the function $f_A^{\min} \colon P_A \to \mathbb{R}$ defined by

$$f_A^{\min}(\boldsymbol{p}) = \inf\{f_A(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{q} \in P_B\}. \qquad \bullet$$

The least gain function tells us that, if player A chooses strategy $\boldsymbol{p} \in P_A$, then, given our assumptions of the malevolence of player B, player A can expect player B to choose a strategy $\boldsymbol{q} \in P_B$ to achieve the gain $f_A^{\min}(\boldsymbol{p})$. Now, given this, player A might be compelled to act in such a way as to maximise the least gain.

1.12.22 Definition (Conservative strategy) For a generalised two-player game (P_A, P_B, f_A, f_B), a ***conservative strategy*** for player A is a strategy $\boldsymbol{p}_0 \in P_A$ such that

$$f_A^{\min}(\boldsymbol{p}_0) = \sup\{f_A^{\min}(\boldsymbol{p}) \mid \boldsymbol{p} \in P_A\}.$$

The value

$$v_A^{\min} = \sup\{f_A^{\min}(\boldsymbol{p}) \mid \boldsymbol{p} \in P_A\} = \sup\{\inf\{f_A(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} \mid \boldsymbol{p} \in P_A\}$$

is the ***conservative value*** for player A. •

The way to think of a conservative strategy is this. Player A may use her conservative strategy as a threat. That is, player A can threaten to reject a bistrategy $(\boldsymbol{p}, \boldsymbol{q})$ for which $f_A(\boldsymbol{p}, \boldsymbol{q}) < v_A^{\min}$ since, by adopting a conservative strategy, player A can ensure that her value will exceed the value $f_A(\boldsymbol{q}, \boldsymbol{p})$. So the point is that player B should consider the possibility of player A using a conservative strategy in selecting his own strategy.

Of course, the preceding two definitions and the discussion surrounding them can be applied to player B, assuming the malevolent intentions of player A. This would define the least gain function $f_B^{\min}\colon P_B \to \mathbb{R}$ for player B and the conservative value

$$v_B^{\min} = \sup\{f_B^{\min}(\boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} = \sup\{\inf\{f_B(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{p} \in P_A\} \mid \boldsymbol{q} \in P_B\}$$

for player B.

The idea for conservative bistrategies is that the only strategies $(\boldsymbol{p}, \boldsymbol{q})$ of interest to both players are those for which

$$f_A(\boldsymbol{p}, \boldsymbol{q}) \geq v_A^{\min}, \quad f_B(\boldsymbol{p}, \boldsymbol{q}) \geq v_B^{\min}.$$

Let us determine the conservative strategies for our examples.

1.12.23 Example (Prisoner's Dilemma (cont'd)) For $\boldsymbol{p} = (p, 1-p)$, since

$$\frac{\partial f_A}{\partial q}(\boldsymbol{p}, \boldsymbol{q}) = (1-p)(S-D) - Cp, \qquad \boldsymbol{q} \in \Pi_2$$

we see that this partial derivative is always negative. Therefore, $\boldsymbol{q} \mapsto f_A(\boldsymbol{p}, \boldsymbol{q})$ will take its minimum at $q = 1$. Using this observation we compute

$$f_A^{\min}(\boldsymbol{p}) = -(1-p)D - Cp.$$

Therefore, the maximum value of $f_A^{\min}$ will always be achieved at $p = 1$. This gives $v_A^{\min} = -C$. A similar analysis gives $v_B^{\min} = -C$. Therefore, the only conservative strategy for Albert is $(1, 0)$ (always confess) and the only conservative strategy for Betty is also $(1, 0)$ (always confess). This is consistent with the "threat" interpretation of a conservative strategy. •

1.12.24 Example (De Montmort's gift (cont'd)) For $\boldsymbol{p} = (p, 1-p)$ we compute $\frac{\partial f_A}{\partial q} = -5p+2$ which is positive for $p \in [0, \frac{2}{5})$ and negative for $p \in (\frac{2}{5}, 1]$. Thus the minimum value of $\boldsymbol{q} \mapsto f_A(\boldsymbol{q}, \boldsymbol{p})$ will be achieved at $q = 0$ for $p \in [0, \frac{2}{5})$ and at $q = 1$ for $p \in (\frac{2}{5}, 1]$. This gives

$$f_A^{\min}(\boldsymbol{p}) = \begin{cases} 2p - 1, & p \in [0, \frac{2}{5}), \\ -3p + 1, & p \in [\frac{2}{5}, 1]. \end{cases}$$

This function achieves its maximum at $p = \frac{2}{5}$ and the value of this maximum is $v_A^{\min} = -\frac{1}{5}$. For Bill, a similar analysis gives

$$f_B^{\min}(\boldsymbol{q}) = \begin{cases} 3q - 1, & q \in [0, \frac{2}{5}), \\ -2q + 1, & q \in [\frac{2}{5}, 1] \end{cases}$$

and $v_B^{\min} = \frac{1}{5}$ which is achieved at $q = \frac{2}{5}$. Thus Anthony and Bill both have the single conservative strategy $(\frac{2}{5}, \frac{2}{5})$. Note that this also agrees with the Nash equilibrium. This is not an accident as we shall see in Proposition 1.12.28. •

We have thus characterised strategy choices where players act in the following four extreme ways:

1. player B acts solely to maximise the gain for player A with no thought to their own gain;
2. player A acts solely to maximise the gain for player B with no though to their own gain;
3. player B acts solely to minimise the gain for player A with no thought to their own gain;
4. player A acts solely to minimise the gain for player B with no thought to their own gain.

This leads to the following definition of the sorts of strategies in which one ought to be interested.

1.12.25 Definition (Viable bistrategy) For a generalised two-player game (P_A, P_B, f_A, f_B), a bistrategy $(\boldsymbol{p}, \boldsymbol{q})$ is ***viable*** if

$$f_A(\boldsymbol{p}, \boldsymbol{q}) \in [v_A^{\min}, v_A^{\max}], \quad f_B(\boldsymbol{p}, \boldsymbol{q}) \in [v_B^{\min}, v_B^{\max}].$$ •

The point is that, as long as both players are playing rationally, no matter what strategy they choose, their joint strategies will be a viable bistrategy.

1.12.3 Zero-sum games

The reader will have noticed an utter lack of results thus far in our discussion. Indeed, one of the challenges of game theory is to come up with criterion for limiting the circumstances of the game so that one can make useful assertions about whether certain strategies exist, etc. One of the instances when one can prove useful theorems is in the case of zero sum games.

1.12.26 Definition (Zero-sum generalised game) A generalised two-player game (P_A, P_B, f_A, f_B) is a ***zero-sum*** if there exists $f\colon P_A \times P_B \to \mathbb{R}$ such that

$$f_A(\boldsymbol{p}, \boldsymbol{q}) = -f_B(\boldsymbol{p}, \boldsymbol{q}) = f(\boldsymbol{p}, \boldsymbol{q}).$$

We will denote such a generalised two-player zero-sum game by (P_A, P_B, f). •

Let us first state an obvious characterisation of some of the general concepts as they appear for zero-sum games.

1.12.27 Proposition (Some properties of zero-sum games) *Let* $(\mathrm{P}_A, \mathrm{P}_B, \mathrm{f})$ *be a generalised two-player zero-sum game and let* $\mathrm{f}_A = \mathrm{f}$ *and* $\mathrm{f}_B = -\mathrm{f}$*. For the generalised two-player game* $(\mathrm{P}_A, \mathrm{P}_B, \mathrm{f}_A, \mathrm{f}_B)$ *the following statements hold:*

(i) $\mathrm{v}_A^{\max} = \sup\{\mathrm{f}(\mathbf{p}, \mathbf{q}) \mid \mathbf{p} \in \mathrm{P}_A,\ \mathbf{q} \in \mathrm{P}_B\}$;

(ii) $\mathrm{v}_B^{\max} = \inf\{\mathrm{f}(\mathbf{p}, \mathbf{q}) \mid \mathbf{p} \in \mathrm{P}_A,\ \mathbf{q} \in \mathrm{P}_B\}$;

(iii) $\mathrm{f}_A^{\min}(\mathbf{p}) = \inf\{\mathrm{f}(\mathbf{p}, \mathbf{q}) \mid \mathbf{q} \in \mathrm{P}_B\}$;

(iv) $\mathrm{f}_B^{\min}(\mathbf{q}) = \sup\{\mathrm{f}(\mathbf{p}, \mathbf{q}) \mid \mathbf{p} \in \mathrm{P}_A\}$;

(v) $\mathrm{v}_A^{\min} = \sup\{\inf\{\mathrm{f}(\mathbf{p}, \mathbf{q}) \mid \mathbf{q} \in \mathrm{P}_B\} \mid \mathbf{p} \in \mathrm{P}_A\}$;

(vi) $\mathrm{v}_B^{\min} = \inf\{\sup\{\mathrm{f}(\mathbf{p}, \mathbf{q}) \mid \mathbf{p} \in \mathrm{P}_A\} \mid \mathbf{q} \in \mathrm{P}_B\}$;

(vii) *all bistrategies are Pareto optimal.*

Proof The only possibly non-obvious statement is the last one, and it is argued in general just as we did in particular in Example 1.12.20. ■

We can also give a convenient characterisation of the non-cooperative equilibria for a zero-sum game.

1.12.28 Proposition (Non-cooperative equilibria for zero-sum games) *Let* $(\mathrm{P}_A, \mathrm{P}_B, \mathrm{f})$ *be a generalised two-player zero-sum game and let* $\mathrm{f}_A = \mathrm{f}$ *and* $\mathrm{f}_B = -\mathrm{f}$*. For a bistrategy* $(\mathbf{p}_q, \mathbf{q}_0) \in \mathrm{P}_A \times \mathrm{P}_B$ *the following statements are equivalent:*

(i) $(\mathbf{p}_0, \mathbf{q}_0)$ *is a non-cooperative equilibrium;*

(ii) $\mathrm{f}(\mathbf{p}, \mathbf{q}_0) \le \mathrm{f}(\mathbf{p}_0, \mathbf{q}_0) \le \mathrm{f}(\mathbf{p}_0, \mathbf{q})$ *for every* $\mathbf{p} \in \mathrm{P}_A$ *and* $\mathbf{q} \in \mathrm{P}_B$;

(iii) *the following three statements hold:*

(a) $\mathrm{v}_A^{\min} = \mathrm{v}_B^{\min}$;

(b) $\mathbf{p}_0$ *is a conservative strategy for player* A;

(c) $\mathbf{q}_0$ *is a conservative strategy for player* B.

Proof (i) $\Longrightarrow$ (ii) If $(\boldsymbol{p}_0, \boldsymbol{q}_0)$ is a non-cooperative equilibrium then

$$f(\boldsymbol{p}_0, \boldsymbol{q}_0) = \sup\{f(\boldsymbol{p}, \boldsymbol{q}_0) \mid \boldsymbol{p} \in P_A\}, \quad f(\boldsymbol{p}_0, \boldsymbol{q}_0) = \inf\{f(\boldsymbol{p}_0, \boldsymbol{q}) \mid \boldsymbol{q} \in P_B\}$$

by definition. From this (ii) immediately follows.

(ii) $\Longrightarrow$ (iii) First we note that

$$\begin{aligned} & \inf\{f(\boldsymbol{p}, \boldsymbol{q}') \mid \boldsymbol{q}' \in P_B\} \le \sup\{f(\boldsymbol{p}', \boldsymbol{q}) \mid \boldsymbol{p}' \in P_A\}, \qquad \boldsymbol{p} \in P_A,\ \boldsymbol{q} \in P_B, \\ \Longrightarrow \quad & \sup\{\inf\{f(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} \mid \boldsymbol{p} \in P_A\} \\ & \le \inf\{\sup\{f(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{p} \in P_A\} \mid \boldsymbol{q} \in P_B\} \end{aligned}$$

giving $v_A^{\min} \le v_B^{\min}$ in all cases. We also always have

$$f_A^{\min}(\boldsymbol{p}) \le v_A^{\min}, \quad f_B^{\min}(\boldsymbol{q}) \ge v_B^{\min}, \qquad \boldsymbol{p} \in P_A,\ \boldsymbol{q} \in P_B.$$

Now assuming (ii) we have

$$\begin{aligned} & f(\boldsymbol{p},\boldsymbol{q}_0) \le f(\boldsymbol{p}_0,\boldsymbol{q}_0) \le f(\boldsymbol{p}_0,\boldsymbol{q}), \qquad \boldsymbol{p} \in P_A,\ \boldsymbol{q} \in P_B, \\ \implies \quad & \sup\{f(\boldsymbol{p},\boldsymbol{q}_0) \mid \boldsymbol{p} \in P_A\} \le f(\boldsymbol{p}_0,\boldsymbol{q}_0) \le \inf\{f(\boldsymbol{p}_0,\boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} \\ \implies \quad & \inf\{\sup\{f(\boldsymbol{p},\boldsymbol{q}) \mid \boldsymbol{p} \in P_A\} \mid \boldsymbol{q} \in P_B\} \le f(\boldsymbol{p}_0,\boldsymbol{q}_0) \\ & \le \sup\{\inf\{f(\boldsymbol{p},\boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} \mid \boldsymbol{p} \in P_A\}, \end{aligned}$$

giving $v_B^{\min} \le v_A^{\min}$, and so $v_A^{\min} = v_B^{\min}$. This also gives

$$f(\boldsymbol{p}_0,\boldsymbol{q}_0) = v_A^{\min} = v_B^{\min}.$$

From this we ascertain that

$$\begin{aligned} & f(\boldsymbol{p}_0,\boldsymbol{q}_0) \le f(\boldsymbol{p}_0,\boldsymbol{q}), \qquad \boldsymbol{q} \in P_B \\ \implies \quad & v_A^{\min} = f(\boldsymbol{p}_0,\boldsymbol{q}_0) \le f_A^{\min}(\boldsymbol{p}_0), \end{aligned}$$

giving $f^{\min}(\boldsymbol{p}_0) = v_A^{\min}$, and so $\boldsymbol{p}_0$ is a conservative strategy for player A. Similarly,

$$\begin{aligned} & f(\boldsymbol{p},\boldsymbol{q}_0) \le f(\boldsymbol{p}_0,\boldsymbol{q}_0), \qquad \boldsymbol{p} \in P_A \\ \implies \quad & f_B^{\min}(\boldsymbol{q}_0) \le f(\boldsymbol{p}_0,\boldsymbol{q}_0) = v_B^{\min}, \end{aligned}$$

giving $f_B^{\min}(\boldsymbol{q}_0) = v_B^{\min}$, and so $\boldsymbol{q}_0$ is a conservative strategy for player B.

(iii) $\implies$ (i) Note that we always have

$$f(\boldsymbol{p},\boldsymbol{q}) \le \sup\{f(\boldsymbol{p}',\boldsymbol{q}) \mid \boldsymbol{p} \in P_A\}, \quad f(\boldsymbol{p},\boldsymbol{q}) \ge \inf\{f(\boldsymbol{p},\boldsymbol{q}') \mid \boldsymbol{q} \in P_B\}$$

for any bistrategy $(\boldsymbol{p},\boldsymbol{q})$. Under the assumptions of part (iii) we have

$$v_A^{\min} = f_A^{\min}(\boldsymbol{p}_0) \le f(\boldsymbol{p}_0,\boldsymbol{q}_0) \le f_B^{\min}(\boldsymbol{q}_0) = v_B^{\min}$$

which gives, upon using the fact that $v_A^{\min} = v_B^{\min}$,

$$f(\boldsymbol{p}_0,\boldsymbol{q}_0) = \inf\{f(\boldsymbol{p}_0,\boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} = \sup\{f(\boldsymbol{p},\boldsymbol{q}_0) \mid \boldsymbol{p} \in P_A\}.$$

or that

$$f(\boldsymbol{p},\boldsymbol{q}_0) \le f(\boldsymbol{p}_0,\boldsymbol{q}_0) \le f(\boldsymbol{p}_0,\boldsymbol{q})$$

for every bistrategy $(\boldsymbol{p},\boldsymbol{q})$ (i.e., we are proving part (i) by first proving (ii)). It, therefore, follows that

$$\sup\{f(\boldsymbol{p},\boldsymbol{q}_0) \mid \boldsymbol{p} \in P_A\} \le f(\boldsymbol{p}_0,\boldsymbol{q}_0) \le \inf\{f(\boldsymbol{p}_0,\boldsymbol{q}) \mid \boldsymbol{q} \in P_B\}$$

which gives

$$f(\boldsymbol{p}_0,\boldsymbol{q}_0) = \sup\{f(\boldsymbol{p},\boldsymbol{q}_0) \mid \boldsymbol{p} \in P_A\}, \quad f(\boldsymbol{p}_0,\boldsymbol{q}_0) = \inf\{f(\boldsymbol{p}_0,\boldsymbol{q}) \mid \boldsymbol{q} \in P_B\},$$

as desired. ■

The Minimax Theorem of von Neumann tells us that there is a large class of games that possess non-cooperative equilibria.

1.12.29 Theorem (von Neumann's[17] Minimax Theorem) *Let* $(\mathrm{P_A}, \mathrm{P_B}, \mathrm{f})$ *be a generalised two-player zero-sum game such that*

(i) $\mathrm{P_A} \subseteq \mathbb{R}^m$ *and* $\mathrm{P_B} \subseteq \mathbb{R}^n$ *are compact and convex,*

(ii) f *is continuous,*

(iii) for each $\mathbf{p}_0 \in \mathrm{P_A}$ *and* $\alpha \in \mathbb{R}$ *the set* $\{\mathbf{q} \in \mathrm{P_B} \mid \mathrm{f}(\mathbf{p}_0, \mathbf{q}) \leq \alpha\}$ *is convex, and*

(iv) for each $\mathbf{q}_0 \in \mathrm{P_B}$ *and* $\beta \in \mathbb{R}$ *the set* $\{\mathbf{p} \in \mathrm{P_A} \mid \mathrm{f}(\mathbf{p}, \mathbf{q}_0) \geq \beta\}$ *is convex.*

Then there are non-cooperative equilibria for the game.

Proof We let $f_A = f$ and $f_B = -f$. Note that $f_A^{\min}$ is upper semicontinuous and $f_B^{\min}$ is lower semicontinuous by Proposition 1.10.17. Denote

$$S = \{(\boldsymbol{p}_0, \boldsymbol{q}_0) \in P_A \times P_B \mid f(\boldsymbol{p}_0, \boldsymbol{q}_0) = f_A^{\min}(\boldsymbol{p}_0)\},$$
$$T = \{(\boldsymbol{p}_0, \boldsymbol{q}_0) \in P_A \times P_B \mid f(\boldsymbol{p}_0, \boldsymbol{q}_0) = f_B^{\min}(\boldsymbol{q}_0)\}.$$

Compactness of $P_A \times P_B$ and continuity of f ensures that these sets are nonempty by Theorem 1.3.31. We also claim that both S and T are closed. Let us show that S is closed. Let $((\boldsymbol{p}_j, \boldsymbol{q}_j))_{j\in\mathbb{Z}_{>0}}$ be a sequence in S converging to $(\boldsymbol{p}_0, \boldsymbol{q}_0) \in P_A \times P_B$. Using Proposition 1.10.13 we have

$$f(\boldsymbol{p}_0, \boldsymbol{q}_0) = \lim_{j\to\infty} f(\boldsymbol{p}_j, \boldsymbol{q}_j) = \lim_{j\to\infty} f_A^{\min}(\boldsymbol{p}_j) \leq \limsup_{j\to\infty} f_A^{\min}(\boldsymbol{p}_j) \leq f_A^{\min}(\boldsymbol{p}_0).$$

By definition of $f_A^{\min}(\boldsymbol{p}_0)$ it, therefore, follows that $f(\boldsymbol{p}_0, \boldsymbol{p}_0) = f_A^{\min}(\boldsymbol{p}_0)$, and so $(\boldsymbol{p}_0, \boldsymbol{q}_0) \in S$. One similarly shows that T is closed. Let us define

$$S_{\boldsymbol{p}_0} = \{\boldsymbol{q} \in P_B \mid (\boldsymbol{p}_0, \boldsymbol{q}) \in S\} = \{\boldsymbol{q} \in P_B \mid f(\boldsymbol{p}_0, \boldsymbol{q}) \leq f_A^{\min}(\boldsymbol{p}_0)\},$$
$$T_{\boldsymbol{q}_0} = \{\boldsymbol{p} \in P_A \mid (\boldsymbol{p}, \boldsymbol{q}_0) \in T\} = \{\boldsymbol{p} \in P_A \mid f(\boldsymbol{p}, \boldsymbol{q}_0) \geq f_B^{\min}(\boldsymbol{q}_0)\}.$$

Then $S_{\boldsymbol{p}_0}$ and $T_{\boldsymbol{q}_0}$ are nonempty, closed, and convex for each $\boldsymbol{p}_0 \in P_A$ and $\boldsymbol{q}_0 \in P_B$.

Define $\boldsymbol{F}\colon P_A \times P_B \twoheadrightarrow P_A \times P_B$ by $\boldsymbol{F}(\boldsymbol{p}, \boldsymbol{q}) = T_{\boldsymbol{q}} \times S_{\boldsymbol{p}}$. We claim that $\boldsymbol{F}$ is upper semicontinuous. By Proposition 1.10.38 it suffices to show that $\boldsymbol{F}$ is closed, i.e., that $\mathrm{graph}(\boldsymbol{F})$ is a closed subset of $(P_A \times P_B) \times (P_A \times P_B)$. Let $((\boldsymbol{z}_j, \boldsymbol{z}'_j))_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathrm{graph}(\boldsymbol{F})$ converging to some $(\boldsymbol{z}_0, \boldsymbol{z}'_0) \in P_A \times P_B$. Let us write $\boldsymbol{z}_0 = (\boldsymbol{p}_0, \boldsymbol{q}_0)$, $\boldsymbol{z}'_0 = (\boldsymbol{p}'_0, \boldsymbol{q}'_0)$, and $\boldsymbol{z}_j = (\boldsymbol{p}_j, \boldsymbol{q}_j)$ and $\boldsymbol{z}'_j = (\boldsymbol{p}'_j, \boldsymbol{q}'_j)$ for $j \in \mathbb{Z}_{>0}$. Note that $\boldsymbol{p}'_j \in T_{\boldsymbol{q}_j}$ and $\boldsymbol{q}'_j = S_{\boldsymbol{p}_j}$. It follows that $(\boldsymbol{p}_j, \boldsymbol{q}'_j) \in S$ and $(\boldsymbol{p}'_j, \boldsymbol{q}_j) \in T$. Closedness of S and T, as shown above, ensures that

$$(\boldsymbol{p}_0, \boldsymbol{q}'_0) = \lim_{j\to\infty} (\boldsymbol{p}_j, \boldsymbol{q}'_j) \in S, \quad (\boldsymbol{p}'_0, \boldsymbol{q}_0) = \lim_{j\to\infty} (\boldsymbol{p}'_j, \boldsymbol{q}_j) \in T.$$

Therefore, $\boldsymbol{p}'_0 \in T_{\boldsymbol{p}_0}$ and $\boldsymbol{q}'_0 \in S_{\boldsymbol{q}_0}$ and so $(\boldsymbol{z}_0, \boldsymbol{z}'_0) \in \mathrm{graph}(\boldsymbol{F})$.

[17] John von Neumann (1903–1957) was Hungarian born, and was one of the leading figures in mathematics, and indeed science, in the twentieth century. He was one of the original members of the Institute for Advanced Study. His significant mathematical contributions include those made to game theory and functional analysis. His functional analysis contributions were made as part of his efforts to put quantum mechanics on a firm mathematical foundation. Von Neumann also was involved in the Manhattan Project, the undertaking by the United States to develop the atomic bomb during the Second World War.

By the Kakutani Fixed Point Theorem it now follows that there exists $(\boldsymbol{p}_0, \boldsymbol{q}_0) \in P_A \times P_B$ such that $\boldsymbol{p}_0 \in T_{\boldsymbol{q}_0}$ and $\boldsymbol{q}_0 \in S_{\boldsymbol{p}_0}$. This means that

$$f(\boldsymbol{p}_0, \boldsymbol{q}_0) = \inf\{f(\boldsymbol{p}_0, \boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} = \sup\{f(\boldsymbol{p}, \boldsymbol{q}_0) \mid \boldsymbol{p} \in P_A\}.$$

Therefore,

$$\begin{aligned}\inf\{\sup\{f(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{p} \in P_A\} \mid \boldsymbol{q} \in P_B\} &\le \sup\{f(\boldsymbol{p}, \boldsymbol{q}_0) \mid \boldsymbol{p} \in P_A\} \\ &= f(\boldsymbol{p}_0, \boldsymbol{p}_0) = \inf\{f(\boldsymbol{p}_0, \boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} \le \sup\{\inf\{f(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} \mid \boldsymbol{p} \in P_A\}.\end{aligned}$$

On the other hand,

$$\begin{aligned} & f(\boldsymbol{p}, \boldsymbol{q}) \le \sup\{f(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{p} \in P_A\}, \qquad \boldsymbol{p} \in P_A,\ \boldsymbol{q} \in P_B \\ \implies\ & \inf\{f(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} \le \inf\{\sup\{f(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{p} \in P_A\} \mid \boldsymbol{q} \in P_B\}, \qquad \boldsymbol{p} \in P_A \\ \implies\ & \sup\{\inf\{f(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{q} \in P_B\} \mid \boldsymbol{p} \in P_A\} \\ & \le \inf\{\sup\{f(\boldsymbol{p}, \boldsymbol{q}) \mid \boldsymbol{p} \in P_A\} \mid \boldsymbol{q} \in P_B\},\end{aligned}$$

and so the theorem follows. ■

We have seen that the game we have been calling De Montmort's gift is a zero sum game. Let us also verify that is satisfies the hypotheses of the von Neumann Minimax Theorem.

1.12.30 Example (De Montmort's gift (cont'd)) Note that $P_A = P_B = \Pi_2$ and so P_A and P_B are certainly compact and convex. As we have been doing, we denote an element of P_A by $(p, 1-p)$ and an element of P_B by $(q, 1-q)$. This gives us natural identifications of P_A and P_B with $[0, 1]$. We then have

$$f((p, 1-p), (q, 1-q)) = -5pq + 2p + 2q - 1,$$

and so f is obviously continuous. Since the functions

$$p \mapsto f((p, 1-p), (q, 1-q)), \quad q \mapsto f((p, 1-p), (q, 1-q)), \qquad p, q \in [0, 1],$$

are linear it follows that the sets

$$\{\boldsymbol{q} \in P_B \mid f(\boldsymbol{p}, \boldsymbol{q}) \le \alpha\}, \quad \{\boldsymbol{p} \in P_A \mid f(\boldsymbol{p}, \boldsymbol{q}) \ge \beta\}$$

are subintervals of identified with $[0, 1]$. In particular, they are convex. This gives us the hypotheses of the von Neumann Minimax Theorem for this game. Therefore, we can conclude the existence of non-cooperative equilibria. Moreover, any non-cooperative equilibria will satisfy the equivalent conditions of Proposition 1.12.28. We have already seen, in fact, that there is a single non-cooperative equilibrium at $((\frac{2}{5}, \frac{3}{5}), (\frac{2}{5}, \frac{3}{5}))$, and that the value of this strategy for Anthony is $\frac{1}{5}$ and the value for Bill is $-\frac{1}{5}$. Not much of a gift.

But we do see that the employment of a mixed strategy, i.e., the introduction of probabilistic aspects into the decision making process, has allowed to arrive at a conclusion, rather than just engage in the circular reasoning we saw in Example 1.12.4. •

1.12.4 Notes

The Minimax Theorem we give as Theorem 1.12.29 was first proved by Neumann [1928]. A clear and rigorous discussion of issues related to this can be found in the book of Aubin [1998]. It is for his generalisation of the Minimax Theorem to more than two players (see [Nash 1951]) that John Nash won a Nobel Prize in Economics. The term "Prisoner's Dilemma" is due to Albert Tucker (1905–1995), a Canadian-born mathematician, is an attempt to make palatable research by mathematicians Melvin Dresher (1911–1992) and Merrill M. Flood (????-????) at the RAND Corporation.

This version: 2022/03/07

Chapter 2

Vector Calculus

The subject of this chapter, vector calculus, is a fundamental tool for many physical models. For example, continuum mechanics and electromagnetism rely essentially on basic (and sometimes not so basic) machinery from vector calculus. We shall do some physical modelling in Chapter V-1 and the reader will see then the rôle of vector calculus. We shall also see in Chapter 3 that ideas from this chapter play an important part in complex function theory where integration over paths provides deep and nontrivial insight into what it means for a function to be holomorphic.

One of the ideas of vector calculus is that one may want to coherently integrate over sets with zero volume. In our development of the Riemann integral in Section 1.6 we saw that if we use the standard theory to integrate over a set of volume zero the result will be zero; see Proposition 1.6.21. However, for example, one may want to integrate over a n-dimensional sphere in $\mathbb{R}^{n+1}$ and obtain something nonzero. We will also need to integrate functions taking values in multi-dimensional Euclidean space. In some sense this is trivial (one just does the integration componentwise), but it does not take long before one needs more sophistication to really understand these integrals.

Do I need to read this chapter? The material in Section 2.2 is essential for reading Chapter 3. However, the remainder of the material in the chapter can be skipped until it is needed. •

Contents

Section 2.1

The Cauchy–Bochner integral

We have thus far only indicated how to integrate functions that are $\mathbb{R}$-valued. It is easy to extend the integral to functions taking values in $\mathbb{R}^n$.

2.1.1 Definition (Integral of a $\mathbb{R}^m$-valued function) Let $A \subseteq \mathbb{R}^n$ and let $\boldsymbol{f}\colon \mathbb{R}^n \to \mathbb{R}^m$ be written as $\boldsymbol{f}(\boldsymbol{x}) = (f_1(\boldsymbol{x}), \dots, f_m(\boldsymbol{x}))$. The function $\boldsymbol{f}$ is ***Riemann integrable*** (in the sense of either of Definitions 1.6.8 or 1.6.22) if each of the functions $f_1, \dots, f_m\colon A \to \mathbb{R}$ are Riemann integrable, and the ***Riemann integral*** of $\boldsymbol{f}$ is

$$\int_A \boldsymbol{f}(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \Big(\int_A f_1(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}, \dots, \int_A f_m(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\Big) \in \mathbb{R}^m. \qquad \bullet$$

That is, to integrate a vector-valued function, one simply integrates each of its components. The following result gives an alternative characterisation that is useful in that it suggests how one might define the integral for functions taking values in vector spaces other than $\mathbb{R}^m$. We shall not explore this further here, but will pick this up in . The statement of the result uses the following notation. Given $A \subseteq \mathbb{R}^n$, $\boldsymbol{f}\colon A \to \mathbb{R}^m$, and $\boldsymbol{\alpha} \in (\mathbb{R}^m)'$, define $f_{\boldsymbol{\alpha}}\colon A \to \mathbb{R}$ by $f_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \boldsymbol{\alpha}(\boldsymbol{f}(\boldsymbol{x}))$. what?

2.1.2 Proposition (Characterisation of $\mathbb{R}^m$-valued Riemann integrable functions) *For $\mathrm{A} \subseteq \mathbb{R}^\mathrm{n}$ and for $\mathbf{f}\colon \mathrm{A} \to \mathbb{R}^\mathrm{m}$ the following statements are equivalent:*

(i) $\mathbf{f}$ is Riemann integrable;

(ii) the function $\mathrm{f}_{\boldsymbol{\alpha}}$ is Riemann integrable for each $\boldsymbol{\alpha} \in (\mathbb{R}^\mathrm{n})'$.

Moreover, if $\mathbf{f}$ is Riemann integrable then

$$\int_\mathrm{A} \mathbf{f}(\mathbf{x})\,\mathrm{d}\mathbf{x} = \Big(\int_\mathrm{A} \mathrm{f}_{\boldsymbol{\alpha}_1}(\mathbf{x})\,\mathrm{d}\mathbf{x}, \dots, \int_\mathrm{A} \mathrm{f}_{\boldsymbol{\alpha}_\mathrm{m}}(\mathbf{x})\,\mathrm{d}\mathbf{x}\Big),$$

where $\{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_\mathrm{m}\}$ is the basis for $(\mathbb{R}^\mathrm{m})'$ dual to the standard basis.

Proof First suppose that $\boldsymbol{f}$ is Riemann integrable and let $\boldsymbol{\alpha} \in (\mathbb{R}^m)'$. Let $\alpha_1, \dots, \alpha_m \in \mathbb{R}$ be the components of $\boldsymbol{\alpha}$ relative to the basis $\{\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m\}$ is the basis for $(\mathbb{R}^m)'$. Then we have

$$f_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \alpha_1 f_1(\boldsymbol{x}) + \dots + \alpha_m f_m(\boldsymbol{x}). \tag{2.1}$$

The Riemann integrability of $f_{\boldsymbol{\alpha}}$ now follows from the linearity of the Riemann integral proved in Proposition 1.6.28.

Conversely, if $f_{\boldsymbol{\alpha}}$ is Riemann integrable for every $\boldsymbol{\alpha} \in (\mathbb{R}^m)'$ then, in particular, $f_{\boldsymbol{\alpha}_j}$ is Riemann integrable for each $j \in \{1, \dots, m\}$. However, from (2.1) we see that $f_{\boldsymbol{\alpha}_j} = f_j$ for each $j \in \{1, \dots, m\}$, and so the Riemann integrability of $\boldsymbol{f}$ follows. ■

We have the following useful inequality which generalises that of Proposition 1.6.31.

2.1.3 Proposition (Riemann integral and Euclidean norm) *If* $\mathrm{A} \subseteq \mathbb{R}^{\mathrm{n}}$ *and if* $\mathbf{f}\colon \mathrm{A} \to \mathbb{R}^{\mathrm{m}}$ *is Riemann integrable then the function* $\mathbf{x} \mapsto \|\mathbf{f}(\mathbf{x})\|_{\mathbb{R}^{\mathrm{m}}}$ *is Riemann integrable and, moreover,*

$$\left\|\int_{\mathrm{A}} \mathbf{f}(\mathbf{x})\,\mathrm{d}\mathbf{x}\right\|_{\mathbb{R}^{\mathrm{m}}} \leq \int_{\mathrm{A}} \|\mathbf{f}(\mathbf{x})\|_{\mathbb{R}^{\mathrm{n}}}\,\mathrm{d}\mathbf{x}.$$

Proof Let $f_1, \dots, f_m \colon A \to \mathbb{R}$ be the components of f which are each Riemann integrable by definition. By Lemma 1.2.67 we have

$$\|f(x)\|_{\mathbb{R}^m} \leq |f_1(x)| + \cdots + |f_m(x)|$$

for every $x \in A$. By Propositions 1.6.28, 1.6.30, and 1.6.31 we conclude that $x \mapsto \|f(x)\|_{\mathbb{R}^m}$ is Riemann integrable.

Then we have

$$\begin{aligned}
\left\|\int_A f(x)\,\mathrm{d}x\right\|_{\mathbb{R}^m}^2 &= \left|\int_A f_1(x)\,\mathrm{d}x\right|^2 + \cdots + \left|\int_A f_m(x)\,\mathrm{d}x\right|^2 \\
&\leq \Big(\int_A |f_1(x)|\,\mathrm{d}x\Big)^2 + \cdots + \Big(\int_A |f_m(x)|\,\mathrm{d}x\Big)^2 \\
&= \left\|\int_A |f_1(x)|\,\mathrm{d}x, \dots, \int_A |f_m(x)|\,\mathrm{d}x\right\|_{\mathbb{R}^m}^2 \\
&\leq
\end{aligned}$$

by Lemma 1.2.67 and

■

Section 2.2

Path integrals

2.2.1 Paths in $\mathbb{R}^n$

In this section, mainly for organisational purposes, we define what we mean by a path and indicate some manipulations one can perform with paths. These ideas will be important in Chapter 3, and will also see other uses in the text.

2.2.1 Definition (Path) A ***path*** in $\mathbb{R}^n$ is a continuous map $\gamma\colon [a,b] \to \mathbb{R}^n$, and a path is said to ***connect*** the points $\gamma(a)$ and $\gamma(b)$. The ***trace*** of a path γ is the subset $\operatorname{image}(\gamma)$ of $\mathbb{R}^n$. •

2.2.2 Remark (Domains of paths) Note that if $\gamma\colon [a,b] \to \mathbb{R}^n$ is a curve connecting x_0 and x_1, then the curve $\tilde{\gamma}\colon [c,d] \to \mathbb{R}^n$ defined by

$$\tilde{\gamma}(s) = \gamma\Big(\tfrac{d-c}{b-a}s + \tfrac{bc-ad}{b-a}\Big)$$

also connects x_0 and x_1. Moreover, the paths γ and $\tilde{\gamma}$ have the same trace. Thus the trace represents the points through which the path passes, and different paths are capable of having the same trace. Sometimes a path is restricted to be defined on $[0,1]$, and our preceding machinations indicate this can be done without loss of generality. •

One often has a picture in one's mind of continuous paths as being far simpler than they actually are. It is actually quite important to understand that continuous paths can be rather wild. Indeed, one of the most important theorems in mathematics is Cauchy's Theorem in complex analysis which has to do with integrals around closed continuous paths in the complex plane. In order to understand the depth of this theorem in its most general form requires an appreciation of the possible complexities of continuous paths. This is resolved for the situation of Cauchy's Theorem by the so-called "Jordan Curve Theorem" which we state as Theorem 3.1.5. Let us here illustrate a rather unfriendly continuous path.

2.2.3 Example (A path surjective onto a two-dimensional region) We construct a path from $[0,1]$ to $[0,1] \times [0,1]$ that is surjective and differentiable except on a set of measure zero.

Let us partition $[0,1]$ into four closed intervals I^1_j, $j \in \{1,2,3,4\}$ of equal length enumerated from left to right. Let us also partition the square $[0,1]\times[0,1]$ into four closed squares S^1_j, $j \in \{1,2,3,4\}$ of equal size, enumerating these in some arbitrary manner. Note that there is a path passing through the four squares in the same order as the order of the intervals as in Figure 2.1.

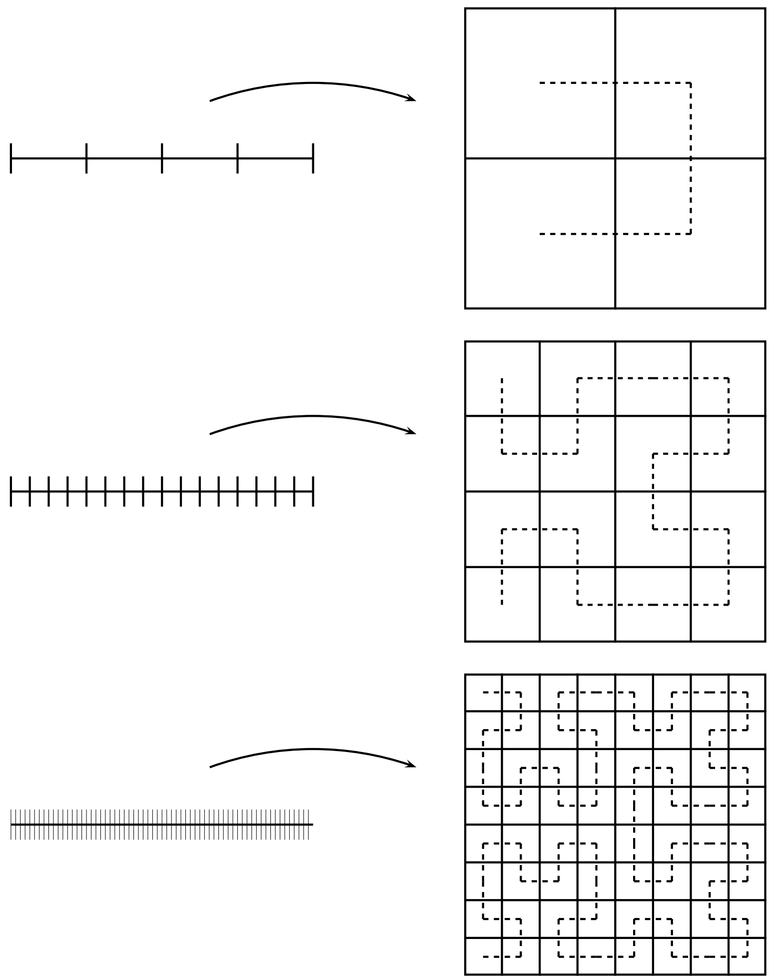

Figure 2.1 The sequential construction of a path from $[0,1]$ onto $[0,1]\times[0,1]$

Now take each of the intervals I_j^1 and partition again it into four closed intervals of equal length. For each $j \in \{1,2,3,4\}$ denote the subintervals of I_j^1 by $I^2_{4(j-1)+1},\dots,I^2_{4j}$, retaining the enumeration from left to right. The result is 4^2 intervals of equal length denoted by $I^2_1,\dots,I^2_{4^2}$. Now partition each of the squares S_j^1, $j\in\{1,2,3,4\}$, into four closed squares of equal size. For each $j\in\{1,2,3,4\}$ denote the subsquares of S_j^1 by $S^2_{4(j-1)+1},\dots,S^2_{4j}$. Order the indices so that the subsquare $S^2_{4(j-1)+1}$ has an edge bordering S^1_{j-1} if $j\in\{2,3,4\}$ and so that the subsquare S^2_{4j} has an edge bordering on S^1_{j+1} for $j\in\{1,2,3\}$. This gives 4^2 squares of equal size which are denoted by $S^2_1,\dots,S^2_{4^2}$. The enumeration of the squares ensures that there is a path passing through the squares in the same order as the intervals; again see Figure 2.1.

This process can be repeated. At step k one ends up with 4^k closed subintervals of $[0,1]$ denoted $I_1^k,\dots,I_{4^k}^k$, ordered from left to right, and 4^k squares $S_1^k,\dots,S_{4^k}^k$ enumerated so that there is a path passing through the squares in the same order as the parameter for the path passes through the intervals. In Figure 2.1 we show the case where $k\in\{1,2,3\}$.

With this sequential construction we now define $\gamma\colon[0,1]\to[0,1]\times[0,1]$ as follows. Let $s\in[0,1]$. For each $k\in\mathbb{Z}_{>0}$ the point s lies in an interval $I_{j_k(s)}^k$ for some $j_k(s)\in\{1,\dots,4^k\}$. If s lies in more than one of the intervals, i.e., when s lies on the boundary of one of the intervals, then choose $j_k(s)$ arbitrarily between the two possibilities. Note that if s lies on the intersection of two intervals I_j^k and I_{j+1}^k then this means that s will lie in the intersection of two intervals at each of the subsequent stages, $k+1,k+2,\dots$. That is, once on encounters an ambiguity in the definition of $j_k(s)$ at some stage, this ambiguity will be present in the choice of $j_{k+1}(s),j_{k+2}(s),\dots$. We then insist that the choices be made so that $I_{j_{k+1}(s)}^{k+1}\subseteq I_{j_k(s)}^k$, $I_{j_{k+1}(s)}^{k+1}\subseteq I_{j_k(s)}^k$, and so on. That is to say, if we adopt this rule, there is at most one step at which there is ambiguity in the definition of $j_k(s)$. Having now chosen a sequence $(j_k(s))_{k\in\mathbb{Z}_{>0}}$, one then has the corresponding sequence $(S_{j_k(s)}^k)_{k\in\mathbb{Z}_{>0}}$ of squares having the property that $S_{j_{k+1}(s)}^{k+1}\subseteq S_{j_k(s)}^k$ for each $k\in\mathbb{Z}_{>0}$ (this is guaranteed by our way of dealing with subsequent ambiguities in the construction of the sequence $(j_k(s))_{k\in\mathbb{Z}}$). Therefore, $\cap_{k=1}^\infty S_{j_k(s)}^k$ is nonempty by Proposition 1.2.39. Moreover, since the squares become arbitrarily small, there exists a unique point in this intersection. Call this point $\gamma(s)$, and this defines γ.

We claim that γ is well-defined in that it is independent of the possible choice made for $j_{k_0}(s)$ when s lies on the boundary of one of the intervals $I_1^{k_0},\dots,I_{4^{k_0}}^{k_0}$ for some $k_0\in\mathbb{Z}_{>0}$. According to our construction, when we encounter the first such case we arbitrarily make a choice between, say, $I_j^{k_0}$ and $I_{j+1}^{k_0}$. Let us compare the consequences of making one choice over the other. In first case we will have $\gamma(s)$ as the unique point in $\cap_{k=k_0}^\infty S_{j_k(s)}^k$ and in the other as the unique point in $\cap_{k=k_0}^\infty S_{j_k(s)+1}^k$. However, since the squares $S_{j_k(s)}^k$ and $S_{j_k(s)+1}^k$ share a common edge for each $k\geq k_0$, and since the common edge for $S_{j_{k+1}(s)}^{k+1}$ (and so also for $S_{j_{k+1}(s)+1}^{k+1}$) is contained in that for $S_{j_k(s)}^k$ (and so also in $S_{j_k(s)+1}^k$) for each $k\geq k_0$, it follows that $\gamma(s)$ is indeed unambiguously defined.

Next we claim that γ is continuous. Let $\epsilon\in\mathbb{R}_{>0}$ and let $s_0\in[0,1]$. Let k be sufficiently large that any square with sides of length 2^{-k+1} containing $\boldsymbol{x}_0\in[0,1]\times[0,1]$ is contained in $\mathsf{B}^2(\epsilon,\boldsymbol{x}_0)$. Take $\delta=4^{-k}$ so that if $|s-s_0|<\delta$ then s and s_0 are either in the same interval or adjacent intervals at the kth step of our sequential construction. Thus $\gamma(s)$ and $\gamma(s_0)$ will lie either in the same square or adjacent squares at the kth step of the sequential construction. Thus $\gamma(s)$ lies in a square with sides of length 2^{-k+1} and so $\gamma(s)\in\mathsf{B}^2(\epsilon,\gamma(s_0))$, giving continuity of γ.

Next we claim that γ is surjective. Let $\boldsymbol{x}_0\in[0,1]\times[0,1]$. By the very construction of γ there then exists a convergent sequence $(s_j)_{j\in\mathbb{Z}_{>0}}$ (with limit denoted by s_0) in $[0,1]$ such that the sequence $(\gamma(s_j))_{j\in\mathbb{Z}_{>0}}$ converges to $\boldsymbol{x}_0$. Since γ is continuous we have $\gamma(s_0) = x_0$ by Theorem 1.3.2.

We have thus far constructed a continuous path γ from $[0,1]$ to $[0,1] \times [0,1]$ that is surjective. With a little modification we shall form a path with all of these properties that is additionally differentiable except on a set of measure zero. Let $f_C\colon [0,1] \to [0,1]$ be Cantor function of Example I-3.2.27. This function, recall (or check), has the following properties:

1. it is continuous;
2. it is differentiable at all points in $[0,1] \setminus C$ since f_C is constant in a neighbourhood of such points;
3. $f_C([0,1] \setminus C) = \{k2^{-j} \mid j, k \in \mathbb{Z}_{>0},\ k < 2^j\}$;
4. $f_C(C) = [0,1]$.

Therefore, the path $\gamma \circ f_C$ has the following properties:

1. it is continuous (being the composition of continuous functions);
2. it is differentiable at all points in $[0,1] \setminus C$ since $\gamma \circ f_C$ is constant in a neighbourhood of such points;
3. $\gamma \circ f_C(C) = [0,1] \times [0,1]$.

Now this is a strange path indeed. At almost all points (i.e., at points in $[0,1] \setminus C$) the path goes nowhere. But at the remaining set of points of zero measure, the path fills out the square $[0,1] \times [0,1]$. •

Paths can be "reversed" and "joined together."

2.2.4 Definition (Inverse of a path, concatenation of two paths) Let $\gamma\colon [a,b] \to \mathbb{R}^n$, $\gamma_1\colon [a_1,b_1] \to \mathbb{R}^n$, and $\gamma_2\colon [a_2,b_2] \to \mathbb{R}^n$ be paths, and suppose that $b_1 = a_2$ and $\gamma_1(b_1) = \gamma_2(a_2)$.

(i) The ***inverse*** of γ is the curve $\gamma^{-1}\colon [a,b] \to \mathbb{R}^n$ defined by

$$\gamma^{-1}(s) = \gamma(b + a - s).$$

(ii) The ***concatenation*** of γ_1 and γ_2 is the curve $\gamma_1 * \gamma_2\colon [a_1,b_2] \to \mathbb{R}^n$ defined by

$$\gamma_1 * \gamma_2(s) = \begin{cases} \gamma_1(s), & s \in [a_1,b_1], \\ \gamma_2(s), & s \in (a_2,b_2]. \end{cases}$$

In Figure 2.2 we depict how one might think of the inverse of a path and the concatenation of two paths.

It is interesting to consider paths with special properties. Let us give some such properties that often arise in applications. In order to do this we first define a very simple class of paths that make use of the structure of $\mathbb{R}^n$.

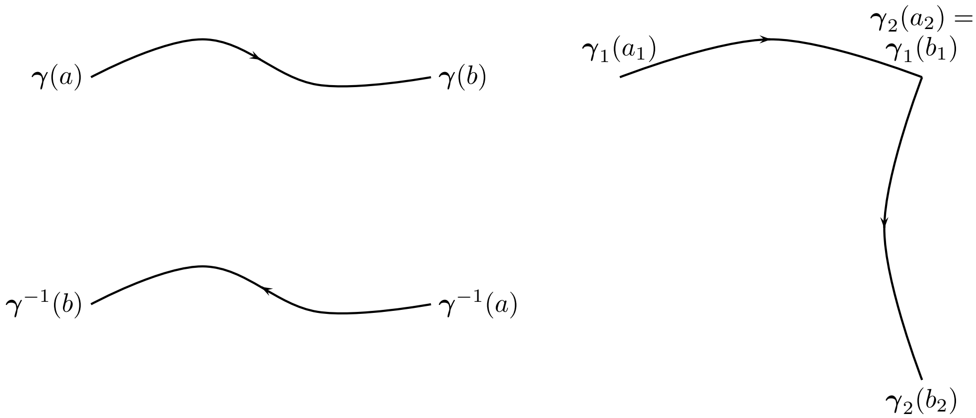

Figure 2.2 A depiction of the inverse of a path (left) and the concatenation of two paths (right)

2.2.5 Definition (Segment, rectangular segment) A ***segment*** in $\mathbb{R}^n$ is a path $\gamma\colon [a, b] \to \mathbb{R}^n$ of the form

$$\gamma(s) = x_0 + sv$$

for $x_0, v \in \mathbb{R}^n$. If $v = e_j$ for some $j \in \{1, \dots, n\}$ (here $\{e_1, \dots, e_n\}$ is the standard basis for $\mathbb{R}^n$) then the segment is ***rectangular***. •

Now we can define our special classes of paths.

2.2.6 Definition (Simple path, closed path, polygonal path, rectangular path) Let $\gamma\colon [a, b] \to \mathbb{R}^n$ be a path.

(i) The path γ is ***simple*** if $\gamma|(a, b)$ is injective.

(ii) The path γ is ***closed*** if $\gamma(a) = \gamma(b)$.

(iii) The path γ is ***polygonal*** if it is a concatenation of a finite number of segments.

(iv) The path γ is ***rectangular*** if it is a concatenation of a finite number of rectangular segments. •

In Figure 2.3 we depict how one might think of these various types of paths.

One of the reasons why special paths are of interest is that one can use them to approximate more general paths. Let us state some results along these lines.

2.2.7 Theorem (Continuous paths can be approximated by rectangular paths) *For a path $\gamma\colon [\mathrm{a}, \mathrm{b}] \to \mathbb{R}^{\mathrm{n}}$ and for $\epsilon \in \mathbb{R}_{>0}$ there exists a rectangular path $\gamma_{\mathrm{rec}}\colon [\mathrm{a}, \mathrm{b}] \to \mathbb{R}^{\mathrm{n}}$ with the property that $\|\gamma(\mathrm{s}) - \gamma_{\mathrm{rec}}(\mathrm{s})\|_{\mathbb{R}^{\mathrm{n}}} < \epsilon$ for every $\mathrm{s} \in [\mathrm{a}, \mathrm{b}]$.*

Proof First note that γ is uniformly continuous by the Heine–Cantor Theorem. Thus, for $\epsilon \in \mathbb{R}_{>0}$ there exists $\delta \in \mathbb{R}_{>0}$ such that if $|s_1 - s_2| < \delta$ then $\|\gamma(s_1) - \gamma(s_2)\|_{\mathbb{R}^n} < \frac{\epsilon}{2}$. Now consider $s_0, s_1, s_k \in [a, b]$ such that

1. $s_0 = a$,
2. $s_k = b$,

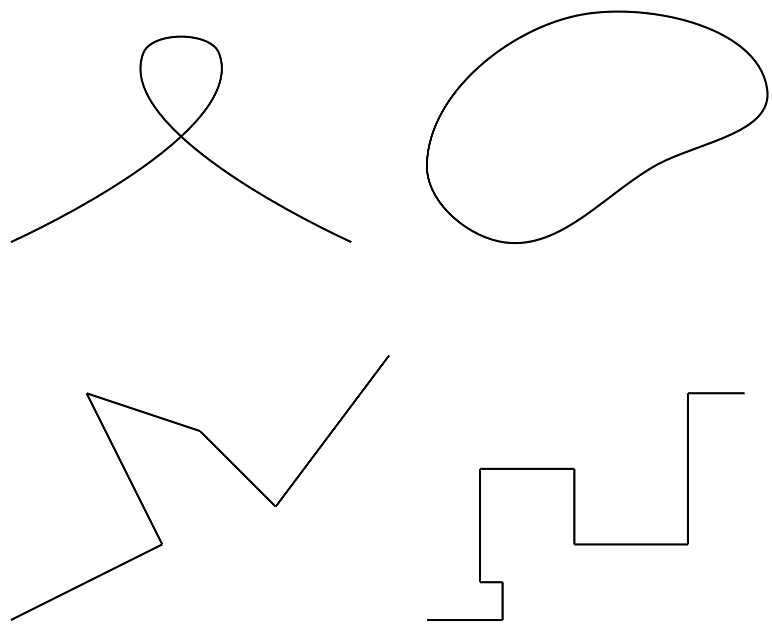

Figure 2.3 A non-simple path (top left), a closed path (top right), a polygonal path (bottom left), and a rectangular path (bottom right)

3. $s_j > s_{j-1}$ for $j \in \{1, \dots, k\}$, and
4. $s_j - s_{j-1} < \delta$ for $j \in \{1, \dots, k\}$.

Denote $I_j = [s_{j-1}, s_j]$, $j \in \{1, \dots, k\}$. Then, for each $j \in \{1, \dots, k\}$ there exists $\boldsymbol{x}_j \in \mathbb{R}^n$ such that $\boldsymbol{\gamma}(I_j) \subseteq \mathsf{B}^n(\frac{\epsilon}{2}, \boldsymbol{x}_j)$. Note that if $\boldsymbol{x}, \boldsymbol{x}' \in \mathsf{B}^n(\frac{\epsilon}{2}, \boldsymbol{x}_j)$ we have

$$\|\boldsymbol{x} - \boldsymbol{x}'\|_{\mathbb{R}^n} \leq \|\boldsymbol{x} - \boldsymbol{x}_j\|_{\mathbb{R}^n} + \|\boldsymbol{x}' - \boldsymbol{x}_j\|_{\mathbb{R}^n} < \epsilon,$$

using the triangle inequality. We now use a simple lemma that captures the essence of what is going on.

1 **Lemma** *For* $\mathrm{r} \in \mathbb{R}_{>0}$, $\mathbf{x}_0 \in \mathbb{R}^{\mathrm{n}}$, *and for* $\mathbf{x}, \mathbf{x}' \in \mathsf{B}^{\mathrm{n}}(\mathrm{r}, \mathbf{x}_0)$ *there exists a rectangular path* $\boldsymbol{\sigma}\colon [\mathrm{a}, \mathrm{b}] \to \mathsf{B}^{\mathrm{n}}(\mathrm{r}, \mathbf{x}_0)$ *connecting* $\mathbf{x}$ *and* $\mathbf{x}'$.

Proof Let us write

$$\boldsymbol{x} = (x_1, \dots, x_n), \quad \boldsymbol{x}' = (x'_1, \dots, x'_n), \quad \boldsymbol{x}_0 = (x_{0,1}, \dots, x_{0,n}).$$

Let us define $\boldsymbol{\gamma}_1\colon [0, 1] \to \mathbb{R}^n$ by

$$\boldsymbol{\gamma}_1(s) = ((1 - s)x_1 + sx_{0,1}, x_2, \dots, x_n)$$

and define $\boldsymbol{x}_1 = (x_{0,1}, x_2, \dots, x_n) = \boldsymbol{\gamma}_1(1)$. Note that $\|\boldsymbol{x}_1 - \boldsymbol{x}_0\|_{\mathbb{R}^n} \leq \|\boldsymbol{x} - \boldsymbol{x}_0\|_{\mathbb{R}^n}$ and so $\boldsymbol{x}_1 \in \mathsf{B}^n(r, \boldsymbol{x}_0)$. Then define $\boldsymbol{\gamma}_2\colon [0, 1] \to \mathbb{R}^n$ by

$$\boldsymbol{\gamma}_2(s) = (x_{0,1}, (1 - s)x_2 + sx_{0,2}, x_3, \dots, x_n)$$

and take $x_2 = (x_{0,1}, x_{0,2}, x_3, \dots, x_n) = \gamma_2(1)$. We have $\|x_2 - x_0\|_{\mathbb{R}^n} \leq \|x - x_0\|_{\mathbb{R}^n}$ giving $x_2 \in \mathsf{B}^n(r, x_0)$. Doing this repeatedly gives rectangular segments $\gamma_j\colon [0,1] \to \mathbb{R}^n$, $j \in \{1, \dots, n\}$, having the following properties:

1. $\gamma_1(0) = x$;
2. $\gamma_j(1) = \gamma_{j+1}(0) \in \mathsf{B}^n(r, x_0)$ for $j \in \{1, \dots, n-1\}$;
3. $\gamma_n(1) = x_0$.

Thus the rectangular path $\gamma = \gamma_1 * \dots * \gamma_n$ connects x and x_0. We claim that this path lives in $\mathsf{B}^n(r, x_0)$. To prove this, we prove something more general. Let $y, y' \in \mathsf{B}^n(r, x_0)$ and let $s \in [0,1]$. Then

$$\|((1-s)y + sy') - x_0\|_{\mathbb{R}^n} \leq (1-s)\|y - x_0\|_{\mathbb{R}^n} + s\|y' - x_0\|_{\mathbb{R}^n} < (1-s)r + sr = r.$$

Thus $(1-s)y + sy' \in \mathsf{B}^n(r, x_0)$. This ensures that each of the paths $\gamma_1, \dots, \gamma_n$ lives in $\mathsf{B}^n(r, x_0)$, and so too, then, does the path γ.

The above process can be repeated to define a rectangular path γ' from x' to x_0 that lies in $\mathsf{B}^n(r, x_0)$. Define $\sigma = \gamma * (\gamma')^{-1}$. The curve σ is defined on the interval $[0, 2n]$, but defines a curve on $[a, b]$ with the same trace, as per Remark 2.2.2. ▼

For $j \in \{1, \dots, k\}$ the lemma gives a rectangular path $\gamma_j\colon I_j \to \mathsf{B}^n(\frac{\epsilon}{2}, x_j)$ connecting $\gamma(s_{j-1})$ with $\gamma(s_j)$. Since $\gamma(s) \in \mathsf{B}^n(\frac{\epsilon}{2}, x_j)$ for each $s \in I_j$ it follows that γ_j is a rectangular path such that $\|\gamma_j(s) - \gamma(s)\|_{\mathbb{R}^n} < \epsilon$ for every $s \in I_j$. Taking $\gamma_{\text{rec}} = \gamma_1 * \dots * \gamma_k$ gives the theorem. ■

Next we consider approximating polygonal paths with "nicer" paths.

2.2.8 Theorem (Polygonal paths can be approximated by differentiable paths) *For a polygonal path* $\gamma\colon [\mathrm{a}, \mathrm{b}] \to \mathbb{R}^\mathrm{n}$ *and for* $\epsilon \in \mathbb{R}_{>0}$ *there exists a differentiable path* $\gamma_{\text{diff}}\colon [\mathrm{a}, \mathrm{b}] \to \mathbb{R}^\mathrm{n}$ *with the property that* $\|\gamma(\mathrm{s}) - \gamma_{\text{diff}}(\mathrm{s})\|_{\mathbb{R}^\mathrm{n}} < \epsilon$ *for every* $\mathrm{s} \in [\mathrm{a}, \mathrm{b}]$.

Proof We consider the case where $\gamma = \gamma_1 * \gamma_2$ is a concatenation of two segments $\gamma_1 * \gamma_2$. We suppose, without loss of generality by Remark 2.2.2, that γ_1 is defined on $[0,1]$ and that γ_2 is defined in $[1,2]$. Since γ is differentiable at all points in $[0,1) \cup (1,2]$, we need only alter γ in a suitable way in a neighbourhood of $1 \in [0,2]$. Let us suppose that $\gamma_1(1) = x_0$ so that

$$\gamma_1(s) = (1-s)x_1 + sx_0, \quad \gamma_2(s) = (s-1)x_2 + (2-s)x_0$$

for some $x_1, x_2 \in \mathbb{R}^n$.

Let us first consider the relatively uninteresting case when $x_1 - x_0$ is collinear with $x_2 - x_0$, i.e., when the vectors $x_1 - x_0$ and $x_2 - x_0$ lie in the same subspace. This uninteresting case admits an even more uninteresting subcase when one or both of $x_1 - x_0$ and $x_2 - x_0$ are zero. In case they are both zero it is immediate that $\gamma(s) = x_0$ for all $s \in [0,2]$ and so is differentiable.

Let us suppose that $x_1 - x_0 \neq \mathbf{0}$ but that $x_2 - x_0 = \mathbf{0}$. Let $\epsilon' \in \mathbb{R}_{>0}$ be such that $\|\gamma_1(1-\epsilon') - x_0\|_{\mathbb{R}^n} < \epsilon$. Then define

$$\begin{aligned} a_0 &= -\tfrac{1}{4\epsilon'}((1-\epsilon')^2 x_0 + (1+\epsilon')^2 x_1), \\ a_1 &= \tfrac{1}{2\epsilon'}(1+\epsilon')(x_0 - x_1), \\ a_2 &= \tfrac{1}{4\epsilon'}(x_1 - x_0). \end{aligned}$$

One can directly and laboriously check that if $\tilde{\gamma}\colon [1-\epsilon', 1+\epsilon'] \to \mathbb{R}^n$ is defined by

$$\tilde{\gamma}(s) = \boldsymbol{a}_0 + \boldsymbol{a}_1 s + \boldsymbol{a}_2 s^2$$

then

1. $\tilde{\gamma}(1-\epsilon') = \gamma_1(1-\epsilon')$,
2. $\tilde{\gamma}(1+\epsilon') = \boldsymbol{x}_0$,
3. $\boldsymbol{D}\tilde{\gamma}(1-\epsilon') = \boldsymbol{D}\gamma_1(1-\epsilon')$, and
4. $\boldsymbol{D}\tilde{\gamma}(1+\epsilon') = \boldsymbol{0}$

(indeed, $\boldsymbol{a}_0$, $\boldsymbol{a}_1$, and $\boldsymbol{a}_2$ were defined by solving these four equations). Thus $\tilde{\gamma}$ is a differentiable path for which the left limit and the left limit of the derivative agree with γ_1 and the right limit and the right limit of the derivative agree with γ_0. Thus the concatenated path $(\gamma_1|[0, 1-\epsilon']) * \tilde{\gamma} * (\gamma_2|[1+\epsilon', 2])$ is differentiable and connects $\gamma_1(0)$ with $\gamma_2(2)$. We need only check that for $s \in [1-\epsilon', 1+\epsilon']$ it holds that $\tilde{\gamma}(s)$ is within ϵ of $\boldsymbol{x}_0$. To see this we compute

$$\boldsymbol{D}\tilde{\gamma}(s) = \tfrac{1}{2\epsilon'}(1+\epsilon'-s)(\boldsymbol{x}_0 - \boldsymbol{x}_1).$$

Since $(1+\epsilon'-s) > 0$ for $s \in [1-\epsilon', 1+\epsilon')$ it follows that $\tilde{\gamma}(s)$ lies in the line segment

$$\{(1-\sigma)\gamma_1(1-\epsilon') + \sigma\boldsymbol{x}_0 \mid \sigma \in [0,1]\}$$

between $\gamma_1(1-\epsilon')$ and $\boldsymbol{x}_0$. By our definition of ϵ' it follows that $\tilde{\gamma}(s)$ is within ϵ of $\boldsymbol{x}_0$, as desired.

The case when $\boldsymbol{x}_1 - \boldsymbol{x} = \boldsymbol{0}$ but $\boldsymbol{x}_2 - \boldsymbol{x}_0 \neq 0$ is handled similarly.

If both $\boldsymbol{x}_1 - \boldsymbol{x}_0$ and $\boldsymbol{x}_2 - \boldsymbol{x}_0$ are nonzero but collinear then we proceed as follows. Let $\epsilon' \in \mathbb{R}_{>0}$ be such that

$$\|\gamma_1(1-\epsilon') - \boldsymbol{x}_0\|_{\mathbb{R}^n}, \quad \|\gamma_2(1+\epsilon') - \boldsymbol{x}_0\|_{\mathbb{R}^n} < \tfrac{\epsilon}{2}.$$

Then define $\tilde{\gamma}\colon [1-\epsilon', 1+\epsilon'] \to \mathbb{R}^n$ by

$$\tilde{\gamma}(s) = \boldsymbol{a}_0 + \boldsymbol{a}_1 s + \boldsymbol{a}_2 s^2,$$

where

$$\begin{aligned}
\boldsymbol{a}_0 &= , \tfrac{1}{4\epsilon'}((1+\epsilon')^2\boldsymbol{x}_1 + (1-\epsilon')^2\boldsymbol{x}_2 - 2(1-\epsilon')^2\boldsymbol{x}_0),\\
\boldsymbol{a}_1 &= \frac{1}{2\epsilon'}((1+\epsilon')\boldsymbol{x}_1 + (1-\epsilon')\boldsymbol{x}_2 - 2\boldsymbol{x}_0),\\
\boldsymbol{a}_2 &= \tfrac{1}{4\epsilon'}(\boldsymbol{x}_1 + \boldsymbol{x}_2 - 2\boldsymbol{x}_0).
\end{aligned}$$

With this definition one directly computes that

1. $\tilde{\gamma}(1-\epsilon') = \gamma_1(1-\epsilon')$,
2. $\tilde{\gamma}(1+\epsilon') = \gamma_2(1+\epsilon')$,
3. $\boldsymbol{D}\tilde{\gamma}(1-\epsilon') = \boldsymbol{D}\gamma_1(1-\epsilon')$, and
4. $\boldsymbol{D}\tilde{\gamma}(1+\epsilon') = \boldsymbol{D}\gamma_2(1+\epsilon')$.

Thus $(\gamma_1|[0,1-\epsilon']) * \tilde{\gamma} * (\gamma_2|[1+\epsilon',2])$ is differentiable and connects $\gamma_1(0)$ with $\gamma_2(2)$. We next claim that $\|\tilde{\gamma}(s) - x_0\|_{\mathbb{R}^n} < \frac{\epsilon}{2}$ for $s \in [1-\epsilon', 1+\epsilon']$. To see this we note that $\boldsymbol{D}\tilde{\gamma}(s)$ is a linear function of s. Thus, as the path $\tilde{\gamma}$ moves along the line connecting $\gamma_1(1-\epsilon')$ to $\gamma(1+\epsilon')$ it changes direction at most once. This is enough to ensure that the distance of $\tilde{\gamma}(s)$ from x_0 is never more than $\frac{\epsilon}{2}$. Since $\gamma_1(s)$, $s \in [1-\epsilon', 1]$, and $\gamma_2(s)$, $s \in [1, 1+\epsilon']$, are also within distance $\frac{\epsilon}{2}$ from x_0, it follows that $\tilde{\gamma}(s)$ is within ϵ of $\gamma(s)$ for all $s \in [1-\epsilon', 1+\epsilon']$, as desired.

The final case to consider is the most interesting one where $x_1 - x_0$ and $x_2 - x_0$ are not collinear. Let $\mathsf{V} = \mathrm{span}_{\mathbb{R}}(x_1 - x_0, x_2 - x_0)$ and note that γ takes its values in the two-dimensional plane

$$\mathsf{P} = \{x_0 + v \mid v \in \mathsf{V}\}$$

that passes through x_0. Choose $\epsilon' \in \mathbb{R}_{>0}$ sufficiently small that

$$\epsilon'\Big(\|x_1 - x_0\|_{\mathbb{R}^n} + \|x_2 - x_0\|_{\mathbb{R}^n}\Big) < \tfrac{\epsilon}{2}.$$

Then define $\tilde{\gamma}\colon [1-\epsilon', 1+\epsilon'] \to \mathbb{R}^n$ by

$$\tilde{\gamma}(s) = x_0 + \tfrac{1}{4\epsilon'}(1-s+\epsilon')^2(x_1 - x_0) + (1-s-\epsilon')^2)(x_2 - x_0),$$

noting that this is a curve in P. One can directly check that

1. $\tilde{\gamma}(1-\epsilon') = \gamma_1(1-\epsilon')$,
2. $\tilde{\gamma}(1+\epsilon') = \gamma_2(1+\epsilon')$,
3. $\boldsymbol{D}\tilde{\gamma}(1-\epsilon') = \boldsymbol{D}\gamma_1(1-\epsilon')$, and
4. $\boldsymbol{D}\tilde{\gamma}(1+\epsilon') = \boldsymbol{D}\gamma_2(1+\epsilon')$.

Thus $(\gamma_1|[0,1-\epsilon']) * \tilde{\gamma} * (\gamma_2|[1+\epsilon',2])$ is differentiable and connects $\gamma_1(0)$ with $\gamma_2(2)$. Next we check that $\|\tilde{\gamma}(s) - x_0\|_{\mathbb{R}^n} < \frac{\epsilon}{2}$ for $s \in [1-\epsilon', 1+\epsilon']$. For such s we have

$$(1-s+\epsilon')^2, (1-s-\epsilon')^2 < 4(\epsilon')^2.$$

Therefore, using the triangle inequality,

$$\|\tilde{\gamma}(s) - x_0\|_{\mathbb{R}^n} \leq \epsilon'\Big(\|x_1 - x_0\|_{\mathbb{R}^n} + \|x_2 - x_0\|_{\mathbb{R}^n}\Big) < \tfrac{\epsilon}{2}$$

by our choice of ϵ'. Thus, for $s \in [1-\epsilon', 1+\epsilon']$, both $\tilde{\gamma}(s)$ and $\gamma(s)$ are with distance $\frac{\epsilon}{2}$ of x_0, and so within distance ϵ of one another, as desired. ■

While the proof of the preceding theorem is lengthy, this is merely because the proof is explicit. The basic idea is very simple, however, and is illustrated in Figure 2.4.

As a final topic in our discussion of paths, let us consider the notion of paths that can be deformed into one another.

2.2.2 Integration along a path

One of the most important

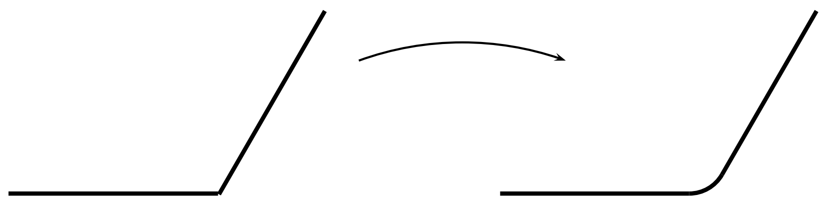

Figure 2.4 Replacing a polygonal path with a differentiable path

2.2.3 Potential functions

2.2.4 Notes

That there exists a continuous path onto a region in $\mathbb{R}^2$ with a nonempty interior was first proved by Peano in 1890. The construction we give in Example 2.2.3 is attributed by Gelbaum and Olmsted [1964] to Hilbert.

ref?
ref?

Section 2.3

Surface integrals

2.3.1 Gaussian regions

2.3.2 Integration on hypersurfaces in $\mathbb{R}^n$

2.3.3 Stokes' Theorem for hypersurfaces

Section 2.4

Stokes' Theorem

This version: 2022/03/07

Chapter 3

Complex analysis

The preceding chapter dealt with (possibly $\mathbb{C}$-valued) functions of a real variable. As we shall see, these sorts of functions naturally arise in our setup as functions of a real time or frequency variable. It is not so evident why one would be interested in studying functions of a *complex* variable. Indeed, the reasons for the appearance of such functions in the theory of signals and systems is somewhat deep, and, at least initially, a little mysterious. Nonetheless, an understanding of at least basic ideas concerning complex function theory will arise in any somewhat complete treatment of signals and systems. And, as we shall see in Chapter III-7, certain elements of not-so-basic complex function theory arise as well. In this chapter we shall provide a self-contained treatment of those ideas from the basic theory of complex functions that will be normally found in a good first course on the subject. Of course, we cannot be as thorough as would be a text for such a course, and thus the reader is invited to look into other suitable books and texts. Good textbooks include [Gamelin 2001, Lang 2003, Needham 1997]. The book by Needham is particularly interesting, treating complex analysis from a geometric perspective. Classic references on complex analysis include [Cartan 1963, Knopp 1996, Titchmarsh 1939]. An excellent introduction to those topics in complex variable theory that are useful in control may be found, complete with all details, in Appendix A of [Seron, Braslavsky, and Goodwin 1997].

Contents

Section 3.1

The geometry of the complex plane

In Chapter I-2 we have already used the notion of a complex number, and so we assume the reader has had some experience at least with arithmetic. In this section we discuss in a systematic way some of the properties of the complex plane that we will use in our treatment of complex function theory.

3.1.1 Complex arithmetic and other simple things

The ***complex plane*** $\mathbb{C}$ is the set of ordered pairs (x, y) of complex numbers. We will follow the usual convention of writing $(x, y) = x + \mathrm{i}y$ where $x, y \in \mathbb{R}$ and where $\mathrm{i} = \sqrt{-1}$. Note that we do not use the symbol j for $\sqrt{-1}$. Only electrical engineers, and those under their influence, use this notation. Complex numbers of the form $x + \mathrm{i}0$ for $x \in \mathbb{R}$ are ***real*** and complex numbers of the form $z = 0 + \mathrm{i}y$ for $y \in \mathbb{R}$ are called ***imaginary***. For $z = x + \mathrm{i}y$, $x, y \in \mathbb{R}$, we denote $\mathrm{Re}(z) = x$ and $\mathrm{Im}(z) = y$ as the ***real part*** and ***imaginary part***, respectively, of z. The complex number $0 + \mathrm{i}0$ will often simply be denoted 0. We will assume the reader knows how to add and multiply complex numbers:

$$(x_1 + \mathrm{i}y_1) + (x_2 + \mathrm{i}y_2) = (x_1 + x_2) + \mathrm{i}(y_1 + y_2)$$
$$(x_1 + \mathrm{i}y_1)(x_2 + \mathrm{i}y_2) = (x_1x_2 - y_1y_2) + \mathrm{i}(x_2y_1 + x_1y_2).$$

We note that this notion of complex arithmetic satisfies the same properties as that for real arithmetic, allowing us to be comfortable with these operations. That is to say, complex arithmetic satisfies

1. $z_1(z_2z_3) = (z_1z_2)z_3$,
2. $z_1z_2 = z_2z_1$,
3. $z_1(z_2 + z_3) = z_1z_2 + z_1z_3$.

Furthermore, $z \in \mathbb{C} \setminus \{0\}$ then it is possible to define $z^{-1} \in \mathbb{C}$ as the unique complex number satisfying the equation $(z^{-1})z = 1$. Explicitly, if $z = x + \mathrm{i}y$ then

$$z^{-1} = \frac{x}{x^2 + y^2} - \mathrm{i}\frac{y}{x^2 + y^2}.$$

The ***complex conjugate*** of $z = x + \mathrm{i}y$ is the complex number $\bar{z} = x - \mathrm{i}y$. By $|x + \mathrm{i}y| = \sqrt{x^2 + y^2}$ we denote the ***modulus*** of z. If z is real, then $|z|$ is the usual absolute value. By $\mathrm{Arg}(x + \mathrm{i}y) = \mathrm{atan2}(x, y)$ we denote the ***argument*** of z. Here $\mathrm{atan2}\colon \mathbb{R}^2 \to (-\pi, \pi]$ is the intelligent arctangent that knows which quadrant one is in. This is illustrated in Figure 3.1.

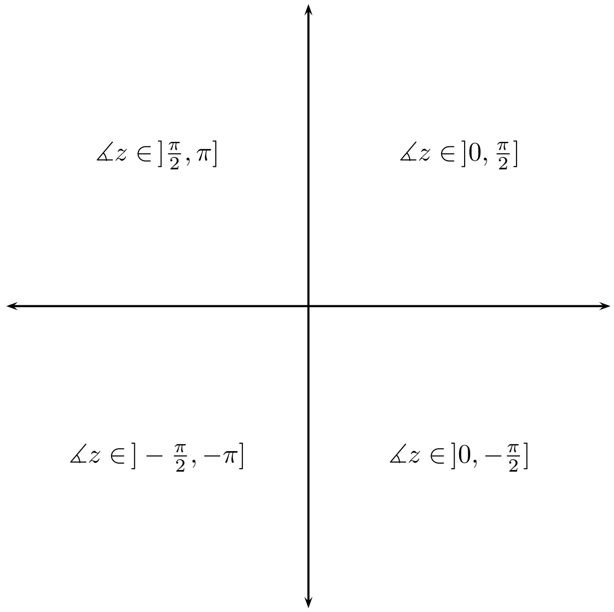

Figure 3.1 The values for the argument of a complex number

It is often useful to represent a complex number $z = x + \mathrm{i}y$ in an alternate form. To do so we define $r = |z|$ and $\theta = \mathrm{Arg}(z)$, then write

$$z = r(\cos\theta + \mathrm{i}\sin\theta).$$

As we shall see in Section 3.2.4, an alternate representation for $\cos\theta + \mathrm{i}\sin\theta$ is $\mathrm{e}^{\mathrm{i}\theta}$. In this case, the expression $z = r\mathrm{e}^{\mathrm{i}\theta}$ is called the ***polar representation*** for z. One of the reasons why the polar representation is so useful is that complex multiplication becomes quite simple: $(r_1\mathrm{e}^{\mathrm{i}\theta_1})(r_2\mathrm{e}^{\mathrm{i}\theta_2}) = r_1 r_2 \mathrm{e}^{\mathrm{i}(\theta_1+\theta_2)}$. Note that when adding the angles $\theta_1 + \theta_2$ one may need to add or subtract 2π to ensure that the result lies in $(-\pi, \pi]$. For instance, if $\theta_1 = \frac{\pi}{2}$ and $\theta_2 = \pi$ then we take $\theta_1 + \theta_2 = -\frac{\pi}{2}(= \frac{3\pi}{2} - 2\pi)$.

3.1.1 Proposition (Existence of roots of unity)

3.1.2 Curves in $\mathbb{C}$

To give all the desired properties of subsets of $\mathbb{C}$, we need some simple concepts concerning curves in $\mathbb{C}$. Furthermore, the notion of a curve will play an important rôle in Cauchy's Theorem, which is one of the most powerful results in complex analysis.

3.1.2 Definition

(i) A ***curve*** is a map $\gamma\colon I \to \mathbb{C}$ from an interval $I \subseteq \mathbb{R}$ that is continuous. That is, the real and imaginary parts of γ are continuous functions.

(ii) A continuous curve $\gamma\colon [a, b] \to \mathbb{C}$ is ***closed*** if $\gamma(a) = \gamma(b)$.

(iii) A ***line*** in $\mathbb{C}$ is a continuous curve $\gamma\colon I \to \mathbb{C}$ of the form $\gamma(t) = (x_0 + \alpha t, y_0 + \beta t)$ for some $\alpha, \beta, x_0, y_0 \in \mathbb{R}$.

(iv) A ***polygonal path*** is a continuous curve $\gamma\colon I \to \mathbb{C}$ with the property that there exists a finite collection $I_1, \dots, I_k$ of subintervals of I with the properties

  (a) $I_i \cap I_j = \emptyset$ for $i, j \in \{1, \dots, k\}$,

  (b) $I = \cup_{j=1}^k I_j$,

  (c) $\gamma|I_j$ is a line for each $j \in \{1, \dots, k\}$.

(v) A continuous curve $\gamma\colon I \to \mathbb{C}$ is ***simple*** if it is injective.

(vi) A ***Jordan curve*** is a continuous simple curve.

(vii) The ***trace*** of a curve $\gamma\colon I \to \mathbb{C}$ is the subset image(γ) of $\mathbb{C}$.

(viii) Two curves $\gamma_1, \gamma_2\colon [a, b] \to \mathbb{C}$ are ***homotopic*** if there exists a continuous map $\Gamma\colon [0, 1] \times [a, b] \to \mathbb{C}$ with the properties that $\Gamma(0, t) = \gamma_1(t)$ and $\Gamma(1, t) = \gamma_2(t)$ for all $t \in [a, b]$. The map Γ is called a ***homotopy map*** from γ_1 to γ_2. •

Note that if one has a continuous curve $\gamma\colon [a, b] \to \mathbb{C}$, one can reparameterise this to give a curve defined on $[0, 1]$ by

$$\tilde{\gamma}(t) = \gamma\Big(\frac{t-a}{b-a}\Big).$$

Thus any curve defined on a compact interval can be thought of as being defined on $[0, 1]$. The trace of the curve remains the same, and all that changes is the speed at which one traverses the curve. The various sorts of curves are illustrated in Figure 3.2.

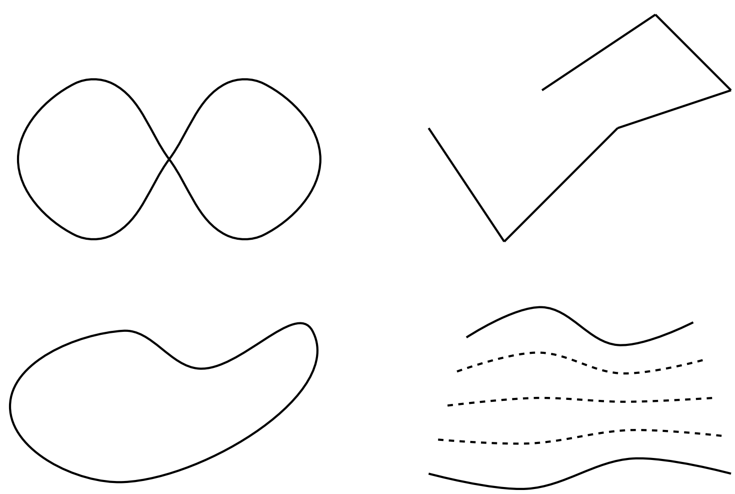

Figure 3.2 Classes of curves in $\mathbb{C}$: closed (top left), polygonal (top right), Jordan (bottom left), and homotopic curves (bottom right)

An important notion associated with a closed curve $\gamma\colon [a, b] \to \mathbb{C}$ and a point z_0 in the complex plane is that of winding number. Suppose that $z_0 \notin \text{image}(\gamma)$. To define the winding number we let $\theta\colon [a, b] \to \mathbb{R}$ be a continuous function with the property that $\gamma(t) = |\gamma(t) - z_0|e^{i\theta(t)}$. Note that this may require allowing θ to take values outside $(-\pi, \pi]$ in order to ensure continuity. We then define the ***winding number*** of γ with respect to z_0 to be $W(\gamma, z_0) = \frac{1}{2\pi}(\theta(b) - \theta(a))$. Note that $W(\gamma, z_0)$ will be an integer since $\gamma(a) = \gamma(b)$, which implies that $\theta(b) - \theta(a) = 2\pi n$ for some $n \in \mathbb{Z}$. Intuitively, $W(\gamma, z_0)$ is the number of times the closed curve γ encircles z_0. A useful means of computing the winding number in examples is the following procedure.

Computation of winding number Let $\gamma\colon [a, b] \to \mathbb{C}$ be a closed curve and let $z_0 \in \mathbb{C} \setminus \text{image}(\gamma)$.

1. Construct a ray ρ emanating from z_0 going to infinity, and assume that ρ intersects image(γ) a finite number of times. Often one can alter ρ to ensure that this is the case.
2. Initialise $W(\gamma, z_0) = 0$.
3. For each counterclockwise intersection of image(γ) with ρ, add 1 to $W(\gamma, z_0)$.
4. For each clockwise intersection of image(γ) with ρ, subtract 1 from $W(\gamma, z_0)$.
5. For intersections of image(γ) with ρ where image(γ) locally lies on the same side of ρ, add 0 to $W(\gamma, z_0)$. •

The reader can apply this procedure in some examples in Exercise 3.5.1.

The following observation concerning the winding number is useful.

3.1.3 Theorem *Let $\gamma\colon [a, b] \to \mathbb{C}$ be a closed curve, and let $\xi\colon I \to \mathbb{C}$ be a curve whose trace does not intersect that of γ. Then $t \mapsto W(\gamma, \xi(t))$ is constant.*

Proof We shall first show that the proposition may be reduced to the case of a closed polygonal path. To do so requires some technical results. First we note that γ may be reparameterised so as to be defined on $[0, 2\pi]$. Then we note that any continuous closed curve in $\mathbb{C}$ defined on $[0, 2\pi]$ can be regarded as a continuous map from

$$\mathbb{S}^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$$

into $\mathbb{C}$. For the remainder of the proof we therefore consider a continuous map $\gamma\colon \mathbb{S}^1 \to \mathbb{C}$.

First we show that compact sets are always separated by a strictly positive distance. To make sense of this, given sets $S_1, S_2 \subseteq \mathbb{C}$ let us define

$$d(S_1, S_2) = \inf\{|z_1 - z_2| \mid z_1 \in S_1,\ z_2 \in S_2\}$$

to be the distance between the sets S_1 and S_2.

1 Lemma *If* $K_1, K_2 \subseteq \mathbb{C}$ *are disjoint compact sets then* $d(K_1, K_2) > 0$.

Proof Let us show that if $S \subseteq \mathbb{C}$ then $z \mapsto d(\{z\}, S)$ is continuous. In fact it holds that $|d(\{z_1\}, S) - d(\{z_2\}, S)| \leq d(\{z_1\}, \{z_2\})$ for $z_1, z_2 \in \mathbb{C}$, and from this the result follows. Let $\epsilon \in \mathbb{R}_{>0}$ and choose $s \in S$ so that $d(\{z_1\}, A) \leq d(\{z_1\}, \{s\}) \leq d(\{z_1\}, A) + \epsilon$. By definition we have $d(\{z_2\}, S) \leq d(\{z_2\}, \{s\})$ and by the triangle inequality we have $d(\{z_2\}, \{s\}) \leq d(\{z_2\}, \{z_1\}) + d(\{z_1\}, \{s\})$. Therefore

$$\begin{aligned} & d(\{z_2\}, S) \leq d(\{z_2\}, \{z_1\}) + d(\{z_1\}, S) + \epsilon \\ \Longrightarrow \quad & d(\{z_2\}, S) - d(\{z_1\}, S) \leq d(\{z_2\}, \{z_1\}) + \epsilon. \end{aligned}$$

Since this is true for every $\epsilon \mathbb{R}_{>0}$ we infer that $d(\{z_2\}, S) - d(\{z_1\}, S) \leq d(\{z_2\}, \{z_1\})$. Symmetry of d gives $|d(\{z_1\}, S) - d(\{z_2\}, S)| \leq d(\{z_1\}, \{z_2\})$, so showing that $z \mapsto d(\{z\}, S)$ is continuous.

In particular, $z \mapsto d(\{z\}, K_1)$ is a bounded continuous function on K_2, and therefore attains its minimum on K_2. This minimum is nonzero since K_1 and K_2 are disjoint. ▼ This seems to not care if K_1 is compact

We now construct a closed polygonal path $\gamma' \colon \mathbb{S}^1 \to \mathbb{C}$ with the property that for each $t \in [0, 1]$ we have $W(\gamma', \xi(t)) = W(\gamma, \xi(t))$. Let $\epsilon = d(\text{image}(\gamma), \text{image}(\xi))$, noting that $\epsilon_1 \in \mathbb{R}_{>0}$ by Lemma 1. The let $N \in \mathbb{Z}_{>0}$ have the property that if $\mathbb{S}^1$ is divided into N arcs of equal length, the image of each arc under γ lies in a disk of radius $\frac{\epsilon}{2}$, this being possible since γ is uniformly continuous by the generalisation of the Heine–Cantor Theorem to higher dimensions. Let $A_1, \ldots, A_N \subseteq \mathbb{S}^1$ denote the equal length arcs on $\mathbb{S}^1$. If we replace $\gamma(S_j)$ with a line connected the images of the endpoints of S_j under γ, this defines a continuous closed polygonal path γ' with N edges, and it is clear that for each $t \in [0, 1]$ we have $W(\gamma', \xi(t)) = W(\gamma, \xi(t))$.

We now prove the theorem for the polygonal path γ'. To do so, we refer ahead to the formula

$$W(\gamma', \xi(t)) = \frac{1}{2\pi \mathrm{i}} \int_{\text{image}(\gamma')} \frac{1}{z - \xi(t)} \, \mathrm{d}z$$

that we give in Proposition 3.6.2. This formula is valid as long as γ' is piecewise differentiable, as it is when it is polygonal. Furthermore, in this case we see that the integrand is bounded uniformly in t since $\text{image}(\gamma')$ and $\text{image}(\xi)$ are disjoint and compact. Therefore, we can infer continuity of the integral from continuity of the integrand from, for example, Theorem III-2.9.16. Since $W(\gamma', \xi(t))$ is integer-valued and continuous, it must be constant. ■

3.1.3 Classification of subsets of $\mathbb{C}$

Let us make a few definitions concerning the nature of subsets of $\mathbb{C}$. First we denote by $\mathsf{B}(r, z)$ the ***open disk*** of radius r centered at z:

$$\mathsf{B}(r, z) = \{z' \in \mathbb{C} \mid |z' - z| < r\}.$$

The ***closed disk*** of radius r centered at z is denoted:

$$\overline{\mathsf{B}}(r, z) = \{z' \in \mathbb{C} \mid |z' - z| \leq r\}.$$

Now we have the following.

3.1.4 Definition (Subsets of $\mathbb{C}$) Let $S \subseteq \mathbb{C}$.

(i) S is ***open*** if for any $z \in S$ there exists an $\epsilon\mathbb{R}_{>0}$ so that $\mathsf{B}(\epsilon, z) \subseteq S$.

(ii) S is ***closed*** if its complement, $\mathbb{C} \setminus S$, is open.

(iii) S is ***bounded*** if there exists $R \in \mathbb{R}_{>0}$ so that $S \subseteq \mathsf{B}(R, 0)$.

(iv) S is compact if it is closed and bounded.

(v) A ***limit point*** for S is a point $z \in \mathbb{C}$ so that for every $\epsilon \in \mathbb{R}_{>0}$ the sets $\mathsf{B}(\epsilon, z) \cap S$ and $\mathsf{B}(\epsilon, z) \cap (\mathbb{C} \setminus S)$ are both nonempty.

(vi) The ***boundary*** of S is the set of all limit points for S and is denoted $\mathrm{bd}(S)$.

(vii) The ***closure*** of S is $S \cup \mathrm{bd}(S)$ and is denoted $\mathrm{cl}(S)$.

(viii) The ***interior*** of S is $\mathrm{cl}(S) \setminus \mathrm{bd}(S)$ and is denoted $\mathrm{int}(S)$.

(ix) An open set $S \subseteq \mathbb{C}$ is ***connected*** if given $z_1, z_2 \in S$ there exists a continuous curve $\gamma\colon [0,1] \to S$ with the property that $\gamma(0) = z_1$ and $\gamma(1) = z_2$.

(x) An open set S that is not connected is ***disconnected***.

(xi) S is ***simply connected*** if for every closed curve $\gamma\colon [0,1] \to S$ there exists a point $z_0 \in S$ and a homotopy map $\Gamma\colon [0,1] \times [0,1] \to S$ from γ to the trivial curve $t \mapsto z_0$.

(xii) An open set S that is not simply connected is ***multiply connected***.

(xiii) S is a ***domain*** if it is open and connected.

(xiv) S is a ***region*** if it is a domain together with a possibly empty subset of its boundary. •

A fundamental result is the following seemingly obvious statement. The proof we give here follows Tverberg [1980]. The proof is quite elementary (meaning it relies on no big machinery), but is nontrivial. The reader interested in understanding the proof would be well-served to make sketches to go along with the various parts of the proof to see why they are true.

3.1.5 Theorem (Jordan Curve Theorem) *If $\gamma\colon [0,1] \to \mathbb{C}$ is a closed continuous Jordan curve then there exists connected open sets $\mathrm{S}_{\mathrm{in}}, \mathrm{S}_{\mathrm{out}} \subseteq \mathbb{C}$ with the following properties:*

(i) S_{in} is bounded;

(ii) $\mathrm{S}_{\mathrm{out}}$ is not bounded;

(iii) $\mathrm{bd}(\mathrm{S}_{\mathrm{in}})$ and $\mathrm{bd}(\mathrm{S}_{\mathrm{out}})$ are both given by the trace of γ.

Furthermore,

(iv) if $\mathrm{z}_0 \in \mathrm{S}_{\mathrm{in}}$ then $|\mathrm{W}(\gamma, \mathrm{z}_0)| = 1$ and

(v) if $\mathrm{z}_0 \in \mathrm{S}_{\mathrm{out}}$ then $\mathrm{W}(\gamma, \mathrm{z}_0) = 0$.

Proof As in the proof of Theorem 3.1.3, we suppose that γ is defined on $\mathbb{S}^1$. We first show that the theorem holds for Jordan polygonal paths, called ***Jordan polygons*** for short.

1 Lemma *If $\gamma\colon \mathbb{S}^1 \to \mathbb{C}$ is a Jordan polygon then the conclusions of the theorem hold for γ.*

Proof The trace of γ is a closed nonintersecting polygon. Let us denote by $E_1, \dots, E_n$ the edges of the polygon and $v_1, \dots, v_n$ the vertices. We assume that the edges are numbered so that E_j and E_{j+1} are adjacent, $j \in \{1, \dots, n\}$ if we take $E_{n+1} = E_1$. We also assume that the vertices are numbered so that v_j is the common endpoint for E_j and E_{j+1}, $j \in \{1, \dots, n\}$. We let δ be the shortest distance between two nonadjacent edges:

$$\delta = \inf\{d(E_j, E_k) \mid E_j \text{ and } E_k \text{ are not adjacent}\}.$$

By Lemma 1 we have $\delta \in \mathbb{R}_{>0}$. Now fix $j \in \{1, \dots, n\}$. Place at a point in $z \in E_j$ a disk $\mathsf{B}(\frac{\delta}{2}, z)$ of radius $\frac{\delta}{2}$. It is clear that $\mathsf{B}(\frac{\delta}{2}, z)$ will intersect at most two edges. One may easily see that image(γ) separates $\mathsf{B}(\frac{\delta}{2}, z)$ into two connected components, say U'_z and U''_z. The edge E_j can be covered by a finite number $\mathsf{B}(\frac{\delta}{2}, z_1), \dots, \mathsf{B}(\frac{\delta}{2}, z_k)$ of such disks, and they may be chosen so that they overlap, and so that the components $U'_{z_1}, \dots, U'_{z_k}$ overlap and so that the components $U''_{z_1}, \dots, U''_{z_k}$ overlap. Note that $\cup_{j=1}^k U'_{z_j}$ and $\cup_{j=1}^k U''_{z_j}$ are connected. One can do this for all edges, so producing (at most) two connected sets U' and U'' (it is possible that $U' = U''$). We now claim that every point in $\mathbb{C} \setminus \text{image}(\gamma)$ can be joined to U' or U'' be a continuous curve. To see this, let $z \in \mathbb{C} \setminus \text{image}(\gamma)$ and draw a line from z that intersects image(γ). Let $z_0 \in \text{image}(\gamma)$ be the first point this line encounters. Prior to intersecting image(γ) this line will pass through at least one of the components U' or U''. Thus we have shown that $\mathbb{C} \setminus \text{image}(\gamma)$ consists of at most two connected components.

Now we show that $\mathbb{C} \setminus \text{image}(\gamma)$ has at least two connected components. We note that one may rotate a Jordan polygon in the complex plane, and that the result will be another Jordan polygon, and that if the theorem holds for a Jordan polygon, it also holds for any rotation of that Jordan polygon. Therefore, without loss of generality (by a suitable rotation, if necessary) we may suppose that image(γ) has the property that the real parts of all vertices are distinct. We then define a map $\sigma\colon \mathbb{C} \setminus \text{image}(\gamma) \to \{-1, 1\}$ as follows:

1. draw a vertical ray ρ_z starting at z and going upwards;
2. if ρ_z intersects image(γ) an even number of times, define $\sigma(z) = 1$;
3. if ρ_z intersects image(γ) an odd number of times, define $\sigma(z) = -1$;
4. if ρ_z passes through a vertex v_j, $j \in \{1, \dots, n\}$, of image(γ), this does not count as an intersection if v_{j-1} and v_{j+1} lie on the same side of ρ_z.

We claim that for each $z \in \mathbb{C} \setminus \text{image}(\gamma)$ there exists $\epsilon \in \mathbb{R}_{>0}$ to that if $\sigma(z') = \sigma(z)$ provided that $|z - z'| < \epsilon$. This is evident if $\text{Re}(z) \neq \text{Re}(v_j)$, $j \in \{1, \dots, n\}$ then $\sigma(z) = \sigma(z')$ for z' sufficiently near z. If $\text{Re}(z) = \text{Re}(v_j)$ for some $j \in \{1, \dots, n\}$, then one of the following two circumstances can arise.

1. $\text{Im}(z) > \text{Im}(v_j)$: In this case $\sigma(z) = \sigma(z')$ for z' sufficiently near z.
2. $\text{Re}(v_{j-1}) < \text{Re}(z) < \text{Re}(v_{j+1})$: In this case $\sigma(z) = \sigma(z')$ for z' sufficiently near z.
3. $\text{Re}(v_{j+1}) < \text{Re}(z) < \text{Re}(v_{j-1})$: In this case $\sigma(z) = \sigma(z')$ for z' sufficiently near z.
4. $\text{Im}(z) < \text{Im}(v_j)$, and v_{j-1} and v_{j+1} lie on the same side of v: In this case the number of intersections of $\rho_{z'}$ with image(γ) will be the same as that for ρ_z, or will be two great than that for ρ_z.

In both case $\sigma(z') = \sigma(z)$ provided that z' is sufficiently close to z. This shows that $\sigma^{-1}(1)$ and $\sigma^{-1}(-1)$ are both open.

We also claim that $\sigma^{-1}(1)$ and $\sigma^{-1}(-1)$ are both nonempty. It is clear that $\sigma^{-1}(1)$ is nonempty since there are points in $\mathbb{C}$ through which the vertical line will not intersect image(γ) (since image(γ) is compact). Now let $y_{\max}$ be the largest imaginary value attained by a point on image(γ). If there is a single vertex with $y_{\max}$ as imaginary part then it is easy to see that there is a point z so that ρ_z intersects image(γ) only once. This is also easily seen (even more so) when there are multiple vertices with $y_{\max}$ as imaginary part.

Finally, we show that any continuous curve in $\xi\colon [0,1] \to \mathbb{C} \setminus \text{image}(\gamma)$ with $\sigma(\xi(0)) = 1$ has the property that $\sigma(\xi(1)) = 1$. Indeed, suppose that there exists $t \in [0,1]$ for which $\sigma(\xi(t)) = -1$, and let t_0 be the infimum over all such points in $[0,1]$. Since $\sigma^{-1}(-1)$ is open, this means that there is a neighbourhood U of $\xi(t_0)$ where $\sigma(z) = -1$ for $z \in U$. By continuity of γ this means that there is a point $t_0' < t_0$ for which $\sigma(\xi(t_0')) = -1$, and this contradicts the definition of t_0.

Thus we have shown that there are at least two disjoint connected sets $\sigma^{-1}(1)$ and $\sigma^{-1}(-1)$ in $\mathbb{C} \setminus \text{image}(\gamma)$. Combined with the first part of the proof, this gives the sets $S_{\text{in}} = \sigma^{-1}(-1)$ and $S_{\text{out}} = \sigma^{-1}(1)$ in the statement of the theorem (it is clear by construction in the first part of the proof that S_{in} and S_{out} have image(γ) as boundary).

The assertions concerning the winding numbers are demonstrated as follows. We first claim that $W(\gamma, z) = 0$ for $z \in S_{\text{out}}$. This follows from Theorem 3.1.3 along with the fact that the winding number is zero for *some* point in S_{out}. To see this, note that if we choose a point z_0 sufficiently far away from image(γ) we can be sure that the continuous function θ defined by $\gamma(t) = |\gamma(t) - z_0|e^{i\theta(t)}$ has as small an image as desired. By definition of winding number, if image(θ) is contained in an interval with length smaller than 2π, the winding number is zero. To see that $W(\gamma, z_0) = \pm 1$ for $z_0 \in S_{\text{in}}$, again by Theorem 3.1.3 it suffices to show that for *some* $z_0 \in S_{\text{in}}$. We choose a point z_0 by asking that it be possible to extend a ray ρ from z_0 to infinity that intersects image(γ) only once. That this is possible was argued above when showing that $\sigma^{-1}(-1)$ is nonempty. We claim that $W(\gamma, z_0) = 1$ if γ intersects ρ going counterclockwise, and $W(\gamma, z_0) = -1$ otherwise. Let us deal with just the counterclockwise case, the clockwise following in the same manner. To see that our assertion holds, consider the ray ρ_- emanating from z_0 and going to infinity in the direction opposite to that of ρ. Consider the number $\alpha(\gamma, z_0)$ defined as follows. For a crossing of ρ_- by γ in the counterclockwise direction, add π, and for a crossing of ρ_- by γ in the counterclockwise direction, subtract π. The winding number is then readily deduced to be $\frac{1}{2\pi}(\alpha(\gamma, z_0) + \pi)$. Note that by definition of z_0 and ρ and since γ is closed, for every clockwise intersection of ρ by γ, there must be one counterclockwise intersection (the polygon must close at z_0). Furthermore, there cannot be more than one counterclockwise intersection that is not matched with a clockwise intersection (this would result in the curve not closing). Finally, there must be at least one counterclockwise intersection in order that the curve close. This then gives $W(\gamma, z_0) = 1$, as claimed. ▼

The next step is to approximate a Jordan curve with a Jordan polygon. We first establish the continuity of the "inverse" of γ.

need Heine-Borel for Rn

2 Lemma *If $\gamma\colon \mathbb{S}^1 \to \mathbb{C}$ is continuous, then for every $\epsilon \in \mathbb{R}_{>0}$ there exists $\delta \in \mathbb{R}_{>0}$ so that if*

$|\gamma(\mathrm{x}_1, \mathrm{y}_1) - \gamma(\mathrm{x}_2, \mathrm{y}_2)| < \delta$ *then* $\|(\mathrm{x}_1, \mathrm{y}_1) - (\mathrm{x}_2, \mathrm{y}_2)\| < \epsilon$.

Proof We shall show that the image of every closed set in $\mathbb{S}^1$ is closed in $\mathbb{C}$. Note that a closed subset $S \subseteq \mathbb{S}^1$ is compact since it is clearly bounded. Now let $(U_a)_{a\in A}$ be a family of open sets covering image(S). The sets $(\gamma^{-1}(U_a))_{a\in A}$ cover S. Since S is compact there is a finite collection $\gamma^{-1}(U_{a_1}), \dots, \gamma^{-1}(U_{a_k})$ of these subsets that cover S. Note then that $U_{a_1}, \dots, U_{a_k}$ then cover $\gamma(S)$ since γ is a bijection. This shows that image(S) is compact. We claim that this implies the closedness of image(S), or equivalently, openness of $\mathbb{C}\setminus$image(S). Let $z_0 \in \mathbb{C}\setminus$image$(S)$. For $z \in$ image(S) let U_z be a neighbourhood of z_0 and V_z be a neighbourhood of z with the property that $U_z \cap V_z = \varnothing$. Note that image$(S) \subseteq \cup_{z\in\mathrm{image}(S)} V_z$. Therefore, since image$(S)$ is compact, there exists a finite collection $z_1, \dots, z_k$ of points in image(S) so that *image*$(S) \subseteq \cup_{j=1}^k V_{z_j}$. Note that the neighbourhood $\cap_{j=1}^k U_{z_k}$ of z_0 does not intersect image(S), thus showing that $\mathbb{C}\setminus$image(S) is open, so image(S) is closed. The result now follows by the generalisation of Theorem I-3.1.3 to functions between Euclidean spaces of dimension greater than one (such a generalisation is performed by a mere change of notation in the proof of Theorem I-3.1.3). ▼

We can now show that every Jordan curve can be approximated arbitrarily well by a Jordan polygon.

3 **Lemma** *Let* $\gamma\colon \mathbb{S}^1 \to \mathbb{C}$ *be a Jordan curve and let* $\epsilon \in \mathbb{R}_{>0}$. *Then there exists a Jordan polygon* $\gamma'\colon \mathbb{S}^1 \to \mathbb{C}$ *with the property that* $|\gamma(\mathrm{x}, \mathrm{y}) - \gamma'(\mathrm{x}, \mathrm{y})| < \epsilon$ *for each* $(\mathrm{x}, \mathrm{y}) \in \mathbb{S}^1$.

Proof Let $\epsilon \in \mathbb{R}_{>0}$ and choose $\delta_1 \in \mathbb{R}_{>0}$ so that if $\|(x_1, y_1) - (x_2, y_2)\| < \delta_1$ gives $|\gamma(x_1, y_1) - \gamma(x_2, y_2)| < \epsilon$. This is possible since γ is uniformly continuous by the Heine–Cantor Theorem. By Lemma 2 let $\delta_2 \in \mathbb{R}_{>0}$ be chosen so that if $|\gamma(x_1, y_1) - \gamma(x_2, y_2)| < \delta_2$ then $\|(x_1, y_1) - (x_2, y_2)\| < \min\{\delta_1, \sqrt{3}\}$. The define $\delta = \min\{\frac{\epsilon}{2}, \delta_2\}$.

Now place on $\mathbb{C}$ a grid of squares whose sides have length δ. Since image(γ) is compact it will intersect only a finite number of these squares, and denote these by $S_1, \dots, S_k$. By definition of δ it holds that $\gamma^{-1}(S_j)$ is contained in an arc on $\mathbb{S}^1$ of length less than $\frac{2\pi}{3}$ (this is where the $\sqrt{3}$ comes in). Let A_j denote the minimal arc containing $\gamma^{-1}(S_j)$. Let $\gamma_0 = \gamma$ and define $\gamma_1(t) = \gamma_0(t)$ for $t \notin A_1$, and define γ_1 on A_1 so that its image is a line, and so that γ_1 is continuous. Note that $\gamma_1^{-1}(S_j) \subseteq \gamma_1^{-1}(S_j)$ for $j \geq 2$. If $\gamma_1^{-1}(S_2) = \varnothing$ then define $\gamma_2 = \gamma_1$. Otherwise, repeat the process above to arrive at γ_2. Repeating this process we arrive at a Jordan polygon γ_k.

It remains to show that γ_k gives the desired approximation of γ. Let $(x, y) \in \mathbb{S}^1$ have the property that $\gamma(x, y) \neq \gamma_k(x, y)$. Then there exists some $j \in \{1, \dots, k\}$ so that $\gamma_k(x, y) = \gamma_j(x, y) \neq \gamma_{j-1}(x, y)$. Our construction ensures that (x, y) lies in an arc A_j on $\mathbb{S}^1$ whose endpoints (x_0, y_0) and (x_1, y_1) satisfy $\gamma_j(x_0, y_0) = \gamma(x_1, y_1)$ and $\gamma_j(x_1, y_1) = \gamma(x_1, y_1)$. Therefore,

$$\begin{aligned}|\gamma(x, y) - \gamma_k(x, y)| &= |\gamma(x, y) - \gamma(x_0, y_0) + \gamma_j(x_0, y_0) - \gamma_j(x, y)| \\ &\leq |\gamma(x, y) - \gamma(x_0, y_0)| + \delta \\ &\leq |\gamma(x, y) - \gamma(x_0, y_0)| + \tfrac{\epsilon}{2}.\end{aligned}$$

Since $\|(x, y) - (x_0, y_0)\| \leq \|(x_1, y_1) - (x_0, y_0)\|$ and since $|\gamma(x_1, y_1) - \gamma(x_0, y_0)| < \frac{\epsilon}{2}$, the result follows. ▼

The idea now if somewhat clear. Given a Jordan curve $\gamma\colon \mathbb{S}^1 \to \mathbb{C}$, one wishes to find a sequence $(\gamma_j)_{j\in\mathbb{Z}_{>0}}$ of Jordan polygons converging to γ (in the sense of Lemma 3), and infer the theorem since it holds for each of the approximating polygons. However, the problem that arises is that the limit of Jordan polygons is not necessarily a Jordan curve. Therefore, one has to ensure that if a sequence of Jordan polygons converge to a *Jordan curve*, then the fact that the theorem holds for each of the approximating polygons implies that the theorem holds for the limit Jordan curve. The following two lemmas have this as their objective.

The first lemma shows the existence of points "far away" from one another in $\mathbb{S}^1$ that map to points "far away" from one another in S_{in}.

4 Lemma *Let $\gamma\colon \mathbb{S}^1 \to \mathbb{C}$ be a Jordan polygon with S_{in} and $\mathrm{S}_{\mathrm{out}}$ the two connected components of $\mathbb{C} \setminus \mathrm{image}(\gamma)$ as in Lemma 1. Then there is a disk $\mathrm{D} \subseteq \mathrm{S}_{\mathrm{in}}$ with the property that there are points $\mathrm{z}_1, \mathrm{z}_2 \in \mathrm{bd}(\mathrm{D})$ for which $\|\gamma^{-1}(\mathrm{z}_1) - \gamma^{-1}(\mathrm{z}_2)\| \geq \sqrt{3}$.*

Proof For a disk $D \subseteq S_{\mathrm{in}}$ let us define

$$\rho(D) = \sup\{\|\gamma^{-1}(z_1) - \gamma^{-1}(z_2)\| \mid z_1, z_2 \in \mathrm{bd}(D)\}.$$

Thought of as a function of the centre and radius of D, ρ is a continuous function by Lemma 2, and since S_{in} is bounded, ρ is a bounded function of radius and centre. Therefore, there exists a disk D for which $\rho(D)$ achieves its maximum. Let $z_1, z_2 \in \mathrm{bd}(D)$ be those points for which $\rho(D) = |z_1 - z_2|$, and suppose that $\|\gamma^{-1}(z_1) - \gamma^{-1}(z_2)\| < \sqrt{3}$. Let A be the longest arc on $\mathbb{S}^1$ joining $\gamma^{-1}(z_1)$ and $\gamma^{-1}(z_2)$, noting that A subtends and angle greater than $\frac{4\pi}{3}$.

We first claim that there are no points in $\mathrm{int}(A)$ whose image under γ intersect D. Indeed, suppose there were such a point (x, y). One easily sees that either $\|\gamma^{-1}(z_1) - (x, y)\|$ or $\|\gamma^{-1}(z_2) - (x, y)\|$ must exceed $\|\gamma^{-1}(z_1) - \gamma^{-1}(z_2)\|$, thereby contradicting the definition of z_1 and z_2.

Now let $v_1, \dots, v_k \in \mathbb{C}$ be the vertices of $\mathrm{image}(\gamma)$ that are images of points in A under γ. Suppose that these are ordered so that they are met as one traverses A from $\gamma^{-1}(z_1)$ to $\gamma^{-1}(z_2)$. We have four cases to consider.

1. First suppose that $v_1 \neq z_1$ and $v_k \neq z_2$. We claim that there exists a disk $D' \subseteq S_{\mathrm{in}}$ whose boundary intersects $\mathrm{image}(\gamma)$ at one point in the edge between z_1 and v_1 and at one point in the edge between v_k and z_2. This follows since $\mathrm{bd}(D)$ is tangent to $\mathrm{image}(\gamma)$ at z_1 and z_2, and at no other place along $\gamma(A)$. The existence of a disk D' that is tangent to $\mathrm{image}(\gamma)$ at points z_1' and z_2' near z_1 and z_2, respectively, follows from continuity of γ. Continuity also ensures that D' does not intersect $\mathrm{image}(\gamma)$ except at z_1' and z_2', thus ensuring that $D' \subseteq S_{\mathrm{in}}$. Now we note that $\|\gamma^{-1}(z_1') - \gamma^{-1}(z_2')\| > \|\gamma^{-1}(z_1) - \gamma^{-1}(z_2)\|$, thus contradicting the definition of z_1 and z_2.
2. Suppose that $v_1 \neq v_1$ and that $v_k = z_2$. Here one argues as in the previous case the existence of a disk D' whose boundary passes through z_2 and which is tangent to $\mathrm{image}(\gamma)$ at a point near z_1 along the edge between z_1 and v_1. One similarly arrives at a contradiction of z_1 and z_2.
3. Suppose that $v_1 = v_1$ and that $v_k \neq z_2$. This case produces a contradiction in a manner similar to the previous case.

4. Finally, suppose that $v_1 = z_1$ and $v_k = z_2$. In this case consider a family of disks $\mathscr{D}$ with the following properties:

   1. if $D' \in \mathscr{D}$ then $z_1, z_2 \in \mathrm{bd}(D')$;

   2. if $D' \in \mathscr{D}$ there exists continuous curves $c\colon [0,1] \to S_{\mathrm{in}}$ and $\rho\colon [0,1] \to \mathbb{R}_+$ for $c(0)$ is the centre of D, $c(1)$ is the centre of D', and the circle of radius $\rho(t)$ and centre $c(t)$ is in S_{in} for $t \in [0,1]$.

   One can see that there will be a curve of circles in $\mathscr{D}$ with the property that either there is a circle in the curve whose boundary meets a point in $\gamma(\mathrm{int}(A))$ or that there is a circle in the curve that becomes tangent to the edges between z_1 and v_1 and v_k and z_2. In the former case we immediately have a contradiction of the definition of z_1 and z_2. In the latter case one can infer the existence of a disk tangent to image(γ) at points close to z_1 and z_2, and that is contained in S_{in}. One then argues as in the first case that this again contradicts the definition of z_1 and z_2. ▼

For the next result, it is convenient to introduce some notation. We are still considering a Jordan polygon $\gamma\colon \mathbb{S}^1 \to \mathbb{C}$. Suppose that z_1 and z_2 both lie in $S \in \{S_{\mathrm{in}}, S_{\mathrm{out}}\}$. If ℓ is a line segment in S that intersects image(γ) only at its endpoints, then this defines two Jordan polygons having ℓ as their common boundary. Therefore ℓ divides S into two connected components by Lemma 1. Let $L(S)$ denote the collection of all segments ℓ such as described above whose length is less than 2.

5 Lemma *With* γ, z_1, z_2, *and* S *as above, suppose that* $\mathrm{d}(\mathrm{image}(\gamma), \{z_1, z_2\}) \geq 1$ *and that for each* $\ell \in \mathrm{L(S)}$, z_1 *and* z_2 *are in the same connected component of* S *defined by* ℓ*. Then there exists a continuous curve* $\xi\colon [0,1] \to \mathrm{S}$ *connecting* z_1 *and* z_2 *and so that* $\mathrm{d}(\mathrm{image}(\gamma), \mathrm{image}(\xi)) \geq 1$.

Proof Note that if z_1' can be connected to z_1 with a curve $\xi'\colon [0,1] \to S$ for which $d(\mathrm{image}(\gamma), \mathrm{image}(\xi')) \geq 1$, then the hypotheses of the lemma hold if z_1 is replaced with z_1' in the statement of the lemma. To see this let $\ell \in L(S)$ and then note that ℓ does not intersect image(ξ'), since if it did this would contradict the assumption that $d(\mathrm{image}(\gamma), \mathrm{image}(\xi')) \geq 1$. Therefore z_1 and z_1' are in the same connected component of S defined by ℓ. Therefore z_1' and z_2 are also in the same connected component of S defined by ℓ, and this is our claim. The claim ensures that we can without loss of generality assume that $d(\{z_1\}, \mathrm{image}(\gamma)) = d(\{z_2\}, \mathrm{image}(\gamma)) = 1$, since if this is not the case we can move z_1 and z_2 via continuous curves to ensure that it is, without changing the outcome of the lemma. Since image(γ) is compact there are points $(x_1, y_1), (x_2, y_2) \in \mathbb{S}^1$ with the property that $\|\gamma(x_1, y_1) - z_1\| = 1$ and $\|\gamma(x_2, y_2) - z_2\| = 1$.

Let D_0 be the unit disk with centre z_1 and touching image(γ) at $\gamma(x_1, y_1)$. For $t \in [0,1]$ let D_t be the unit disk obtained by continuously rolling D_0 along image(γ) until D_1 touches image(γ) at $\gamma(x_2, y_2)$. Let $c_t \in \mathbb{C}$ denote the centre of D_t.

We first claim that $D_t \subseteq S$ for each $t \in [0,1]$. If not, then there is $t_0 \in [0,1]$ so that there is a line segment across D_{t_0} with points in image(γ) as endpoints. Let us denote these endpoints as z_1' and z_2'. This line segment ℓ will necessarily have length less than 2, and so will partition S into two connected components, say S_1 and S_2. Suppose that $c_{t_0} \in S_1$. As we argued at the beginning of the proof, this shows that $z_2 \in S_1$ (take $z_1' = c_{t_0}$ in the first part of the proof). Also, $\gamma(x_2, y_2) \in \mathrm{bd}(S_2)$ since we are assuming that $c_t \neq z_2$ for every $t \in [0, t_0]$. Let D' be the unit disk with centre z_2. The line segment from z_2 to $\gamma(x_2, y_2)$ must therefore intersect ℓ as z_2 and $\gamma(x_2, y_2)$ lie on different sides of ℓ. Thus we arrive at the following argument.

1. We have two unit disks D_t and D' with centres c_t and z_2 on the same side of the line segment ℓ.
2. The line segment ℓ runs across D_t.
3. D' cannot contain the endpoints of ℓ since this would violate the condition $D' \subseteq S$.
4. The line from z_2 to the point $\gamma(x_2, y_2) \in \mathrm{bd}(D')$ intersects ℓ.
5. The point $\gamma(x_2, y_2) \in \mathrm{bd}(D')$ lies on the opposite side of ℓ than z_2.
6. This means that $\mathrm{bd}(D')$ intersects ℓ in two places.
7. The last three statements are contradictory.

Thus we have shown that $D_t \subseteq S$ for each $t \in [0, 1]$.

The preceding argument shows that the point $\gamma(x_2, y_2)$ must lie on the boundary of some disk D_1 constructed as above. It does not necessarily follow that $D_1 = D'$. Thus we need to show that z_2 can be reached from c_1 by a curve that remains a distance at least 1 from image(γ). If image(γ) is tangent to D_1 at $\gamma(x_2, y_2)$ then we have $D' = D_1$. However, if $\gamma(x_2, y_2)$ is a vertex of image(γ) then we have to do some extra work to see that D_1 can be connected to D' as desired. In this case a problem arises when one cannot "pivot" D_1 around $\gamma(x_2, y_2)$ to D'. If this is the case then there exists a unit disk D'' with $\gamma(x_2, y_2)$ as a point on its boundary, and with a line segment ℓ' of length less than two running across it from $\gamma(x_2, y_2)$ to a point $z_1'' \in \mathrm{image}(\gamma)$. If c'' is the centre of D'' then the points c'' and z_2 lie on opposite sides of the line segment ℓ'. However, this is in contradiction with the argument in the first part of the lemma. ▼

Now we proceed more or less directly with the proof, using the above technical results.

6 Lemma *If $\gamma\colon \mathbb{S}^1 \to \mathbb{C}$ is a Jordan curve then $\mathbb{C} \setminus \mathrm{image}(\gamma)$ consists of at least two connected components, one of which is bounded, another of which is unbounded.*

Proof That there is one connected component that is unbounded is clear since image(γ) is compact. We therefore show the existence of a bounded component. Let $D_0 \subseteq \mathbb{C}$ be a closed disk containing image(γ) in its interior. Let $(\gamma_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence of Jordan polygons converging to γ. Thus

$$\lim_{j \to \infty} \sup\{|\gamma(x, y) - \gamma_j(x, y)| \mid (x, y) \in \mathbb{S}^1\} = 0.$$

By Lemma 4 there is a sequence of disks D_j whose boundaries contain points $\gamma(x_{1,j}, y_{1,j})$ and $\gamma(x_{2,j}, y_{2,j})$ where $\|(x_{1,j}, y_{1,j}) - (x_{2,j}, y_{2,j})\| \geq \sqrt{3}$. Let z_j be the centre of the disk D_j for $j \in \mathbb{Z}_{>0}$. Since the sequence $(\gamma_j)_{j \in \mathbb{Z}_{>0}}$ converges to γ we may choose a subsequence $(\gamma_{j_k})_{k \in \mathbb{Z}_{>0}}$ with the property that $\mathrm{image}(\gamma_{j_k}) \subseteq D_0$ for $k \in \mathbb{Z}_{>0}$. Then, since the sequence $\{z_{j_k}\}$ is bounded, by the Bolzano–Weierstrass Theorem we may choose a subsequence $(\gamma_{j_{k_m}})_{m \in \mathbb{Z}_{>0}}$ for which $(z_{j_{k_m}})_{m \in \mathbb{Z}_{>0}}$ converges. Let us relabel the sequence $(\gamma_{j_{k_m}})_{m \in \mathbb{Z}_{>0}}$ as $(\gamma_j)_{j \in \mathbb{Z}_{>0}}$.

Now let $\epsilon \in \mathbb{R}_{>0}$ be chosen so that if $\|(x_1, y_1) - (x_2, y_2)\| \geq \sqrt{3}$ then $|\gamma(x_1, y_1) - \gamma(x_2, y_2)| \geq \epsilon$. Thus $|\gamma(x_{1,j}, y_{1,j}) - \gamma(x_{2,j}, y_{2,j})| \geq \epsilon$, from which we deduce that there exists $N \in \mathbb{Z}_{>0}$ so that $|\gamma_j(x_{1,j}, y_{1,j}) - \gamma_j(x_{2,j}, y_{2,j})| > \frac{\epsilon}{2}$ for $j \geq N$. This implies that the disk D_j has diameter greater than $\frac{\epsilon}{2}$, and therefore that $d(\{z_j\}, \mathrm{image}(\gamma_j)) > \frac{\epsilon}{4}$, provided that $j \geq N$. Thus we have a sequence of disks $(D_j)_{j\in\mathbb{Z}_{>0}}$ of diameter greater than $\frac{\epsilon}{2}$ whose centres $(z_j)_{j\in\mathbb{Z}_{>0}}$ converge to a point z. Also, we note that $\mathrm{int}(D_j)$ lies in the bounded connected component of $\mathbb{C} \setminus \mathrm{image}(\gamma_j)$, and that $(\gamma_j)_{j\in\mathbb{Z}_{>0}}$. Thus the points z_j, $j \geq N$, and z all lie in the same connected component of $\mathbb{C} \setminus \mathrm{image}(\gamma_j)$ and in the same connected component of $\mathbb{C} \setminus \mathrm{image}(\gamma)$.

Next, suppose that the point z constructed above can be connected to a point z_0 lying outside the disk D_0 by a continuous curve $\xi\colon [0,1] \to \mathbb{C} \setminus \mathrm{image}(\gamma)$. Let $\delta = d(\mathrm{image}(\xi), \mathrm{image}(\gamma))$ and let $N \in \mathbb{Z}_{>0}$ have the property that

$$\sup\{|\gamma(x,y) - \gamma_j(x,y)| \mid (x,y) \in \mathbb{S}^1\} < \tfrac{\delta}{2}, \qquad j \geq N.$$

It follows that $d(\mathrm{image}(\xi), \mathrm{image}(\gamma_j)) > \frac{\delta}{2}$ for $j \geq N$. This implies that z and z_j lie in the unbounded connected component of $\mathbb{C} \setminus (\gamma_j)$ for all sufficiently large j, and this is a contradiction. ▼

7 Lemma *If* $\gamma\colon \mathbb{S}^1 \to \mathbb{C}$ *is a Jordan curve then* $\mathbb{C} \setminus \mathrm{image}(\gamma)$ *consists of at most two connected components.*

Proof Suppose that $\mathbb{C} \setminus \mathrm{image}(\gamma)$ possess three distinct connected components S_1, s_2, and S_3, and let $z_a \in S_a$, $a \in \{1,2,3\}$. Denote $\epsilon = d(\{z_1, z_2, z_3\}, \mathrm{image}(\gamma))$. Let $(\gamma_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence of Jordan polygons converging to γ as in the proof of Lemma 6. Choose $N \in \mathbb{Z}_{>0}$ so that $d(\{z_1, z_2, z_3\}, \mathrm{image}(\gamma_j)) \geq \frac{\epsilon}{2}$ for $j \geq N$. Since $\mathbb{C} \setminus \mathrm{image}(\gamma_j)$ consists of two connected components, for each j at least two of z_1, z_2, and z_3 will lie in the same connected component of $\mathbb{C} \setminus \mathrm{image}(\gamma)$. We may choose a subsequence $(\gamma_{j_k})_{k\in\mathbb{Z}_{>0}}$ with the property that for each $k \in \mathbb{Z}_{>0}$ the points z_1 and z_2 lie in the same connected component of $\mathbb{C} \setminus \mathrm{image}(\gamma_{j_k})$. Let us reindex this subsequence to be $(\gamma_j)_{j\in\mathbb{Z}_{>0}}$, and let us denote S_j as the connected component of $\mathbb{C} \setminus \mathrm{image}(\gamma_j)$ in which z_1 and z_2 lie.

We next show that for $\delta \in (0, \epsilon)$ there are at most finitely many $j \in \mathbb{Z}_{>0}$ for which there exists a continuous curve $\xi_j\colon [0,1] \to \mathbb{C} \setminus \mathrm{image}(\gamma_j)$ connecting z_1 to z_2 with the property that $d(\mathrm{image}(\xi_j), \mathrm{image}(\gamma_j)) \geq \delta$. Indeed, if there is an infinite sequence $(j_k)_{k\in\mathbb{Z}_{>0}}$ of such j's, then this would mean that for sufficiently large j we have $d(\mathrm{image}(\xi_{j_k}), \mathrm{image}(\gamma)) > \frac{\delta}{2}$, which then gives a continuous curve in $\mathbb{C} \setminus \mathrm{image}(\gamma)$ connected z_1 and z_2. This contradicts our assumption that z_1 and z_2 lie in distinct connected components of $\mathbb{C} \setminus \mathrm{image}(\gamma)$.

Let $\delta_k = \frac{\epsilon}{2k}$, $k \in \mathbb{Z}_{>0}$. Our preceding argument ensures the existence of $m_k \in \mathbb{Z}_{>0}$ for which there is no continuous curve $\xi_{m_k}\colon [0,1] \to \mathbb{C} \setminus \mathrm{image}(\gamma_{m_k})$ connecting z_1 to z_2 with the property that $d(\mathrm{image}(\xi_{m_k}), \mathrm{image}(\gamma_{m_k})) \geq \delta$. By a scaled version of Lemma 5 this ensures that there is a line segment ℓ_k with the following properties:

1. the interior points of the line $\mathrm{int}(\ell_k)$ lie in S_{m_k};
2. S_{m_k} has length less that $2\delta_k$;
3. the endpoints of ℓ_k touch $\mathrm{image}(\gamma_{m_k})$;
4. z_1 and z_2 lie in different connected components of $S_{m_k} \setminus \ell_k$.

Let $(x_{1,k}, y_{1,k}), (x_{2,k}, y_{2,k}) \in \mathbb{S}^1$, $k \in \mathbb{Z}_{>0}$ have the property that $\gamma_{m_k}(x_{1,k}, y_{1,k})$ and $\gamma_{m_k}(x_{2,k}, y_{2,k})$ are the endpoints of ℓ_k. Since the length of ℓ_k tends to zero as $k \to \infty$, and since γ is continuous, it follows that $\lim_{k\to\infty}\|(x_{1,k}, y_{1,k}) - (x_{2,k}, y_{2,k})\| = 0$. Let $A_k \subseteq \mathbb{S}^1$ be the shortest arc between $(x_{1,k}, y_{1,k})$ and $(x_{2,k}, y_{2,k})$. For infinitely many $k \in \mathbb{Z}_{>0}$ one of z_1 and z_2, let us say z_1 for concreteness, lies in that connected component of $S_{m_k} \setminus \ell_k$ bounded by $\ell_k \cup \mathrm{image}(A_k)$. As $k \to \infty$ the size of this connected component will shrink to zero since the length of ℓ_k shrinks to zero and since the length of A_k shrinks to zero. In particular, this means that $\lim_{k\to\infty}|z_1 - \gamma_{m_k}(x_{1,k}, y_{1,k})| = 0$. However, we began by assuming that z_1 lies a distance $\epsilon \in \mathbb{R}_{>0}$ from $\mathrm{image}(\gamma)$, and so we have achieved a contradiction since $(\gamma_{m_k})_{k\in\mathbb{Z}_{>0}}$ converges to γ. ▼

8 Lemma *Let $\gamma\colon \mathbb{S}^1 \to \mathbb{C}$ be a Jordan curve and let S_{in} and S_{out} be the connected components of $\mathbb{C} \setminus \mathrm{image}(\gamma)$. Then $W(\gamma, z_0) = 0$ for $z_0 \in S_{\mathrm{out}}$ and $|W(\gamma, z_0)| = 1$ for $z_0 \in S_{\mathrm{in}}$.*

Proof This follows immediately from Theorem 3.1.3 since S_{in} and S_{out} are connected. ▼

This completes the proof. ■

This theorem was stated first by Jordan [1887],[1] although Jordan's proof was not satisfactory. It is felt that the first correct proof was due to Veblen [1905]. Since then, many versions of the proof have been presented. Some, like the one we give here, are long but elementary (another such can be found in [Whyburn 1942]), while others, as can be found in texts on algebraic topology [e.g., Munkres 1984], are shorter, but rely on heavy machinery.

get number for OV:05

Exercises

3.1.1 Indicate why the name "triangle inequality" makes sense for the complex magnitude function.

be sure this is referenced

3.1.2

[1]Marie Ennemond Camille Jordan, 1838–1922.

Section 3.2

Functions

3.2.1 $\mathbb{C}$-valued functions

For $I \in \mathscr{I}$ we may talk about functions $f\colon I \to \mathbb{C}$ taking values in $\mathbb{C}$ rather than $\mathbb{R}$. For each $t \in I$ we may write $f(t) = \operatorname{Re}(f)(t) + i\operatorname{Im}(f)(t)$, so defining $\mathbb{R}$-valued functions $\operatorname{Re}(f), \operatorname{Im}(f)\colon I \to \mathbb{R}$ called the ***real part*** and the ***imaginary part*** of f, respectively. Many of the properties of $\mathbb{R}$-valued functions can be defined by ascribing these same properties to the real and imaginary part of the function. Let us summarise this for the various sorts of function properties we described for $\mathbb{R}$-valued functions.

1. A function $f\colon I \to \mathbb{C}$ is ***continuous at*** $\mathbf{t_0}$ if $\operatorname{Re}(f)$ and $\operatorname{Im}(f)$ are both continuous at t_0. One can verify, using the fact the complex magnitude $|\cdot|$ is continuous, that if a function $f\colon I \to \mathbb{C}$ is continuous at t_0 then the $\mathbb{R}$-valued function $t \mapsto |f(t)|$ is continuous at t_0. The converse is clearly false (think of a counterexample).
2. A function $f\colon I \to \mathbb{C}$ is ***continuous*** if $\operatorname{Re}(f)$ and $\operatorname{Im}(f)$ are both continuous. If a function $f\colon I \to \mathbb{C}$ is continuous then the $\mathbb{R}$-valued function $t \mapsto |f(t)|$ is continuous.
3. A function $f\colon I \to \mathbb{C}$ is ***uniformly continuous*** if $\operatorname{Re}(f)$ and $\operatorname{Im}(f)$ are both uniformly continuous. If a function $f\colon I \to \mathbb{C}$ is uniformly continuous then the $\mathbb{R}$-valued function $t \mapsto |f(t)|$ is uniformly continuous.
4. A function $f\colon I \to \mathbb{C}$ is ***bounded*** if there exists $M > 0$ so that $|f(t)| \le M$ for each $t \in I$. It is pretty clear that f is bounded if and only if $\operatorname{Re}(f)$ and $\operatorname{Im}(f)$ are bounded. As with $\mathbb{R}$-valued functions, it is true that a continuous $\mathbb{C}$-valued function defined on a compact interval I is bounded. It is also true that if f is a continuous $\mathbb{C}$-valued function on a compact interval I then $|f|$ attains its maximum and minimum in I.
5. A function $f\colon I \to \mathbb{C}$ is ***differentiable at*** $\mathbf{t_0}$ if both $\operatorname{Re}(f)$ and $\operatorname{Im}(f)$ are differentiable at t_0. The ***derivative*** is defined to be $f'(t_0) = \operatorname{Re}(f)'(t_0) + i\operatorname{Im}(f)'(t_0)$.
6. A function $f\colon I \to \mathbb{C}$ is ***continuously differentiable*** if it is differentiable at each $t \in I$ and if the function $t \mapsto f'(t)$ is continuous.
7. A function $f\colon I \to \mathbb{C}$ is $\mathbf{r}$ ***times continuously differentiable*** if it can be differentiated r times, and if the rth derivative is continuous. The rth derivative is denoted $f^{(r)}$.
8. For $f\colon I \to \mathbb{C}$ and $t_0 \in I$ define
$$f(t_0-) = \lim_{\epsilon \downarrow 0} f(t_0 - \epsilon), \quad f(t_0+) = \lim_{\epsilon \downarrow 0} f(t_0 + \epsilon).$$
9. A function $f\colon [a,b] \to \mathbb{C}$ is ***piecewise continuous*** if there exists a partition $\{t_0, t_1, \dots, t_n\}$ of $[a,b]$ with the properties

(i) f is continuous on each of the intervals (t_j, t_{j-1}) for $j \in \{1, \dots, n\}$;
(ii) for $j \in \{1, \dots, n-1\}$ the limits $f(t_j+)$ and $f(t_j-)$ exist;
(iii) the limits $f(a+)$ and $f(b-)$ exist.

10. Define

$$f'(t_0-) = \lim_{\epsilon \uparrow 0} \frac{f(t_0+\epsilon) - f(t_0)}{\epsilon}, \quad f'(t_0+) = \lim_{\epsilon \downarrow 0} \frac{f(t_0+\epsilon) - f(t_0)}{\epsilon}.$$

11. A piecewise continuous function is ***piecewise differentiable*** if there exists a partition $\{t_0, t_1, \dots, t_n\}$ of $[a, b]$ with the properties
   (i) f is continuously differentiable on each of the intervals (t_j, t_{j-1}) for $j \in \{1, \dots, n\}$;
   (ii) for $j \in \{1, \dots, n-1\}$ the limits $f(t_j+)$, $f(t_j-)$, $f'(t_j+)$, and $f'(t_j-)$, exist;
   (iii) the limits $f(a+)$, $f(b-)$, $f'(a+)$, and $f'(b-)$ exist.

3.2.2 $\mathbb{C}$-valued functions of bounded variation

First let us make the routine extension from the $\mathbb{R}$-valued functions of bounded variation considered in Section I-3.3 to the case of $\mathbb{C}$-valued functions. First we make the routine extension of the definition.

3.2.1 Definition ($\mathbb{C}$-valued function of bounded variation) For $I = [a, b]$ a compact interval and $f: I \to \mathbb{C}$ a $\mathbb{C}$-valued function on I, the ***total variation*** of f is given by

$$\mathrm{TV}(f) = \sup\Big\{ \sum_{j=1}^{k} |f(x_j) - f(x_{j-1})| \;\Big|\; \{x_0, x_1, \dots, x_k\} = \mathrm{EP}(P),\ P \in \mathrm{Part}([a, b]) \Big\}.$$

If $\mathrm{TV}(f) < \infty$ then f has ***bounded variation***. We denote by $\mathsf{BV}([a, b]; \mathbb{C})$ (resp. $\mathsf{BV}([a, b]; \mathbb{R})$) the set of $\mathbb{C}$-valued (resp. $\mathbb{R}$-valued) functions of bounded variation on $[a, b]$. •

Now we show that (what anyone would agree is) the natural alternative definition for a $\mathbb{C}$-valued function of bounded variation is equivalent to the one we give.

3.2.2 Proposition ($\mathbb{C}$-valued functions of bounded variation) *Let* $\mathrm{I} = [\mathrm{a}, \mathrm{b}]$ *be a compact interval. A* $\mathbb{C}$*-valued function* $\mathrm{f}: [\mathrm{a}, \mathrm{b}] \to \mathbb{C}$ *has bounded variation if and only if the* $\mathbb{R}$*-valued functions* $\mathrm{Re(f)}, \mathrm{Im(f)}: [\mathrm{a}, \mathrm{b}] \to \mathbb{R}$ *have bounded variation.*

move this result to a better place?

Proof For $z = x + iy \in \mathbb{C}$ in Corollary III-2.7.50 we established the formulae

$$\sqrt{x^2 + y^2} \le |x| + |y|, \quad |x| + |y| \le \sqrt{2}\sqrt{x^2 + y^2}.$$

Now suppose that $\mathrm{Re}(f)$ and $\mathrm{Im}(f)$ have bounded variation and let $(x_0, x_1, \dots, x_k)$ be the endpoints of a partition of $[a, b]$. Then

$$\begin{aligned}\sum_{j=1}^{k}|f(x_j) - f(x_{j-1})| &= \sum_{j=1}^{k}|\mathrm{Re}(f)(x_j) - \mathrm{Re}(f)(x_{j-1}) + \mathrm{i}(\mathrm{Im}(f)(x_j) - \mathrm{Im}(f)(x_{j-1}))| \\ &\le \sum_{j=1}^{k}\Big(|\mathrm{Re}(f)(x_j) - \mathrm{Re}(f)(x_{j-1})| + |\mathrm{Im}(f)(x_j) - \mathrm{Im}(f)(x_{j-1})|\Big) \\ &\le \Big(\mathrm{TV}(\mathrm{Re}(f)) + \mathrm{TV}(\mathrm{Im}(f))\Big).\end{aligned}$$

Thus $\mathrm{TV}(f) < \mathrm{TV}(\mathrm{Re}(f)) + \mathrm{TV}(\mathrm{Im}(f))$.

Next suppose that $\mathrm{TV}(f) < \infty$. Then, for $(x_0, x_1, \dots, x_k)$ the endpoints of a partition of $[a, b]$, we have

$$\begin{aligned}\sum_{j=1}^{k}|\,\mathrm{Re}(f)(x_j) - \mathrm{Re}(f)(x_{j-1})| \\ &\le \sum_{j=1}^{k}\Big(|\mathrm{Re}(f)(x_j) - \mathrm{Re}(f)(x_{j-1})| + |\mathrm{Im}(f)(x_j) - \mathrm{Im}(f)(x_{j-1})|\Big) \\ &\le \sqrt{2}\sum_{j=1}^{k}|\mathrm{Re}(f)(x_j) - \mathrm{Re}(f)(x_{j-1}) + \mathrm{i}(\mathrm{Im}(f)(x_j) - \mathrm{Im}(f)(x_{j-1}))| \\ &\le \sqrt{2}\,\mathrm{TV}(f).\end{aligned}$$

Thus $\mathrm{TV}(\mathrm{Re}(f)) \le \sqrt{2}\,\mathrm{TV}(f)$, and an entirely similar argument shows that $\mathrm{TV}(\mathrm{Im}(f)) \le \sqrt{2}\,\mathrm{TV}(f)$, so giving the result. ■

3.2.3 Absolutely continuous $\mathbb{C}$-valued functions

In exactly the same manner as we proved Proposition 3.2.2 we prove the following result.

3.2.3 Proposition *A signal* $\mathrm{f}\colon \mathbb{T} \to \mathbb{C}$ *on a compact continuous time-domain is absolutely continuous if and only if the signals* $\mathrm{Re}(\mathrm{f}), \mathrm{Im}(\mathrm{f})\colon \mathbb{T} \to \mathbb{R}$ *are absolutely continuous.*

3.2.4 Elementary functions

3.2.4 Definition (Properties of $\mathbb{C}$- valued functions)

(i)

Let $D \subseteq \mathbb{C}$ be a domain. A function $f\colon D \to \mathbb{C}$ is ***continuous at $\mathbf{z_0}$*** if for every $\delta > 0$ there exists $\epsilon > 0$ so that $|z - z_0| < \epsilon$ implies that $|f(z) - f(z_0)| < \delta$. If f is continuous at every point in D then f is simply ***continuous***. The function f is ***differentiable*** at z_0 if the limit

$$\lim_{z\to z_0}\frac{f(z) - f(z_0)}{z - z_0}$$

exists and is independent of the manner in which the limit is taken.[2] The limit when it so exists is the ***derivative*** and denoted $f'(z_0)$. If we write $z = x + iy$ and $f(z) = u(x, y) + iv(x, y)$ for $\mathbb{R}$-valued functions u and v, then it may be shown that f is differentiable at $z_0 = x_0 + iy_0$ if and only if (1) u and v are differentiable at (x_0, y_0) and (2) the ***Cauchy-Riemann equations*** are satisfied at z_0:

$$\frac{\partial u}{\partial x}(x_0, y_0) = \frac{\partial v}{\partial y}(x_0, y_0), \quad \frac{\partial u}{\partial y}(x_0, y_0) = -\frac{\partial v}{\partial x}(x_0, y_0).$$

A function $f\colon D \to \mathbb{C}$ on a domain D is ***analytic*** at $z_0 \in D$ if there exists $\epsilon > 0$ so that f is differentiable at every point in $D(z_0, \epsilon)$. If $R \subseteq D$ is a region, we say f is analytic in R if it is analytic at each point in R. Note that this may necessitate differentiability of f at points outside R.

Analytic functions may fail to be defined at isolated points. Let us be systematic about characterising such points.

3.2.5 Definition Let $f\colon D \to \mathbb{C}$ be analytic.

(i) A point $z_0 \in D$ is an ***isolated singularity*** for f if there exists $\epsilon > 0$ so that f is defined and analytic on $D(z_0, r) \setminus \{z_0\}$ but is not defined on $D(z_0, r)$.

(ii) An isolated singularity z_0 for f is ***removable*** if there exists an $r > 0$ and an analytic function $g\colon D(z_0, r) \to \mathbb{C}$ so that $g(z) = f(z)$ for $z \neq z_0$.

(iii) An isolated singularity z_0 for f is a ***pole*** if

  (a) $\lim_{z\to z_0}|f(z)| = \infty$ and

  (b) there exists $k > 0$ so that the function g defined by $g(z) = (z - z_0)^k f(z)$ is analytic at z_0. The smallest $k \in \mathbb{Z}$ for which this is true is called the ***order*** of the pole.

(iv) An isolated singularity z_0 for f is ***essential*** if it is neither a pole nor a removable singularity.

(v) A function $f\colon D \to \mathbb{C}$ is ***meromorphic*** if it analytic except possibly at a finite set of poles. •

Another important topic in the theory of complex functions is that of series expansions. Let D be a domain. If $f\colon D \to \mathbb{C}$ is analytic at $z_0 \in D$ then one can show that all derivatives of f exist at z_0. The ***Taylor series*** for f at z_0 is then the series

$$f(z) = \sum_{j=0}^{\infty} a_j (z - z_0)^j.$$

where the coefficients are defined by

$$a_j = \frac{f^{(j)}(z_0)}{j!}.$$

[2] Thus for any sequence $\{z_k\}$ converging to z_0, the sequence $\left\{\frac{f(z_k)-f(z_0)}{z_k-z_0}\right\}$ should converge, and should converge to the same complex number.

Analyticity of f guarantees pointwise convergence of the Taylor series in a closed disk of positive radius. If z_0 is an isolated singularity for f then the Taylor series is not a promising approach to representing the function. However, one can instead use the ***Laurent series*** given by

$$f(z) = \sum_{j=0}^{\infty} a_j(z - z_0)^j + \sum_{j=1}^{\infty} \frac{b_j}{(z - z_0)^j}.$$

The matter of expressing the coefficients in terms of f obviously cannot be done by evaluations of f and its derivatives at z_0. However, there are formulas for the coefficients involving contour integrals. So...

3.2.5 Roots of unity

3.2.6 Proposition *If ζ is a primitive nth root of unity and if $\mathrm{k} \in \mathbb{Z}_{>0}$, then*

$$\sum_{\mathrm{j}=0}^{\mathrm{n}-1} \zeta^{\mathrm{jk}} = \begin{cases} \mathrm{n}, & \mathrm{n}|\mathrm{k}, \\ 0, & \textit{otherwise}. \end{cases}$$

Proof Suppose that $n \nmid k$. Then we can write $k = qn + r$ for $q \in \mathbb{Z}_{\geq 0}$ and $r \in \{1, \dots, n-1\}$. Therefore,

$$\zeta^k = \zeta^{qn+r} = \zeta^{qn}\zeta^r = \zeta^r \neq 1$$

since ζ is a primitive nth root of unity. We have

$$\begin{aligned} 1 + \zeta^k + \zeta^{2k} + \dots + \zeta^{(n-1)k} &= 1 + \zeta^k + \zeta^{2k} + \dots + \zeta^{(n-1)k} \frac{1 - \zeta^k}{1 - \zeta^k} \\ &= \frac{1 - z^{nk}}{1 - z^k} = 0. \end{aligned}$$

If $n|k$ then we have $\zeta^{jk} = 1$ for each $j \in \{0, 1, \dots, N - 1\}$, and so $\sum_{j=0}^{n-1} \zeta^{jk} = n$. ■

Move this theorem

3.2.7 Theorem (Approximation of finitely many points in the unit circle) *Let $\mathrm{a}_1, \dots, \mathrm{a}_\mathrm{k} \in \mathbb{R}$ have the property that $\{2\pi, \mathrm{a}_1, \dots, \mathrm{a}_\mathrm{k}\}$ are linearly independent over $\mathbb{Q}$. Then, for any $\mathrm{z}_1, \dots, \mathrm{z}_\mathrm{k} \in \mathbb{S}^1$ and for any $\epsilon \in \mathbb{R}_{>0}$, there exists $\mathrm{b} \in \mathbb{Z}_{>0}$ such that*

is this the notation I am using?

$$\max\{|\mathrm{e}^{\mathrm{iba}_1} - \mathrm{z}_1|, \dots, |\mathrm{e}^{\mathrm{iba}_\mathrm{k}} - \mathrm{z}_\mathrm{k}|\} < \epsilon.$$

Proof We claim that the map

$$\mathbb{R} \ni \theta \mapsto \mathrm{e}^{\mathrm{i}\theta} \in \mathbb{S}^1 \subseteq \mathbb{C}$$

is uniformly continuous. Indeed, this map is written in its real and imaginary components as

$$\theta \mapsto (\cos\theta, \sin\theta).$$

By Proposition I-3.2.8 and Proposition I-3.8.19 it follows that each of the components is uniformly continuous, and so the uniform continuity of the map follows. Therefore, there exists $\delta \in \mathbb{R}_{>0}$ such that, if $|\theta - \theta'| < \delta$, then $|e^{i\theta} - e^{i\theta'}| < \epsilon$.

For each $j \in \{1, \dots, k\}$, let us write $z_j = e^{i\theta_j}$ for some $\theta_j \in \mathbb{R}$. By the Kronecker Approximation Theorem, Theorem I-2.2.20, there exists $b \in \mathbb{Z}_{>0}$ and $m_1, \dots, m_k \in \mathbb{Z}$ such that

$$|ba_j - \theta_j - 2\pi m_j| < \delta, \qquad j \in \{1, \dots, k\}.$$

By our observation from the first paragraph of the proof it follows that

$$|e^{iba_j} - e^{i(\theta_j + 2\pi m_j)}| = |e^{iba_j} - e^{i\theta_j}| = |e^{iba_j} - z_j| < \epsilon$$

for each $j \in \{1, \dots, k\}$, as desired. ■

move this

3.2.8 Theorem (Annuli admitting a biholomorphic mapping) *Let* $r_1, r_2, R_1, R_2 \in \mathbb{R}_{>0}$ *satisfy* $r_1 < R_1$ *and* $r_2 < R_2$ *and consider two annuli in* $\mathbb{C}$*:*

$$\mathbb{A}(r_1, R_1) = \{z \in \mathbb{C} \mid |z| \in (r_1, R_1)\}, \qquad \mathbb{A}(r_2, R_2) = \{z \in \mathbb{C} \mid |z| \in (r_2, R_2)\}.$$

Then the following statements are equivalent:

(i) there exists a biholomorphic bijection $\Phi\colon \mathbb{A}(r_1, R_1) \to \mathbb{A}(r_2, R_2)$*;*

(ii) $\frac{r_1}{R_1} = \frac{r_2}{R_2}$.

Proof If $\frac{r_1}{R_1} = \frac{r_2}{R_2}$, then the mapping

$$z \mapsto \frac{r_2}{r_1} z$$

gives the desired biholomorphic bijection when restricted to $\mathbb{A}(r_1, R_1)$. So we prove the converse assertion.

We can, without loss of generality, assume that $r_1 = r_2 = 1$. Let us thus abbreviate

$$\mathbb{A}_1 = \mathbb{A}(1, R_1), \qquad \mathbb{A}_2 = \mathbb{A}(1, R_2).$$

Let $\Phi\colon \mathbb{A}_1 \to \mathbb{A}_2$ be a biholomorphic bijection and let $r \in (1, R_2)$. Denote

$$C = \{z \in \mathbb{C} \mid |z| = r\}.$$

Continuity of Φ^{-1} ensures that $\Phi^{-1}(C)$ is compact (Proposition 1.3.29). Since $r > 1$, for $z \in \mathrm{bd}(\mathbb{D}10)$ there exists $\delta_z \in \mathbb{R}_{>0}$ such that $\mathbb{D}\delta_z z \cap \Phi^{-1}(C) = \varnothing$. Since $\mathrm{bd}(\mathbb{D}10)$ is compact, there exists $z_1, \dots, z_k \in \mathrm{bd}(\mathbb{D}10)$ such that

$$\mathrm{bd}(\mathbb{D}10) \subseteq \cup_{j=1}^k \mathbb{D}\delta_{z_j} z_j$$

and

$$C \cap (\cup_{j=1}^k \mathbb{D}\delta_{z_j} z_j) = \varnothing.$$

Thus there exists $\delta \in \mathbb{R}_{>0}$ such that $\mathbb{A}(1, 1 + \delta) \cap C = \varnothing$. Let $V = \Phi(\mathbb{A}(1, 1 + \delta))$. Since $\mathbb{A}(1, 1 + \delta)$ us connected, V is connected by Proposition 1.3.34. Since $V \cap C = \varnothing$, this means that either $V \subseteq \mathbb{A}(1,r)$ or $V \subseteq \mathbb{A}(r,R_2)$. Note that $z \mapsto \frac{R_2}{\Phi(z)}$ is a biholomorphic bijection of $\mathbb{A}_1$ with $\mathbb{A}_2$. Thus, we can assume without loss of generality (by replacing Φ with $\frac{R_2}{\Phi}$ if necessary) that $V \subseteq \mathbb{A}(1,r)$.

We claim that, if $(z_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence in $\mathbb{A}_1$ satisfying $\lim_{j\to\infty}|z_j| = 1$, then $\lim_{j\to\infty}|\Phi(z_j)| = 1$. We can assume, with δ defined as above, that $z_j \in \mathbb{A}(1, 1+r)$ for every $z \in \mathbb{Z}_{>0}$. Note that $(\Phi(z_j))_{j\in\mathbb{Z}_{>0}}$ does not have a limit point in $\mathbb{A}_2$. Indeed, if it did, then continuity of Φ ensures that $(z_j)_{j\in\mathbb{Z}_{>0}}$ has a limit point in $\mathbb{A}_1$ (by Theorem 1.3.2) which is contradicted by the fact that $\lim_{j\to\infty}|z_j| = 1$. The sequence $(|\Phi(z_j)|)_{j\in\mathbb{Z}_{>0}}$ converges by Theorem 1.3.2 and continuity of Φ and $|\cdot|$. Therefore, we either have $\lim_{j\to\infty}|\Phi(z_j)| = 1$ or $\lim_{j\to\infty}|\Phi(z_j)| = R_2$. By our assumption that $V \subseteq \mathbb{A}(1,r)$, it must be the former of these limits that is valid.

One similarly shows that, if $(z_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence in $\mathbb{A}_1$ satisfying $\lim_{j\to\infty}|z_j| = R_1$, then $\lim_{j\to\infty}|\Phi(z_j)| = R_2$.

Now let $\alpha = \frac{\log(R_2)}{\log(R_1)}$ and define

$$\Psi(z) = 2(\log(|\Phi(z)|) - \alpha\log(|z|).$$

By Example 3.8.4–1, Ψ is harmonic on $\mathbb{A}_1$ since Φ is nonzero on $\mathbb{A}_1$. By the conclusions of the preceding two paragraphs, Ψ extends to a function on $\mathrm{cl}(\mathbb{A}_1)$ that vanishes on $\mathrm{bd}(\mathbb{A}_1)$. Therefore, by , Ψ is zero on $\mathbb{A}_1$. In particular, what

$$0 = \Psi'(z) = \frac{\Phi'(z)}{z} - \frac{\alpha}{z}. \tag{3.1}$$

Let $\rho \in (1, R_1)$ and let Γ_ρ be the contour

$$\Gamma_\rho = \{\rho e^{i\theta} \mid \theta \in [0, 2\pi]\}.$$

By Cauchy's Integral Theorem we have get the reference right

$$\alpha = \frac{1}{2\pi i}\int_{\Gamma_\rho} \frac{\alpha}{z}\,dz = \frac{1}{2\pi i}\int_{\Gamma_\rho} \frac{\Phi'(z)}{\Phi(z)}\,dz,$$

from which we conclude that $\alpha \in \mathbb{Z}$. Now compute

$$\frac{d}{dz}(z^{-\alpha}\Phi(z)) = z^{-\alpha-1}(-\alpha\Phi(z) + z\Phi'(z)) = 0$$

by (3.1). Thus $\Phi(z) = \beta z^\alpha$ for some $\beta \in \mathbb{C}$. Since Φ is injective, we must have $\alpha = 1$, and so $R_1 = R_2$. ■

Exercises

3.2.1

Section 3.3

$\mathbb{C}$-power series

Exercises

3.3.1

Section 3.4

Some $\mathbb{C}$-valued functions of interest

3.4.1 The exponential function

3.4.1 Proposition (Properties of the exponential function)

3.4.2 Power functions and general logarithmic functions

Alexandrian on continuous dependence of roots of polynomial on coefficients

Section 3.5

Integration

Much of what interests in complex variable theory centres around integration. In this section we give a rapid overview of the essential facts.

A ***curve*** in $\mathbb{C}$ is a continuous map $c\colon [a,b] \to \mathbb{C}$. A ***closed curve*** in $\mathbb{C}$ is a curve $c\colon [a,b] \to \mathbb{C}$ for which $c(a) = c(b)$. Thus a closed curve forms a loop with no intersections (see Figure 3.3). A curve c defined on $[a,b]$ is ***simple*** if the restriction of

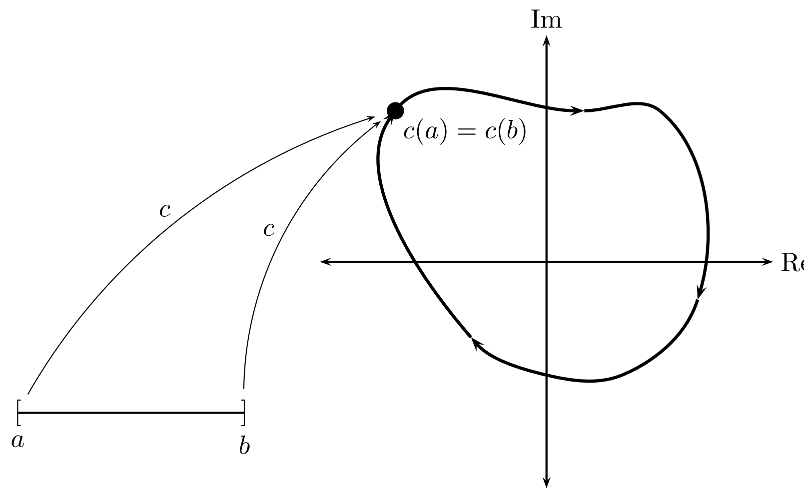

Figure 3.3 A closed curve in $\mathbb{C}$

c to (a,b) is injective. Thus for each $t_1, t_2 \in (a,b)$ the points $c(t_1)$ and $c(t_2)$ are distinct. Sometimes a simple closed curve is called a ***Jordan curve***. The Jordan Curve Theorem then states that a simple closed curve separates $\mathbb{C}$ into two domains, the interior and the exterior. This also allows us to make sense of the ***orientation*** of a simple closed curve. We shall speak of simple closed curves as having "clockwise orientation" or "counterclockwise orientation." Let us agree not to give these precise notation as the meaning will be obvious in any application we encounter.

Sometimes we will wish for a curve to have more smoothness, and so speak of a ***differentiable curve*** as one where the functions $u, v\colon [a,b] \to \mathbb{R}$ defined by $c(t) = u(t) + iv(t)$ are differentiable. For short, we shall call a differentiable curve an ***arc***. In such cases we denote

$$c'(t) = \frac{\mathrm{d}u}{\mathrm{d}t} + i\frac{\mathrm{d}v}{\mathrm{d}t}.$$

The ***length*** of a differentiable curve $c\colon [a, b] \to \mathbb{C}$ is given by

$$\int_a^b |c'(t)| \, \mathrm{d}t.$$

A ***contour*** is a curve that is a concatenation of a finite collection of disjoint differentiable curves.

If $c\colon [a, b] \to \mathbb{C}$ is a curve then we define

$$\int_a^b c(t) \, \mathrm{d}t = \int_a^b u(t) \, \mathrm{d}t + i \int_a^b v(t) \, \mathrm{d}t,$$

where u and v are defined by $c(t) = u(t) + iv(t)$. Now we let D be a domain in $\mathbb{C}$, $c\colon [a, b] \to D$ be an arc, and $f\colon D \to \mathbb{C}$ be a continuous function. We define

$$\int_c f(z) \, \mathrm{d}z = \int_a^b f(c(t))c'(t) \, \mathrm{d}t. \tag{3.2}$$

One may verify that this integral does not in fact depend on the parameterisation of c, and so really only depends on the "shape" of the image of c in $U \subseteq \mathbb{C}$. We shall typically denote $C = \mathrm{image}(c)$ and write $\int_C = \int_c$. If c is a contour, then one may similarly define the integral by defining it over each of the finite arcs comprising c. If $F\colon D \to \mathbb{C}$ is differentiable with continuous derivative f, then one verifies that

$$\int_c f(z) \, \mathrm{d}z = F(c(b)) - F(c(a)),$$

for a contour $c\colon [a, b] \to \mathbb{C}$.

The following theorem lies at the heart of much of complex analysis, and will be useful for us here.

3.5.1 Theorem (Cauchy's Integral Theorem) *Let* $\mathrm{D} \subseteq \mathbb{C}$ *be a simply connected domain, suppose that* $\mathrm{f}\colon \mathrm{D} \to \mathbb{C}$ *is analytic on the closure of* D*, and let* C *be a simple closed contour contained in* D*. Then*

$$\int_{\mathrm{C}} \mathrm{f}(z) \, \mathrm{d}z = 0.$$

Exercises

finish

3.5.1

3.5.2 Graphical calculation of residues from Truxal (page 27)

Section 3.6

Applications of Cauchy's Integral Theorem

Cauchy's Integral Theorem forms the basis for much that is special in the theory of complex variables. We shall give a few of the applications that are of interest to us in this book.

Let us begin by providing formulas for the coefficients in the Laurent expansion in terms of contour integrals. The following result does the job.

3.6.1 Proposition *Let* $f\colon D \to \mathbb{C}$ *be analytic and let* $z_0 \in D$ *be an isolated singularity for* f*. Let* C_0 *and* C_1 *be circular contours centred at* z_0 *with* C_1 *smaller than* C_0 *(see Figure 3.4). If*

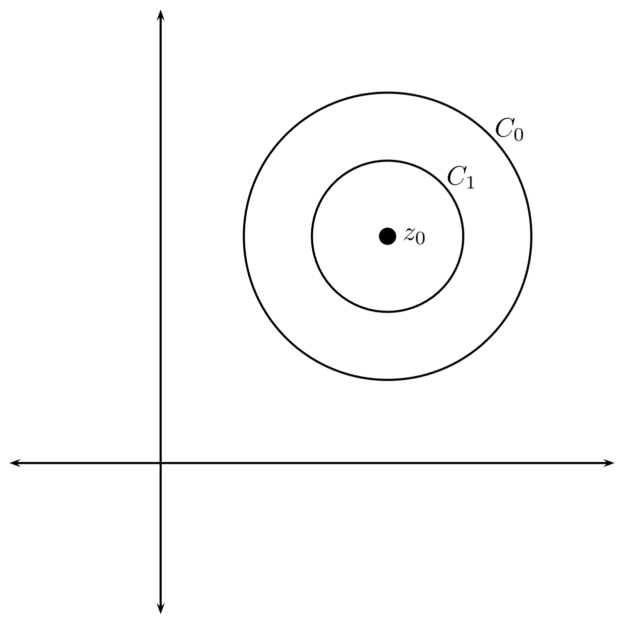

Figure 3.4 Contours for definition of Laurent series coefficients

$$f(z) = \sum_{j=-\infty}^{\infty} c_j (z - z_0)^j$$

is the Laurent series for f *at* z_0 *then we have*

$$c_j = \frac{1}{2\pi i} \int_{C_0} \frac{f(z)}{(z - z_0)^{j+1}}\, dz, \quad j = 0, 1, \dots$$
$$c_j = \frac{1}{2\pi i} \int_{C_1} \frac{f(z)}{(z - z_0)^{j+1}}\, dz, \quad j = -1, -2, \dots$$

3.6.2 Proposition $W(\gamma, z_0) = \frac{1}{2\pi i}\int_{\mathrm{image}(\gamma)} \frac{1}{z - z_0}\,dz.$

The ***residue*** of an analytic function f at an isolated singularity z_0 is the coefficient c_{-1} in the Laurent series for f at z_0. We denote the residue by

$$\mathrm{Res}_{z=p_j} f(z) = \frac{1}{2\pi i}\int_C f(z)\,dz,$$

where C is some sufficiently small circular contour centred at z_0. The Residue Theorem is also important for us.

3.6.3 Theorem (Residue Theorem) *Let* $D \subseteq \mathbb{C}$ *be a domain with* C *a simple, clockwise-oriented, closed contour in* D*. Let* $f\colon D \to \mathbb{C}$ *be meromorphic in the interior of* C *and analytic on* C*. Denote the poles of* f *in the interior of* C *by* $p_1, \dots, p_k$*. Then*

$$\int_C f(s)\,ds = 2\pi i \sum_{j=1}^{k} \mathrm{Res}_{s=p_j} f(s).$$

3.6.4 Theorem

Another useful result is the ***Poisson Integral Formula***.

3.6.5 Theorem (Poisson Integral Formula) *Let* $D \subseteq \mathbb{C}$ *be a domain containing the positive complex plane* $\overline{\mathbb{C}}_+$ *and let* $f\colon D \to \mathbb{C}$ *be analytic in* $\overline{\mathbb{C}}_+$*. Additionally, we will suppose that if for* $R > 0$ *we define* $m(R) > 0$ *by*

$$m(R) = \sup_{\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}]} |f(Re^{i\theta})|, \tag{3.3}$$

then f *has the property that*

$$\lim_{R\to\infty} \frac{m(R)}{R} = 0.$$

If $z_0 = x_0 + iy_0 \in \mathbb{C}_+$ *then we have*

$$f(z_0) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(iy)\frac{x_0}{x_0^2 + (x - x_0)^2}\,dy.$$

The Poisson Integral Formula has the following useful corollary, stated by Freudenberg and Looze [1985].

3.6.6 Corollary *Suppose that* D *is a domain containing* $\overline{\mathbb{C}}$ *and that* $f\colon D \to \mathbb{C}$ *is analytic and nonzero in* $\overline{\mathbb{C}}$*, with the possible exception of zeros on the imaginary axis. Also, assume that* $\log f$ *satisfies the equality (3.3). Then for each* $z_0 = x_0 + iy_0 \in \mathbb{C}_+$ *we have*

$$\log|f(z_0)| = \frac{1}{\pi}\int_{-\infty}^{\infty} \log|f(iy)|\frac{x_0}{x_0^2 + (x - x_0)^2}\,dy.$$

Finally, we state a sort of stray result, but one that is standard in complex variable theory, the ***Maximum Modulus Principle***.

3.6.7 Theorem *If* $\mathrm{f}\colon \mathrm{D} \to \mathbb{C}$ *is an analytic function on a domain* D*, then* $|\mathrm{f}|$ *has no maximum on* D *unless* f *is constant.*

From this result it follows that if f is analytic in a closed bounded region R, then the maximum value taken by $|f|$ must occur on the boundary of R.

Exercises

cayley-hamilton-cauchy.pdf

3.6.1

Section 3.7

Analytic continuation

Exercises

3.7.1

Section 3.8

Harmonic and subharmonic functions of a complex variable

3.8.1 Harmonic functions

First we discuss harmonic functions.
Now we turn to harmonic and subharmonic functions.

3.8.1 Definition (Harmonic function) Let $\Omega \subseteq \mathbb{C}$ be open and let $u\colon \Omega \to \mathbb{R}$. The function u is ***harmonic*** if it is of class C^2 and if $\frac{\partial^2 u}{\partial z \partial \bar{z}}(z) = 0$ for every $z \in \Omega$. •

It will be convenient on occasion to use the notation

$$\Delta u(z) = 4\frac{\partial^2 u}{\partial z \partial \bar{z}}(z),$$

this being the ***Laplacian*** of u.

Let us give some of the basic properties of harmonic functions.

3.8.2 Theorem (Properties of harmonic functions) *If $\Omega \subseteq \mathbb{C}$ is open, the following statements hold:*

(i) *if $\Phi \in \mathrm{C}^{\mathrm{hol}}(\Omega)$ then $\mathrm{Re}(\Phi)$ is harmonic;*

(ii) *if Ω is an open disk and if $\mathrm{u}\colon \Omega \to \mathbb{R}$ is harmonic, then there exists $\Phi \in \mathrm{C}^{\mathrm{hol}}(\Omega)$ such that $\mathrm{u} = \mathrm{Re}(\Phi)$;*

(iii) *if $\mathrm{u}\colon \Omega \to \mathbb{R}$ is harmonic then, for each $\mathrm{z}_0 \in \Omega$, there exists $\rho \in \mathbb{R}_{>0}$ such that $\overline{\mathsf{D}}(\rho, \mathrm{z}_0) \subseteq \Omega$ and such that*

$$\mathrm{u}(\mathrm{z}_0) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{u}(\mathrm{z}_0 + \mathrm{r}e^{\mathrm{i}\theta})\,\mathrm{d}\theta$$

for every $\mathrm{r} \in (0, \rho]$;

(iv) *if $\mathrm{u}\colon \Omega \to \mathbb{R}$ is continuous and if, for every $\mathrm{z}_0 \in \Omega$, there exists $\rho \in \mathbb{R}_{>0}$ such that $\overline{\mathsf{D}}(\mathrm{r}, \mathrm{z}_0) \subseteq \Omega$ and such that*

$$\mathrm{u}(\mathrm{z}_0) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{u}(\mathrm{z}_0 + \mathrm{r}e^{\mathrm{i}\theta})\,\mathrm{d}\theta$$

for every $\mathrm{r} \in (0, \rho]$, then u is harmonic.

Proof (i) Write $\Phi(z) = u(z) + \mathrm{i}v(z)$ for $\mathbb{R}$-valued smooth functions u and v on Ω.

wirtinger Referring to we have

$$\frac{\partial^2 u}{\partial z \partial \bar{z}} = \frac{1}{4}\left(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2}\right).$$

The Cauchy–Riemann equations are

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x},$$

and from this we immediately see that $\frac{\partial^2 u}{\partial z \partial \bar{z}} = 0$ by equality of mixed partials.

(ii) We will define $v\colon \Omega \to \mathbb{R}$ such that $\Phi \triangleq u + \mathrm{i}v$ is holomorphic. Let $z_0 = x_0 + \mathrm{i}y_0 \in \Omega$ be the centre of the disk Ω and let $r \in \mathbb{R}_{>0}$ be the radius. For $z = x + \mathrm{i}y \in \Omega$ define

$$v(z) = \int_{y_0}^{y} \frac{\partial u}{\partial x}(x, \eta)\,\mathrm{d}\eta - \int_{x_0}^{x} \frac{\partial u}{\partial y}(\xi, y_0)\,\mathrm{d}\xi.$$

One can verify by direct computation that $\Phi = u + \mathrm{i}v$ satisfies the Cauchy–Riemann equations, and so is holomorphic.

(iii) Let Φ be holomorphic in a neighbourhood of z containing $\overline{\mathsf{D}}(r, z)$. By the Cauchy integral formula,

$$\Phi(z) = \frac{1}{2\pi\mathrm{i}} \int_{\mathrm{bd}(\mathsf{D}(r,z))} \frac{\Phi(\zeta)}{\zeta - z}\,\mathrm{d}\zeta.$$

Letting $\zeta = z_0 + re^{\mathrm{i}\theta}$ and taking real parts gives the result.

(iv) Follows from properties of circular convolution ■ finish

3.8.3 Remarks (Harmonic functions)

1. Note that it is clear that the imaginary part of a holomorphic function is also a harmonic function (since multiplication of a holomorphic function by $-\mathrm{i}$ produces another holomorphic function). Given a harmonic function u and a holomorphic function Φ for which $u = \mathrm{Re}(\Phi)$, we say that $\mathrm{Im}(\Phi)$ is the ***harmonic conjugate*** of u.
2. Since a harmonic function is the real part of a holomorphic function, it follows that harmonic functions are infinitely differentiable, although their definition only requires them to be of class C^2. •

The preceding result then allows us to construct some examples of harmonic functions.

3.8.4 Examples (Harmonic functions)

1. If $\Omega \subseteq \mathbb{C}$ is open and if $f \in C^{\mathrm{hol}}(\Omega)$ then the function

$$\Omega \ni z \mapsto \log(|f|)(z) \triangleq \begin{cases} \log(|f(z)|), & f(z) \neq 0, \\ -\infty, & f(z) = 0, \end{cases}$$

is harmonic on $\Omega \setminus f^{-1}(0)$. To see that $\log(|f|)$ is harmonic on $\Omega \setminus f^{-1}(0)$, let $z_0 \in \Omega \setminus f^{-1}(0)$ and in a neighbourhood of z_0 write

$$\log(f(z)) = \log(|f(z)|) + \mathrm{i}\theta,$$

where $f(z) = re^{\mathrm{i}\theta}$. Thus $\log(|f|)$ is the real part of a holomorphic function, and so harmonic in a neighbourhood of z. •

3.8.2 Subharmonic functions

In this section we discuss subharmonic functions of a single complex variable. We can get some insight by thinking first about standard notions of convexity. A function $u: I \to \mathbb{R}$ defined on an interval I is ***convex*** if

$$u((1-s)x_1 + sx_2) \leq (1-s)u(x_1) + su(x_2)$$

for every distinct $x_1, x_2 \in I$ and for every $s \in (0,1)$. In Figure 3.5 we depict how

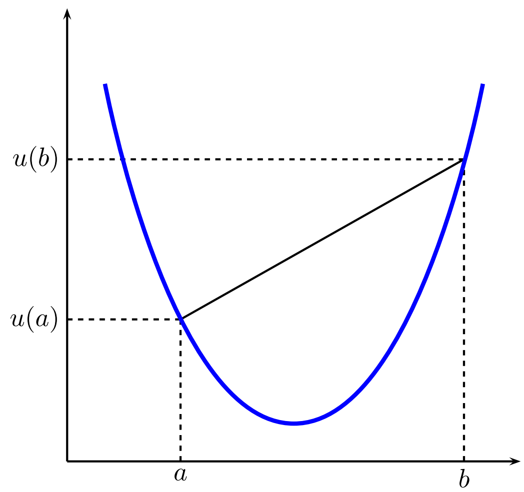

Figure 3.5 A convex function

the definition works. The idea—and one that relates to how we will think of how subharmonic relates to harmonic—is that if u agrees with a linear function at points a and b, then u does not exceed the linear function on (a,b). Said in this way, it is perhaps not unreasonable to think of "convex" functions as being "sublinear." Moreover, the linear functions can be thought of those twice differentiable functions with zero second derivative. It is a classical result that a function of class C^2 is convex if and only if $u''(x) \geq 0$ for every $x \in I$ [Webster 1994, Theorem 5.5.5]. We shall see in Theorem 3.8.6(vii) below that a similar interpretation holds for subharmonic functions, but with "second derivative" being replaced with "Laplacian."

We recall from Section 1.10.2 that a function $f: S \to [-\infty, \infty)$ is ***upper semicontinuous*** if $f^{-1}([-\infty, \alpha))$ is open for every $\alpha \in \mathbb{R}$.

3.8.5 Definition (Subharmonic function) Let $\Omega \subseteq \mathbb{C}$ be open and let $u: \Omega \to [-\infty, \infty)$. The function u is ***subharmonic*** if

(i) it is upper semicontinuous;

(ii) for every $r \in \mathbb{R}_{>0}$ and $z_0 \in \Omega$ for which $\overline{\mathsf{D}}(r, z_0) \subseteq \Omega$ and for every continuous $\sigma: \overline{\mathsf{D}}(r, z_0) \to \mathbb{R}$ such that (1) $\sigma|\mathsf{D}(r, z_0)$ is harmonic, and (2) $\sigma(z) \geq u(z)$ for $z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$, we have $\sigma(z) \geq u(z)$ for every $z \in \overline{\mathsf{D}}(r, z_0)$. •

Let us give some of the basic properties of harmonic functions.

3.8.6 Theorem (Properties of harmonic functions) *If $\Omega \subseteq \mathbb{C}$ is open, the following statements hold:*

(i) *if* u *is harmonic then it is subharmonic;*

(ii) *if* $(u_j)_{j\in\mathbb{Z}_{>0}}$ *is a sequence of subharmonic functions on* Ω *such that* $u_{j+1}(z) \leq u_j(z)$ *for each* $j \in \mathbb{Z}_{>0}$ *and* $z \in \Omega$*, then the function* u *on* Ω *defined by* $u(z) = \lim_{j\to\infty} u_j(z)$ *is subharmonic;*

(iii) *if* $(u_a)_{a\in A}$ *is a family of subharmonic functions on* Ω *then the function* u *on* Ω *defined by*

$$u(z) = \sup\{u_a(z) \mid a \in A\}$$

is subharmonic if it is upper semicontinuous and everywhere finite;

(iv) *if* $u_1, \dots, u_k \colon \Omega \to [-\infty, \infty)$ *are subharmonic and if* $F \colon \mathbb{R}^k \to \mathbb{R}$ *is continuous, convex, and nondecreasing in each component, and if we extend* F *to* $\bar{F} \colon ([-\infty, \infty))^k \to [-\infty, \infty)$ *by continuity,*[3] *then the function*

$$z \mapsto F(u_1(z), \dots, u_k(z))$$

is subharmonic;

(v) *if* $u \colon \Omega \to \mathbb{R}$ *is upper semicontinuous, then it is subharmonic if and only if, for each* $z_0 \in \Omega$*, there exists* $\rho \in \mathbb{R}_{>0}$ *such that* $\overline{\mathsf{D}}(\rho, z_0) \subseteq \Omega$ *and such that*

$$u(z_0) \leq \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{i\theta})\, d\theta$$

for every $r \in (0, \rho]$*;*

(vi) *if* Ω *is connected and if* u *is subharmonic and has a global maximum in* Ω*, then* u *is constant;*

(vii) *if* u *is of class* C^2*, then it is subharmonic if and only if* $\frac{\partial^2 u}{\partial z \partial \bar{z}}(z) \geq 0$ *for every* $z \in \Omega$*.*

Proof (i) If u is harmonic it is continuous and so upper semicontinuous. By Theorem 3.8.2(iii) and by (v) below, it then follows that if u is harmonic it is subharmonic.

(ii) By [Aliprantis and Border 2006, Lemma 2.41] we have that u is upper semicontinuous. Let $z_0 \in \Omega$ and $r \in \mathbb{R}_{>0}$ be such that $\overline{\mathsf{D}}(r, z_0) \subseteq \Omega$. Let σ be a continuous function on $\overline{\mathsf{D}}(r, z_0)$ that is harmonic on $\mathsf{D}(r, z_0)$ and is such that $\sigma(z) \geq u(z)$ for all $z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$. Let $\epsilon \in \mathbb{R}_{>0}$ and for $j \in \mathbb{Z}_{>0}$ define

$$K_{j,\epsilon} = \{z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0)) \mid u_j(z) \geq \sigma(z) + \epsilon\}.$$

Note that $K_{j,\epsilon}$ is compact, that $K_{j+1,\epsilon} \subseteq K_{j,\epsilon}$, and that $\cap_{j\in\mathbb{Z}_{>0}} K_{j,\epsilon} = \emptyset$, the latter since $\lim_{j\to\infty} u_j(x) = u(x) \leq \sigma(x)$. It follows, since the intersection of a nested sequence of nonempty compact sets is nonempty [Rudin 1976, Corollary to Theorem 2.36], that

[3] A little precisely, we define $\bar{F}$ as follows. Suppose that we wish to evaluate $\bar{F}$ at a point where $x_{j_1} = \cdots = x_{j_m} = -\infty$ for and only for some $j_1, \dots, j_m \in \{1, \dots, k\}$. We then let each of the coordinates $x_{j_1}, \dots, x_{j_m}$ tend together monotonically to ∞, while fixing the remaining coordinates at their desired values. The value of $\bar{F}$ at this point is then the limit of the values of F.

there exists $N \in \mathbb{Z}_{>0}$ such that $K_{N,\epsilon} = \emptyset$. Thus, for $j \geq N$, $u_j(z) < \sigma(z) + \epsilon$ for every $z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$ and so $u_j(z) < \sigma(z) + \epsilon$ for every $z \in \overline{\mathsf{D}}(r, z_0)$. It follows that $u(z) < \sigma(z) + \epsilon$ for every $z \in \overline{\mathsf{D}}(r, z_0)$, and so $u(z) \leq \sigma(z)$ for every $z \in \overline{\mathsf{D}}(r, z_0)$, as desired.

(iii) Let $z_0 \in \Omega$ and $r \in \mathbb{R}_{>0}$ be such that $\overline{\mathsf{D}}(r, z_0) \subseteq \Omega$. Let $\sigma \colon \overline{\mathsf{D}}(r, z_0) \to \mathbb{R}$ be continuous, harmonic on $\mathsf{D}(r, z_0)$, and satisfying $\sigma(z) \geq u(z)$ for $z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$. We then have $\sigma(z) \geq u_a(z)$ for $z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$ and $a \in A$. It follows that $\sigma(z) \geq u_a(z)$ for $z \in \overline{\mathsf{D}}(r, z_0)$ and $a \in A$, and so $\sigma(z) \geq u(z)$ for $z \in \overline{\mathsf{D}}(r, z_0)$, as desired.

(iv) Let $(\phi_a)_{a \in A}$ be a family of affine functions $\phi_a \colon \mathbb{R}^k \to \mathbb{R}$ such that

$$\{(\boldsymbol{x}, y) \in \mathbb{R}^k \times \mathbb{R} \mid y \geq F(\boldsymbol{x})\} = \cap_{a \in A} \{(\boldsymbol{x}, y) \in \mathbb{R}^k \times \mathbb{R} \mid y \geq \phi_a(\boldsymbol{x}) \text{ for all } a \in A\}$$

what (this is possible since the epigraph of a convex function is convex). Then we have

$$F(\boldsymbol{x}) = \sup\{\phi_a(\boldsymbol{x}) \mid a \in A\}$$

for every $\boldsymbol{x} \in \mathbb{R}^k$ [Webster 1994, Theorem 5.4.2]. If we write $\phi_a(\boldsymbol{x}) = \langle \boldsymbol{m}_a, \boldsymbol{x} \rangle + b_a$, $a \in A$, the fact that F is increasing implies that the components of $\boldsymbol{m}$ are nonnegative. By subharmonicity of $u_1, \dots, u_k$ and part (v) below we thus have

$$\begin{aligned} \sum_{j=1}^k m_{a,j} u_j(z) + b_a &\leq \frac{1}{2\pi} \sum_{j=1}^k m_{a,j} \int_0^{2\pi} (u_j(z + r\mathrm{e}^{\mathrm{i}\theta}) + b_a) \, \mathrm{d}\theta \\ &\leq \frac{1}{2\pi} \int_0^{2\pi} F(u_1(z + r\mathrm{e}^{i\theta}), \dots, u_k(z + r\mathrm{e}^{\mathrm{i}\theta})) \, \mathrm{d}\theta \end{aligned}$$

for sufficiently small $r \in \mathbb{R}_{>0}$ and for all $a \in A$. This part of the result follows by taking the supremum over $a \in A$ and again applying part (v) below.

(v) First consider a general upper semicontinuous function $v \colon \Omega \to \mathbb{R}$ that satisfies

$$v(z_0) \leq \frac{1}{2\pi} \int_0^{2\pi} v(z_0 + r\mathrm{e}^{\mathrm{i}\theta}) \, \mathrm{d}\theta$$

for $r \in \mathbb{R}_{>0}$ and $z_0 \in \Omega$ such that $\overline{\mathsf{D}}(r, z_0) \subseteq \Omega$. A look through the proof of part (vi) below shows that this implies that v is constant on any connected component of Ω on which it attains its maximum.

Now suppose that

$$u(z_0) \leq \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + r\mathrm{e}^{\mathrm{i}\theta}) \, \mathrm{d}\theta$$

for $r \in \mathbb{R}_{>0}$ and $z_0 \in \Omega$ such that $\overline{\mathsf{D}}(r, z_0) \subseteq \Omega$. Now let $r \in \mathbb{R}_{>0}$ and $z_0 \in \Omega$ be such that $\overline{\mathsf{D}}(r, z_0) \subseteq \Omega$ and let $\sigma \colon \overline{\mathsf{D}}(r, z_0) \to \mathbb{R}$ be continuous, harmonic on $\mathsf{D}(r, z_0)$, and be such that $u(z) \leq \sigma(z)$ for $z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$. If we take $v = u - \sigma$ then we have, by hypothesis and harmonicity of σ,

$$v(z_0) \leq \frac{1}{2\pi} \int_0^{2\pi} v(z_0 + r\mathrm{e}^{\mathrm{i}\theta}) \, \mathrm{d}\theta.$$

As mentioned in the preceding paragraph, this implies that v attains its maximum on $\mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$. That is,

$$u(z) - \sigma(z) \leq \sup\{u(\zeta) - \sigma(\zeta) \mid \zeta \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))\} \leq 0$$

for every $\zeta \in \overline{\mathsf{D}}(r, z_0)$. Thus u is subharmonic.

For the converse assertion, we use a lemma which makes reference to the so-called ***Poisson kernel***. This is a family of maps P_r defined for each $r \in \mathbb{R}_{>0}$ by

$$\begin{aligned} P_r \colon \mathsf{D}(r, 0) \times \mathrm{bd}(\mathsf{D}(r, 0)) &\to \mathbb{R} \\ (z, r\mathrm{e}^{\mathrm{i}\theta}) &\mapsto \frac{1}{2\pi} \frac{r^2 - |z|^2}{|z - r\mathrm{e}^{\mathrm{i}\theta}|^2}. \end{aligned}$$

We shall require the following fact that is rather important in its own right, as it is the solution to the so-called ***Dirichlet Problem*** for the unit disk.

1 Lemma *If* $\mathrm{u} \colon \mathrm{bd}(\overline{\mathsf{D}}(1, 0)) \to \mathbb{R}$ *is continuous, then the function* $\sigma \colon \overline{\mathsf{D}}(1, 0) \to \mathbb{R}$ *defined by*

$$\sigma(\mathrm{z}) = \begin{cases} \int_0^{2\pi} \mathrm{u}(\mathrm{e}^{\mathrm{i}\phi}) \mathrm{P}_1(\mathrm{z}, \mathrm{e}^{\mathrm{i}\phi})\, \mathrm{d}\phi, & \mathrm{z} \in \mathsf{D}(1, 0), \\ \mathrm{u}(\mathrm{z}), & \mathrm{z} \in \mathrm{bd}(\mathsf{D}(1, 0)), \end{cases}$$

is continuous and harmonic on $\mathsf{D}(1, 0)$*.*

Proof First we prove that σ is continuous at points on $\mathrm{bd}(\overline{\mathsf{D}}(1, 0))$. Let $z_0 = \mathrm{e}^{\mathrm{i}\theta_0} \in \mathrm{bd}(\overline{\mathsf{D}}(1, 0))$. Denote

$$M = \sup\{|u(z)| \mid z \in \mathrm{bd}(\mathsf{D}(1, 0))\}.$$

Let $\epsilon \in \mathbb{R}_{>0}$ and use uniform continuity of u to choose $\delta \in \mathbb{R}_{>0}$ such that if $|a - b| < \delta$ then $|u(\mathrm{e}^{\mathrm{i}a}) - u(\mathrm{e}^{\mathrm{i}b})| < \frac{\epsilon}{2}$. Let $z = r\mathrm{e}^{\mathrm{i}\theta} \in \mathsf{D}(1, 0)$ be chosen sufficiently close to z_0 so that $|\theta - \theta_0| < \frac{\delta}{3}$ and $r \in [\frac{1}{2}, 1)$ and $1 - r < \frac{\delta^2 \epsilon}{100M}$. We then perform a couple of preliminary estimates.

First we note that

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{1 - |r\mathrm{e}^{\mathrm{i}\theta}|^2}{|r\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{\mathrm{i}\phi}|^2}\, \mathrm{d}\phi = 1. \tag{3.4}$$

(This can be proved by using the Poisson Integral Formula; I used Mathematica®.) We have

$$\left| \frac{1}{2\pi} \int_{|\phi - \theta_0| < \delta} (u(\mathrm{e}^{\mathrm{i}\theta_0}) - u(\mathrm{e}^{\mathrm{i}\phi})) \frac{1 - r^2}{|r\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{\mathrm{i}\phi}|^2}\, \mathrm{d}\phi \right| \le \frac{\epsilon}{2} \frac{1}{2\pi} \int_0^{2\pi} \frac{1 - r^2}{|r\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{\mathrm{i}\phi}|^2}\, \mathrm{d}\phi \le \frac{\epsilon}{2},$$

using (3.4) and the fact that

$$\frac{1 - r^2}{|r\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{\mathrm{i}\phi}|^2} \ge 0.$$

By elementary computations we have

$$|r\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{\mathrm{i}\phi}|^2 = |1 - r\mathrm{e}^{\mathrm{i}(\theta - \phi)}|^2 = 1 - 2r\cos(\theta - \phi) + r^2$$

Now we estimate

$$\begin{aligned} |r\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{\mathrm{i}\phi}|^2 &= (1 - r)^2 + 2r(1 - \cos(\theta - \phi)) \\ &\ge 2r(1 - \cos(\theta - \phi)) \ge 2r\frac{(\theta - \phi)^2}{2} \ge \frac{(\theta - \phi)^2}{4}, \end{aligned}$$

using the Taylor expansion of cos for small angles and using the definition of r. Given that $|\theta - \theta_0| < \frac{\delta}{3}$ and if we take $|\phi - \theta_0| \geq \delta$ we have that $|\theta - \phi| \geq \frac{2\delta}{3}$. Thus we have

$$\begin{aligned}\left|\frac{1}{2\pi}\int_{|\phi-\theta_0|\geq\delta}(u(\mathrm{e}^{\mathrm{i}\theta_0}) - u(\mathrm{e}^{\mathrm{i}\phi}))\frac{1-r^2}{|r\mathrm{e}^{\mathrm{i}\theta} - \mathrm{e}^{\mathrm{i}\phi}|^2}\,\mathrm{d}\phi\right| &\leq \frac{1}{2\pi}8M\int_{|\phi-\theta_0|\geq\delta}\frac{1-r^2}{(\theta-\phi)^2}\,\mathrm{d}\phi \\ &\leq \frac{1}{2\pi}\frac{72M}{4\delta^2}\int_0^{2\pi}(1+r)(1-r)\,\mathrm{d}\phi \\ &\leq \frac{1}{2\pi}\frac{72M}{4\delta^2}2\frac{\delta\epsilon}{100M} \leq \frac{\epsilon}{2}.\end{aligned}$$

Now let us put the preceding estimates together. Using (3.4) we have

$$\sigma(z_0) - \sigma(z) = \frac{1}{2\pi}\int_0^{2\pi}(u(z_0) - u(z))\frac{1-|z^2|}{|z-\mathrm{e}^{\mathrm{i}\phi}|^2}\,\mathrm{d}\phi.$$

Then

$$\begin{aligned}|\sigma(z_0) - \sigma(z)| &= \left|\frac{1}{2\pi}\int_0^{2\pi}(u(z_0) - u(z))\frac{1-|z^2|}{|z-\mathrm{e}^{\mathrm{i}\phi}|^2}\,\mathrm{d}\phi\right| \\ &\leq \left|\frac{1}{2\pi}\int_{|\phi-\theta_0|\leq\delta}u(z_0) - u(z))\frac{1-|z^2|}{|z-\mathrm{e}^{\mathrm{i}\phi}|^2}\,\mathrm{d}\phi\right| \\ &\quad + \left|\frac{1}{2\pi}\int_{|\phi-\theta_0|\geq\delta}u(z_0) - u(z))\frac{1-|z^2|}{|z-\mathrm{e}^{\mathrm{i}\phi}|^2}\,\mathrm{d}\phi\right| \leq \epsilon,\end{aligned}$$

giving continuity at boundary points, as desired.

Now we show that σ is harmonic on $\mathsf{D}(1,0)$. Here we use the directly verified identity

$$\frac{1-|z|^2}{|z-\mathrm{e}^{\mathrm{i}\phi}|^2} = \frac{\mathrm{e}^{\mathrm{i}\phi}}{\mathrm{e}^{\mathrm{i}\phi} - z} + \frac{\mathrm{e}^{-\mathrm{i}\phi}}{\mathrm{e}^{-\mathrm{i}\phi} - z} - 1.$$

Thus, for $z \in \mathsf{D}(1,0)$,

$$\sigma(z) = \frac{1}{2\pi}\int_0^{2\pi}u(\mathrm{e}^{\mathrm{i}\phi})\frac{\mathrm{e}^{\mathrm{i}\phi}}{\mathrm{e}^{\mathrm{i}\phi}-z}\,\mathrm{d}\phi + \frac{1}{2\pi}\int_0^{2\pi}u(\mathrm{e}^{\mathrm{i}\phi})\frac{\mathrm{e}^{-\mathrm{i}\phi}}{\mathrm{e}^{-\mathrm{i}\phi}-z}\,\mathrm{d}\phi - \frac{1}{2\pi}\int_0^{2\pi}u(\mathrm{e}^{\mathrm{i}\phi})\,\mathrm{d}\phi.$$

The first term on the right is holomorphic in z. The second term can be verified to be harmonic (i.e., its real and imaginary parts are harmonic) by simply differentiating under the integral sign to verify the conditions for a harmonic function. The last term is constant and so harmonic. Since u is real, we can take real parts to see that the right-hand side, each of which will be harmonic, to see that u is harmonic. ▼

Now suppose that u is continuous and subharmonic. Let $z_0 \in \Omega$ and $r \in \mathbb{R}_{>0}$ be such that $\overline{\mathsf{D}}(r, z_0) \subseteq \Omega$. By the lemma (and an elementary change of variable to translate $1 \to r$ and $0 \to z_0$), if we define

$$\sigma(z) = \frac{1}{2\pi}\int_0^{2\pi}P_r(z - z_0, r\mathrm{e}^{\mathrm{i}\theta})(u(z_0 + r\mathrm{e}^{\mathrm{i}\theta}))\,\mathrm{d}\theta,$$

then σ is harmonic on $\mathsf{D}(r, z_0)$ for each $\epsilon \in \mathbb{R}_{>0}$. Moreover, $\sigma(z) = u(z)$ for each $z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$. Since u is subharmonic, this implies that $u(z) \leq \sigma(z)$ for $z \in \mathsf{D}(r, z_0)$. This implies, for example, that

$$u(z_0) \leq \sigma(z_0) = \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{i\theta}) \, d\theta.$$

It remains to show that if u is upper semicontinuous and subharmonic, then

$$u(z_0) \leq \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{i\theta}) \, d\theta.$$

for $z_0 \in \Omega$ and for sufficiently small r. By [Aliprantis and Border 2006, Theorem 3.13] we let $(u_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence of continuous functions converging pointwise from above to u on $\overline{\mathsf{D}}(r, z_0)$. Then we have

$$u(z_0) = \lim_{j \to \infty} u_j(z_0) \leq \lim_{j \to \infty} \frac{1}{2\pi} \int_0^{2\pi} u_j(z_0 + re^{i\theta}) \, d\theta = \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{i\theta}) \, d\theta,$$

as desired.

(vi) Let

$$M = \sup\{u(z) \mid z \in \Omega\},$$

noting that $M < \infty$ by hypothesis. Indeed, there is $z_0 \in \Omega$ such that $u(z_0) = M$. We first claim that u is constant in some neighbourhood of z_0. Suppose otherwise, and let $r \in \mathbb{R}_{>0}$ be such that, for some $z_1 \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$ we have $u(z_0) > u(z_1)$. Since u is upper semicontinuous, let $(v_j)_{j \in \mathbb{Z}_{>0}}$ be a sequence of continuous functions on $\mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$ converging pointwise to u and such that $v_j(z) \geq u(z)$ for every $z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$ [Aliprantis and Border 2006, Theorem 3.13]. Choose N sufficiently large that $v_N(z_1) < M$. Then the function

$$\sigma(z) = \min\{v_N(z), M\}, \qquad z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0)),$$

is continuous. By the lemma above, we can extend σ to a harmonic function, which we also denote by σ, on $\overline{\mathsf{D}}(r, z_0)$. We then have, by part Theorem 3.8.2(iii) and noting that $\sigma(z_1) < M$,

$$\sigma(z_0) = \frac{1}{2\pi} \int_0^{2\pi} \sigma(z_0 + re^{i\theta}) \, d\theta < M = u(z_0),$$

contradicting the fact that u is subharmonic.

Thus u is constant in any neighbourhood of a point where it attains its maximum. Thus the set of points where u attains its maximum is open. As this set is clearly closed (its complement is $u^{-1}([-\infty, M))$ which is open since u is upper semicontinuous) and since Ω is connected, u is everywhere equal to M.

(vii) We first prove a lemma known as ***Green's third formula***. We use the following vector calculus notation. If $I \subseteq \mathbb{R}$ is an interval, if $\gamma\colon I \to \mathbb{R}^2$ is a differentiable curve for which $\|\gamma'(s)\| = 1$ for each $s \in I$, and if $u\colon \mathbb{R}^2 \to \mathbb{R}$ is differentiable, we denote

$$\frac{\partial u}{\partial \boldsymbol{n}_\gamma}(s) = \mathrm{grad}\, u(\gamma(s)) \cdot \boldsymbol{n}_\gamma(s),$$

where $\boldsymbol{n}_\gamma(s) = (\gamma_2'(s), -\gamma_1'(s))$ is the normal vector to γ at s. With this notation, if I is compact we denote

$$\int_{\mathrm{image}(\gamma)} \frac{\partial u}{\partial \boldsymbol{n}}\, \mathrm{d}s \triangleq \int_I \frac{\partial u}{\partial \boldsymbol{n}_\gamma}(s)\, \mathrm{d}s.$$

With this notation we have the following result.

2 Lemma *Let $\Omega \subseteq \mathbb{C}$ be a connected open set for which $\mathrm{bd}(\Omega)$ is the image of a finite number of differentiable curves and let $z_0 \in \Omega$. Let $u\colon \mathrm{cl}(\Omega) \to \mathbb{R}$ be continuous on $\mathrm{cl}(\Omega)$, of class C^2 on Ω, and such that $\mathbf{D}u$ extends to a continuous function on $\mathrm{bd}(\Omega)$. Let $v\colon \mathrm{cl}(\Omega) \setminus \{z_0\} \to \mathbb{R}$ be continuous, harmonic on $\Omega \setminus \{z_0\}$, be such that $\mathbf{D}v$ extends to a continuous function on $\mathrm{cl}(\Omega) \setminus \{z_0\}$, and such that $z \mapsto v(z) - \log(|z - z_0|^{-1})$ is harmonic in a neighbourhood of z_0. Then, denoting $z = x + iy$,*

$$u(z_0) = -\frac{1}{2\pi}\int_\Omega v(x,y)\left(\frac{\partial^2 u}{\partial x^2}(x,y) + \frac{\partial^2 u}{\partial y^2}(x,y)\right) \mathrm{d}x\mathrm{d}y - \frac{1}{2\pi}\int_{\mathrm{bd}(\Omega)}\left(u\frac{\partial v}{\partial \mathbf{n}} - v\frac{\partial u}{\partial \mathbf{n}}\right)\mathrm{d}s. \tag{3.5}$$

Proof Let $\gamma_j\colon [0, L_j] \to \mathbb{C}$, $j \in \{1, \dots, k\}$, be differentiable curves for which (1) $\gamma_j|[0, L_j)$ is a injection into $\mathrm{bd}(\Omega)$ for each $j \in \{1, \dots, k\}$ and (2) $\mathrm{bd}(\Omega)$ is a disjoint union of $\gamma([0, L_j))$, $j \in \{1, \dots, k\}$. Let u be as in the statement of the lemma, and let σ also have the same properties. Then, using Green's Theorem [Lang 1987, Chapter XIV],

$$\begin{aligned}\int_{\mathrm{bd}(\Omega)} u\frac{\partial \sigma}{\partial \boldsymbol{n}}\, \mathrm{d}s &= \sum_{j=1}^k \int_0^{L_j} u(\gamma_j(s))\left(\frac{\partial \sigma}{\partial x}(\gamma_j(s))\gamma_{j,2}'(s) - \frac{\partial \sigma}{\partial y}(\gamma_j(s))\gamma_{j,1}'(s)\right)\mathrm{d}s\\ &= \int_\Omega \left(\frac{\partial u}{\partial x}(x,y)\frac{\partial \sigma}{\partial x}(x,y) + \frac{\partial u}{\partial y}(x,y)\frac{\partial \sigma}{\partial y}(x,y) + u\frac{\partial^2 v}{\partial x^2}(x,y) + u\frac{\partial^2 \sigma}{\partial y^2}(x,y)\right)\mathrm{d}x\mathrm{d}y\\ &= \int_\Omega (\mathrm{grad}\, u(x,y) \cdot \mathrm{grad}\, \sigma(x,y) + u(x,y)\Delta\sigma(x,y))\, \mathrm{d}x\mathrm{d}y.\end{aligned}$$

This is ***Green's first formula***. Swapping the rôles of u and σ and subtracting then gives

$$\int_{\mathrm{bd}(\Omega)}\left(u\frac{\partial \sigma}{\partial \boldsymbol{n}} - \sigma\frac{\partial u}{\partial \boldsymbol{n}}\right)\mathrm{d}s = \int_\Omega (u(x,y)\Delta\sigma(x,y) - \sigma(x,y)\Delta u(x,y))\, \mathrm{d}x\mathrm{d}y,$$

which is ***Green's second formula***.

Let u and v be as in the statement of the lemma and let σ satisfy the same conditions as u, plus the condition that σ is harmonic on Ω. By Green's second formula we then have

$$\int_{\mathrm{bd}(\Omega)}\left(u\frac{\partial \sigma}{\partial \boldsymbol{n}} - \sigma\frac{\partial u}{\partial \boldsymbol{n}}\right)\mathrm{d}s + \int_\Omega \sigma(x,y)\Delta u(x,y)\, \mathrm{d}x\mathrm{d}y = 0.$$

Thus, in the formula (3.5), we can add a harmonic function to v and the formula still holds. In particular, we can add to v the harmonic function $z \mapsto -v(z) + \log(|z - z_0|^{-1})$ to conclude that, without loss of generality, we may take $v(z) = \log(|z - z_0|^{-1})$. Thus, in the remainder of the proof, we take this as v. One easily verifies that v is harmonic on $\mathbb{C} \setminus \{z_0\}$. Indeed, if we write $(z - z_0)^{-1} = re^{i\theta}$,

$$\log((z - z_0)^{-1}) = \log(|z - z_0|^{-1}) + i\theta,$$

and so, in any neighbourhood of any point in $\mathbb{C} \setminus \{z_0\}$, v is the real part of a holomorphic function, and so is harmonic by part Theorem 3.8.2(i).

Now let $r \in \mathbb{R}_{>0}$ be such that $\overline{\mathsf{D}}(r, z_0) \subseteq \Omega$ and define $\Omega_r = \Omega \setminus \overline{\mathsf{D}}(r, z_0)$. One then applies Green's second formula on Ω_r:

$$\int_{\mathrm{bd}(\Omega)} \left(u\frac{\partial v}{\partial \boldsymbol{n}} - v\frac{\partial u}{\partial \boldsymbol{n}}\right) \mathrm{d}s - \int_{\mathrm{bd}(\overline{\mathsf{D}}(r,z_0))} \left(u\frac{\partial v}{\partial \boldsymbol{n}} - v\frac{\partial u}{\partial \boldsymbol{n}}\right) \mathrm{d}s = -\int_{\Omega_r} v(x,y)\Delta u(x,y)\,\mathrm{d}x\mathrm{d}y. \quad (3.6)$$

Note that the singularity of v at z_0 is integrable. Indeed, making the change of variables to polar coordinates,

$$\int_{\mathsf{D}(1,0)} \log(|z - z_0|^{-1})\,\mathrm{d}x\mathrm{d}y = \int_0^{2\pi}\int_0^1 \log(r^{-1})r\,\mathrm{d}r\mathrm{d}\theta.$$

Since $\lim_{r\to 0} r\log(r^{-1}) = 0$, the integral is finite. From this we conclude that the right-hand side of (3.6) tends to 0 as r tends to 0. Let us now turn to the other terms in (3.6). First we denote by M a bound for $\mathrm{grad}\, u$ in a neighbourhood of z_0 containing $\overline{\mathsf{D}}(r, z_0)$. Then we have

$$\int_{\mathrm{bd}(\overline{\mathsf{D}}(r,z_0))} v\frac{\partial u}{\partial \boldsymbol{n}}\,\mathrm{d}s \le 2\pi r M\log(r^{-1}).$$

Thus

$$\lim_{r\to 0}\int_{\mathrm{bd}(\overline{\mathsf{D}}(r,z_0))} v\frac{\partial u}{\partial \boldsymbol{n}}\,\mathrm{d}s = 0.$$

On $\mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$), writing $z = z_0 + r\mathrm{e}^{\mathrm{i}\theta}$, we have

$$\frac{\partial v}{\partial \boldsymbol{n}} = \frac{\partial}{\partial r}\log(r^{-1}) = -r^{-1}$$

and $\mathrm{d}s = r\mathrm{d}\theta$. Thus

$$-\int_{\mathrm{bd}(\Omega)} u\frac{\partial v}{\partial \boldsymbol{n}}\,\mathrm{d}s = \int_0^{2\pi} u(z_0 + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta.$$

Now let $\epsilon \in \mathbb{R}_{>0}$ and choose r sufficiently small that $|u(z) - u(z_0)| < \frac{\epsilon}{2\pi}$ for $z \in \overline{\mathsf{D}}(r, z_0)$. Then we have

$$\left|2\pi u(z_0) - \int_0^{2\pi} u(z_0 + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta\right| \le \int_0^{2\pi} |u(z_0) - u(z_0 + r\mathrm{e}^{\mathrm{i}\theta})|\,\mathrm{d}\theta < \epsilon.$$

Thus

$$\lim_{r\to 0}\left(-\int_{\mathrm{bd}(\Omega)} u\frac{\partial v}{\partial \boldsymbol{n}}\,\mathrm{d}s\right) = 2\pi u(z_0).$$

Putting this all together gives the lemma. ▼

Proceeding with the proof of this part of the result, let $r \in \mathbb{R}_{>0}$ and z_0 be such that $\overline{\mathsf{B}}^1(r, z_0) \subseteq \Omega$. By the lemma and the computations from the proof of the lemma we have

$$u(z_0) = -\frac{1}{2\pi}\int_{\overline{\mathsf{D}}(r,z_0)} \Delta u(x,y)\log(r|z - z_0|^{-1})\,\mathrm{d}x\,\mathrm{d}y + \frac{1}{2\pi}\int_0^{2\pi} u(z_0 + r\mathrm{e}^{\mathrm{i}\theta})\,\mathrm{d}\theta. \quad (3.7)$$

Note that $\log(r|z-z_0|^{-1})$ is nonnegative on $\overline{\mathsf{D}}(r, z_0)$ and only zero on the boundary.

Now suppose that $\Delta u(z_0) < 0$ for some $z_0 \in \Omega$. We then choose $r \in \mathbb{R}_{>0}$ such that $\Delta u(z) < 0$ for all $z \in \overline{\mathsf{D}}(r, z_0)$, and we then see from (3.7) that

$$u(z_0) > \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{i\theta})\,d\theta$$

By part (v) it follows that u is not subharmonic.

Conversely, suppose that u is not subharmonic. By part (v) there exists $z_0 \in \Omega$ and $r \in \mathbb{R}_{>0}$ such that

$$u(z_0) > \frac{1}{2\pi} \int_0^{2\pi} u(z_0 + re^{i\theta})\,d\theta.$$

By (3.7) we conclude that Δu must be negative at points in a neighbourhood of z_0. ■

3.8.7 Remark (Subharmonic functions) The condition of upper semicontinuity for subharmonic functions might seem a little unmotivated. Many natural subharmonic functions are continuous (and in fact many authors assume continuity in their definitions of subharmonic functions.) However, it comes as a consequence of properties (ii) and (iii) that upper semicontinuity can arise in limiting processes where continuity is present. •

3.8.8 Examples (Subharmonic functions)

1. (Example 3.8.4–1 cont'd) If $\Omega \subseteq \mathbb{C}$ is open and if $f \in C^{\mathrm{hol}}(\Omega)$ then the function

$$\Omega \ni z \mapsto \log(|f|)(z) \triangleq \begin{cases} \log(|f(z)|), & f(z) \neq 0, \\ -\infty, & f(z) = 0, \end{cases}$$

is subharmonic on Ω. If f is identically zero in a neighbourhood of $z \in \Omega$, it is immediate that $\log(|f|)$ is subharmonic on this neighbourhood. It remains to consider points $z_0 \in \Omega$ such that $f(z_0) = 0$ but, on any neighbourhood of z_0, f is not identically zero. Let $r \in \mathbb{R}_{>0}$ be such that $\overline{\mathsf{D}}(r, z_0) \subseteq \Omega$ and let $\sigma\colon \overline{\mathsf{D}}(r, z_0) \to \mathbb{R}$ be a continuous function, harmonic on $\mathsf{D}(r, z_0)$, such that $\log(|f(z)|) \leq \sigma(z)$ for $z \in \mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$. Note that we clearly have

$$\log(|f(z_0)|) \leq \int_0^{2\pi} \log(|f(z_0 + re^{i\theta})|)\,d\theta.$$

The same condition holds for $\log(|f|) - \sigma$. Referring to the proof of part (vi) of Theorem 3.8.6, we see that this implies that $\log(|f|) - \sigma$, not being a constant function, has the property that it must achieve its maximum on $\mathrm{bd}(\overline{\mathsf{D}}(r, z_0))$. This implies that $\log(|f(z)|) \leq \sigma(z)$ for $z \in \mathsf{D}(r, z_0)$, giving subharmonicity of $\log(|f|)$. •

Summary

Bibliography

Abraham, R., Marsden, J. E., and Ratiu, T. S. [1988] *Manifolds, Tensor Analysis, and Applications*, number 75 in Applied Mathematical Sciences, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-96790-5.

Aliprantis, C. D. and Border, K. C. [2006] *Infinite-Dimensional Analysis, A Hitchhiker's Guide*, 3rd edition, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-540-32696-0.

Arzelà, C. [1885] *Sulla integrazione per serie*, Atti della Accademia Nazionale dei Lincei. Memorie. Classe di Scienze Fisiche, Matematiche e Naturali. Sezione Ia. Matematica, Meccanica, Astronomia, Geodesia e, **4**(1), pages 532–537, ISSN: 0391-8149.

— [1900] *Sulle serie di funzioni*, Atti della Accademia delle Scienze dell'Istituto di Bologna. Classe di Scienze Fisiche. Rendiconti. Serie XIII, **5**(8), pages 701–704, ISSN: 1122-4142.

Aubin, J.-P. [1998] *Optima and Equilibria, An Introduction to Nonlinear Analysis*, 2nd edition, number 140 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-540-64983-0.

Berge, C. [1959] *Espaces topologiques: fonctions multivoques*, Collection Universitaire de Mathematiques, Dunod: Paris.

Brouwer, L. E. J. [1912] *Beweis zur Invarianz des n-dimensionalen Gebiets*, Mathematische Annalen, **72**(1), pages 55–56, ISSN: 0025-5831, DOI: 10.1007/BF01456846.

Cartan, H. [1963] *Elementary Theory of Analytic Functions of One or Several Complex Variables*, Addison Wesley: Reading, MA, Reprint: [Cartan 1995].

— [1995] *Elementary Theory of Analytic Functions of One or Several Complex Variables*, Dover Publications, Inc.: New York, NY, ISBN: 978-0-486-68543-4, Original: [Cartan 1963].

Chellaboina, V.-S. and Haddad, W. M. [1995] *Is the Frobenius matrix norm induced?*, Institute of Electrical and Electronics Engineers. Transactions on Automatic Control, **40**(12), pages 2137–2139, ISSN: 0018-9286, DOI: 10.1109/9.478340.

Drakakis, K. and Pearlmutter, B. A. [2009] *On the calculation of the $\ell_2 \to \ell_1$ induced matrix norm*, International Journal of Algebra, **3**(5), pages 231–240, ISSN: 1312-8868, URL: http://www.m-hikari.com/ija/ija-password-2009/ija-password5-8-2009/drakakisIJA5-8-2009.pdf.

Edgar, G. A. [2007] *Measure, Topology, and Fractal Geometry*, 2nd edition, Undergraduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-74748-4.

Farkas, J. [1902] *Über die Theorie der einfachen Ungleichungen*, Journal für die Reine und Angewandte Mathematik, **124**, pages 1–24, ISSN: 0075-4102, DOI: 10.1515/crll.1902.124.1.

Freudenberg, J. S. and Looze, D. P. [1985] *Right half plane poles and zeros and design tradeoffs in feedback systems*, Institute of Electrical and Electronics Engineers. Transactions on Automatic Control, **30**(6), pages 555–565, ISSN: 0018-9286, DOI: 10.1109/TAC.1985.1104004.

Gamelin, T. W. [2001] *Complex Analysis*, Undergraduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-95069-3.

Gelbaum, B. R. and Olmsted, J. M. H. [1964] *Counterexamples in Analysis*, The Mathesis Series, Holden-Day, Inc.: San Francisco/London/Amsterdam, Reprint: [Gelbaum and Olmsted 2003].

— [2003] *Counterexamples in Analysis*, Dover Publications, Inc.: New York, NY, ISBN: 978-0-486-42875-8, Original: [Gelbaum and Olmsted 1964].

Gordon, R. A. [2000] *A convergence theorem for the Riemann integral*, Mathematics Magazine, **73**(2), pages 141–147, ISSN: 0025-570X, DOI: 10.2307/2691086.

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.

Hausdorff, F. [1937] *Mengenlehre*, Walter de Gruyter: Berlin/New York, Translation: [Hausdorff 2005].

— [2005] *Set Theory*, 2nd edition, AMS Chelsea Publishing: Providence, RI, ISBN: 978-0-8218-3835-8, Original: [Hausdorff 1937].

Horn, R. A. and Johnson, C. R. [2013] *Matrix Analysis*, 2nd edition, Cambridge University Press: New York/Port Chester/Melbourne/Sydney, ISBN: 978-0-521-83940-2.

Hurewicz, W. and Wallman, H. [1941] *Dimension Theory*, number 4 in Princeton Mathematical Series, Princeton University Press: Princeton, NJ, ISBN: 978-0-691-07947-9.

Jordan, C. [1887] *Cours d'Analyse de l'École Polytechnique*, volume 3, Gauthier-Villars: Paris.

Kakutani, S. [1941] *A generalization of Brouwer's fixed point theorem*, Duke Mathematical Journal, **8**(3), pages 457–459, ISSN: 0012-7094, DOI: 10.1215/S0012-7094-41-00838-4.

Klein, E. and Thompson, A. C. [1984] *Theory of Correspondences, Including Applications to Mathematical Economics*, Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley and Sons: NewYork, NY, ISBN: 978-0-471-88016-5.

Knopp, K. [1931] *Funktionentheorie*, 2 volumes, Göschen: Berlin/Leipzig, Reprint: [Knopp 1996].

— [1996] *Theory of Functions*, translated by F. Bagemihl, 2 volumes, Dover Publications, Inc.: New York, NY, ISBN: 978-0-486-69219-7, Original: [Knopp 1931].

Kulpa, W. and Tursański, M. [1988] *A proof of the domain invariance theorem*, Radovi Matematiçki, **4**(1), pages 337–341, ISSN: 0352-6100.

Lang, S. [1987] *Calculus of Several Variables*, 3rd edition, Undergraduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-96405-8.

— [2003] *Complex Analysis*, 4th edition, number 103 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-98592-3.

Lewin, J. [1986] *A truly elementary approach to the bounded convergence theorem*, The American Mathematical Monthly, **93**(5), pages 395–397, ISSN: 0002-9890, DOI: 10.2307/2323608.

Mann, H. B. [1943] *Quadratic forms with linear constraints*, The American Mathematical Monthly, **50**(7), pages 430–433, ISSN: 0002-9890, DOI: 10.2307/2303666.

McShane, E. J. [1973] *The Lagrange multiplier rule*, The American Mathematical Monthly, **80**(8), pages 922–925, ISSN: 0002-9890, DOI: 10.2307/2319406.

Milnor, J. W. [1978] *Analytic proofs of the "Hairy Ball Theorem" and the Brouwer Fixed Point Theorem*, The American Mathematical Monthly, **85**(7), pages 521–524, ISSN: 0002-9890, DOI: 10.2307/1969983.

Minkowski, H. [1897] *Allgemeine Lehrsätze über die konvexe Polyeder*, Nachrichten der Akademie der Wissenschaften in Göttingen. II. Mathematisch-Physikalische Klasse, pages 198–219, ISSN: 0065-5295, URL: https://eudml.org/doc/58391 (visited on 07/23/2014).

— [1903] *Volumen und Oberfläche*, Mathematische Annalen, **57**(4), pages 447–495, ISSN: 0025-5831, DOI: 10.1007/BF01445180.

Munkres, J. R. [1984] *Elements of Algebraic Topology*, Addison Wesley: Reading, MA, ISBN: 978-0-201-04586-4.

Nash, J. F. [1951] *Non-cooperative games*, Annals of Mathematics. Second Series, **54**(2), pages 286–295, ISSN: 0003-486X, DOI: 10.2307/1969529.

Needham, T. [1997] *Visual Complex Analysis*, Oxford University Press: Oxford, ISBN: 978-0-19-853446-4.

Neumann, J. von [1928] *Zur Theorie der Gesellschaftsspiele*, Mathematische Annalen, **100**(1), pages 295–320, ISSN: 0025-5831, DOI: 10.1007/BF01448847.

Rademacher, H. [1919] *Über partielle und totale Differenzierbarkheit von Funktionen mehrerer Variabelen und über die transformation der Doppelintegrale*, Mathematische Annalen, **79**(1), pages 340–359, ISSN: 0025-5831, DOI: 10.1007/BF01498415.

Rohn, J. [2000] *Computing the norm* $\|A\|_{\infty,1}$ *is NP-hard*, Linear and Multilinear Algebra, **47**(3), ISSN: 0308-1087, DOI: 10.1080/03081080008818644.

Rudin, W. [1976] *Principles of Mathematical Analysis*, 3rd edition, International Series in Pure & Applied Mathematics, McGraw-Hill: New York, NY, ISBN: 978-0-07-054235-8.

Searcóid, M. Ó. [2007] *Metric Spaces*, Springer Undergraduate Mathematics Series, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-1-84628-369-7.

Seron, M. M., Braslavsky, J. H., and Goodwin, G. C. [1997] *Fundamental Limitations in Filtering and Control*, Communications and Control Engineering Series, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-3-540-76126-6.

Spring, D. [1985] *On the second derivative test for constrained local extrema*, The American Mathematical Monthly, **92**(9), pages 631–643, ISSN: 0002-9890, DOI: 10.2307/2323709.

Szabó, L. [1997] *A simple proof for the Jordan measurability of convex sets*, Elemente der Mathematik, **52**(2), pages 84–86, ISSN: 0013-6018, DOI: 10.1007/s000170050014.

Titchmarsh, E. C. [1939] *The Theory of Functions*, 2nd edition, Oxford University Press: Oxford, Reprint: [Titchmarsh 1976].

— [1976] *The Theory of Functions*, 2nd edition, Oxford University Press: Oxford, ISBN: 978-0-19-853349-8, Original: [Titchmarsh 1939].

Tverberg, H. [1980] *A proof of the Jordan curve theorem*, The Bulletin of the London Mathematical Society, **12**(1), pages 34–38, ISSN: 0024-6093, DOI: 10.1112/blms/12.1.34.

Veblen, O. [1905] *Theory of place curves in non-metrical analysis situs*, Transactions of the American Mathematical Society, **6**(2), pages 83–98, ISSN: 0002-9947, DOI: 10.1090/S0002-9947-1905-1500697-4.

Webster, R. J. [1994] *Convexity*, Oxford Science Publications, Oxford University Press: Oxford, ISBN: 978-0-19-853147-0.

Weyl, H. [1934] *Elementare Theorie der konvexen Polyeder*, Commentarii Mathematici Helvetici, A Journal of the Swiss Mathematical Society, **7**, pages 290–306, ISSN: 0010-2571, DOI: 10.1007/BF01292722.

Whyburn, G. T. [1942] *Analytic Topology*, number 28 in American Mathematical Society Colloquium Publications, American Mathematical Society: Providence, RI.

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.

Ziemer, W. P. [1989] *Weakly Differentiable Functions, Sobolev Spaces and Functions of Bounded Variation*, number 120 in Graduate Texts in Mathematics, Springer-Verlag: New York/Heidelberg/Berlin, ISBN: 978-0-387-97017-2.