This version: 28/02/2014

Appendix B
Convex analysis

In this appendix we review a few basic notions of convexity and related notions that will be important for us at various times.

B.1 The Hausdorff distance

We begin with a fairly simple measure of "closeness" of sets.

B.1.1 Definition (Distance between sets) Let $(\mathcal{M}, \mathrm{d})$ be a metric space. For nonempty subsets A and B of $\mathcal{M}$ the ***distance*** between A and B is

$$\mathrm{dist}(A, B) = \inf\{\mathrm{d}(x, y) \mid x \in A, \ y \in B\}.$$

If $A = \{x\}$ for some $x \in \mathcal{M}$, then we denote $\mathrm{dist}(x, B) = \mathrm{dist}(\{x\}, B)$ and, if $B = \{y\}$ for some $y \in \mathcal{M}$, then we denote $\mathrm{dist}(A, y) = \mathrm{dist}(A, \{y\})$. •

For general sets A and B there is not much useful one can say about $\mathrm{dist}(A, B)$. However, if we make some assumptions about the sets, then there is some structure here. Let us explore some of this.

B.1.2 Proposition (Continuity of distance to a set) *Let $(\mathcal{M}, \mathrm{d})$ be a metric space. If $B \subseteq \mathcal{M}$ then the function $x \mapsto \mathrm{dist}(x, B)$ on $\mathcal{M}$ is uniformly continuous in the metric topology.*

Proof Let $\epsilon \in \mathbb{R}_{>0}$ and take $\delta = \frac{\epsilon}{2}$. Let $y \in B$ be such that $\mathrm{d}(x_1, y) - \mathrm{dist}(x_1, B) < \frac{\epsilon}{2}$. Then, if $\mathrm{d}(x_1, x_2) < \delta$,

$$\mathrm{dist}(x_2, B) \leq \mathrm{d}(x_2, y) \leq \mathrm{d}(x_2, x_1) + \mathrm{d}(x_1, y) \leq \mathrm{dist}(x_1, B) + \epsilon.$$

In a symmetric manner one shows that

$$\mathrm{dist}(x_1, B) \leq \mathrm{dist}(x_2, B) + \epsilon,$$

provided that $\mathrm{d}(x_1, x_2) < \delta$. Therefore,

$$|\mathrm{dist}(x_1, B) - \mathrm{dist}(x_2, B)| < \epsilon,$$

provided that $\mathrm{d}(x_1, x_2) < \delta$, giving uniform continuity, as desired. ■

Now let us consider some properties of the distance function for closed sets. It is convenient to have at this point the notion of a ***Heine–Borel*** metric space, by which we mean one in which closed and bounded sets are compact.

B.1.3 Proposition (Set distance and closed sets) *Let* $(\mathcal{M}, \mathrm{d})$ *be a metric space. If* $\mathrm{A}, \mathrm{B} \subseteq \mathcal{M}$ *are closed sets then the following statements hold:*

(i) if $\mathrm{A} \cap \mathrm{B} = \emptyset$ *then* $\mathrm{dist}(\mathrm{x}, \mathrm{B}), \mathrm{dist}(\mathrm{A}, \mathrm{y}) > 0$ *for all* $\mathrm{x} \in \mathrm{A}$ *and* $\mathrm{y} \in \mathrm{B}$*;*

(ii) if $(\mathcal{M}, \mathrm{d})$ *is Heine–Borel and if* A *is compact, then there exists* $\mathrm{x}_0 \in \mathrm{A}$ *and* $\mathrm{y}_0 \in \mathrm{B}$ *such that* $\mathrm{dist}(\mathrm{A}, \mathrm{B}) = \mathrm{d}(\mathrm{x}_0, \mathrm{y}_0)$.

Proof (i) Suppose that $\mathrm{dist}(x, B) = 0$. Then there exists a sequence $(y_j)_{j\in\mathbb{Z}_{>0}}$ in B such that $\mathrm{d}(y_j, x) < \frac{1}{j}$ for each $j \in \mathbb{Z}_{>0}$. Thus the sequence $(y_j)_{j\in\mathbb{Z}_{>0}}$ converges to x and so $x \in \mathrm{cl}(B) = B$. Therefore, if $A \cap B = \emptyset$ we can conclude that, if $\mathrm{dist}(x, B) = 0$, then $x \notin A$. That is, $\mathrm{dist}(x, B) > 0$ for every $x \in A$, and similarly $\mathrm{dist}(A, y) > 0$ for every $y \in B$.

(ii) By Proposition B.1.2 the function $x \mapsto \mathrm{dist}(x, B)$ is continuous and so too then is its restriction to the compact set A. Thus, since continuous functions on compact sets obtain their minimum [Abraham, Marsden, and Ratiu 1988, Corollary 1.5.8], there exists $x_0 \in A$ such that $\mathrm{dist}(A, B) = \mathrm{dist}(x_0, B)$. Now there exists a sequence $(y_j)_{j\in\mathbb{Z}_{>0}}$ in B such that $\mathrm{d}(y_j, x_0) < \mathrm{dist}(x_0, B) + \frac{1}{j}$ for each $j \in \mathbb{Z}_{>0}$. Abbreviate $r = \mathrm{dist}(x_0, B)$. The sequence $(y_j)_{j\in\mathbb{Z}_{>0}}$ is contained in the closed ball $\overline{\mathsf{B}}(r + 1, x_0)$, which is compact since $\mathcal{M}$ is Heine–Borel. Therefore, by the Bolzano–Weierstrass Theorem [Abraham, Marsden, and Ratiu 1988, Theorem 1.5.4], there exists a convergent subsequence $(y_{j_k})_{k\in\mathbb{Z}_{>0}}$ converging to y_0. Since B is closed we necessarily have $y_0 \in B$. We claim that $\mathrm{dist}(A, B) = \mathrm{d}(x_0, y_0)$. Indeed, continuity of the metric ensures that

$$\mathrm{dist}(A, B) = \mathrm{dist}(x_0, B) = \lim_{k\to\infty} \mathrm{d}(y_{j_k}, x_0) = \mathrm{d}(y_0, x_0),$$

as desired. ■

B.2 Convex sets and affine subspaces

In this section we review the basic notions of convexity we use. A standard text with additional information along these lines is [Rockafellar 1970].

B.2.1 Definitions

We begin by defining subsets of a $\mathbb{R}$-vector space that have the properties we shall study.

B.2.1 Definition (Convex set, affine subspace) Let V be a $\mathbb{R}$-vector space.

(i) A subset $C \subseteq \mathsf{V}$ is ***convex*** if, for each $x_1, x_2 \in C$, we have

$$\{sx_1 + (1 - s)x_2 \mid s \in [0, 1]\} \subseteq C.$$

(ii) A subset $A \subseteq \mathsf{V}$ is an ***affine subspace*** if, for each $x_1, x_2 \in A$, we have

$$\{sx_1 + (1 - s)x_2 \mid s \in \mathbb{R}\} \subseteq A. \qquad \bullet$$

Note that the set $\{sx_1 + (1 - s)x_2 \mid s \in [0, 1]\}$ is the line segment in V between x_1 and 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, $\{\lambda x \mid \lambda \in \mathbb{R}_{\geq 0}\}$ is the ray emanating from $0 \in \mathsf{V}$ through the point x. An affine subspace is a set where the (bi-infinite) line through any two points in the set remains in the set.

B.2.2 Combinations and hulls

We shall be interested in generating convex sets and affine subspaces containing given sets.

B.2.2 Definition (Convex hull, affine hull) Let V be a $\mathbb{R}$-vector space and let $S \subseteq \mathsf{V}$ be nonempty.

(i) A ***convex combination*** from S is a linear combination in V of the form

$$\sum_{j=1}^{k} \lambda_j v_j, \qquad k \in \mathbb{Z}_{>0},\ \lambda_1, \dots, \lambda_k \in \mathbb{R}_{\geq 0},\ \sum_{j=1}^{k} \lambda_j = 1,\ v_1, \dots, v_k \in S.$$

(ii) The ***convex hull*** of S, denoted by $\operatorname{conv}(S)$, is the smallest convex subset of V containing S.

(iii) An ***affine combination*** from S is a linear combination in V of the form

$$\sum_{j=1}^{k} \lambda_j v_j, \qquad k \in \mathbb{Z}_{>0},\ \lambda_1, \dots, \lambda_k \in \mathbb{R},\ \sum_{j=1}^{k} \lambda_j = 1,\ v_1, \dots, v_k \in S.$$

(iv) The ***affine hull*** of S, denoted by $\operatorname{aff}(S)$, is the smallest affine subspace of V containing S. •

B.2.3 Remark (Sensibility of hull definitions) The definitions of $\operatorname{conv}(S)$ and $\operatorname{aff}(S)$ make sense because intersections of convex sets are convex and intersections of affine subspaces are affine subspaces. •

Convex combinations have the following useful property which also describes the convex hull.

B.2.4 Proposition (The convex hull is the set of convex combinations) *Let* V *be a* $\mathbb{R}$*-vector space, let* $\mathsf{S} \subseteq \mathsf{V}$ *be nonempty, and denote by* $\mathsf{C}(\mathsf{S})$ *the set of convex combinations from* S*. Then* $\mathsf{C}(\mathsf{S}) = \operatorname{conv}(\mathsf{S})$.

Proof First we show that $C(S)$ is convex. Consider two elements of $C(S)$ given by

$$x = \sum_{j=1}^{k} \lambda_j u_j, \quad y = \sum_{l=1}^{m} \mu_l v_l.$$

Then, for $s \in [0, 1]$ we have

$$sx + (1-s)y = \sum_{j=1}^{k} s\lambda_j u_j + \sum_{l=1}^{m} (1-s)\mu_j v_j.$$

For $r \in \{1, \dots, k+m\}$ define

$$w_r = \begin{cases} u_r, & r \in \{1, \dots, k\}, \\ v_{r-k}, & r \in \{k+1, \dots, k+m\} \end{cases}$$

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 $w_r \in S$ and $\rho_r \geq 0$ 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 $sx + (1-s)y \in C(S)$, and so $C(S)$ is convex.

This necessarily implies that $\operatorname{conv}(S) \subseteq C(S)$ since $\operatorname{conv}(S)$ is the smallest convex set containing S. To show that $C(S) \subseteq \operatorname{conv}(S)$ 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

$$\sum_{j=1}^{k} \lambda_j u_j, \qquad k \in \{1, \dots, m\},$$

is in $\operatorname{conv}(S)$, and consider a convex combination from S of the form

$$y = \sum_{l=1}^{m+1} \mu_l v_l = \sum_{l=1}^{m} \mu_l v_l + \mu_{m+1} v_{m+1}.$$

If $\sum_{l=1}^{m} \mu_l = 0$ then $\mu_l = 0$ for each $l \in \{1, \dots, m\}$. Thus $y \in \operatorname{conv}(S)$ 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,

$$\sum_{l=1}^{m} \mu'_l v_l \in \operatorname{conv}(S)$$

by the induction hypothesis. But we also have

$$y = (1 - \mu_{m+1}) \sum_{l=1}^{m} \mu'_l v_l + \mu_{m+1} v_{m+1}$$

by direct computation. Therefore, y is a convex combination of two elements of $\operatorname{conv}(S)$. Since $\operatorname{conv}(S)$ is convex, this means that $y \in \operatorname{conv}(S)$, 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.

B.2.5 Proposition (Characterisation of an affine subspace) *A nonempty subset* A *of a* $\mathbb{R}$-*vector space* V *is an affine subspace if and only if there exists* $\mathrm{x}_0 \in \mathsf{V}$ *and a subspace* $\mathsf{U} \subseteq \mathsf{V}$ *such that*

$$\mathrm{A} = \{\mathrm{x}_0 + \mathrm{u} \mid \mathrm{u} \in \mathsf{U}\}.$$

Proof Let $x_0 \in A$ and define $\mathsf{U} = \{x - x_0 \mid x \in A\}$. The result will be proved if we prove that U is a subspace. Let $x - x_0 \in \mathsf{U}$ for some $x \in A$ 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 $ax + (1-a)x_0 \in A$. For $x_1 - x_0, x_2 - x_0 \in \mathsf{U}$ with $x_1, x_2 \in A$ 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 $x_1 + x_2 - x_0 \in A$. However, we have

$$\begin{aligned}
& x_1 - x_0, x_2 - x_0 \in \mathsf{U}, \\
\implies\ & 2(x_1 - x_0), 2(x_2 - x_0) \in \mathsf{U}, \\
\implies\ & 2(x_1 - x_0) + x_0, 2(x_2 - x_0) + x_0 \in A, \\
\implies\ & \tfrac{1}{2}(2(x_1 - x_0) + x_0) + \tfrac{1}{2}(2(x_2 - x_0) + x_0) \in A,
\end{aligned}$$

which gives the result after we notice that

$$\tfrac{1}{2}(2(x_1 - x_0) + x_0) + \tfrac{1}{2}(2(x_2 - x_0) + x_0) = x_1 + x_2 - x_0.$$ ■

We now make the following definition, corresponding to the preceding result.

B.2.6 Definition (Linear part of an affine subspace) Let A be an affine subspace of a $\mathbb{R}$-vector space V, and let $x_0 \in \mathsf{V}$ and a subspace $\mathsf{U} \subseteq \mathsf{V}$ satisfy $\mathsf{A} = x_0 + \mathsf{U}$ as in the preceding proposition. The subspace U is called the ***linear part*** of A and denoted by $L(\mathsf{A})$. •

Now we can characterise the affine hull as the set of affine combinations.

B.2.7 Proposition (The affine hull is the set of affine combinations) *Let* V *be a* $\mathbb{R}$-*vector space, let* $\mathrm{S} \subseteq \mathsf{V}$ *be nonempty, and denote by* $\mathrm{A}(\mathrm{S})$ *the set of affine combinations from* S. *Then* $\mathrm{A}(\mathrm{S}) = \operatorname{aff}(\mathrm{S})$.

Proof We first show that the set of affine combinations is an affine subspace. Choose $x_0 \in S$ and define

$$\mathsf{U}(S) = \{v - x_0 \mid v \in A(S)\}.$$

We first claim that $\mathsf{U}(S)$ is the set of linear combinations of the form

$$\sum_{j=1}^{k} \lambda_j v_j, \qquad k \in \mathbb{Z}_{>0},\ \lambda_1, \dots, \lambda_k \in \mathbb{R},\ \sum_{j=1}^{k} \lambda_j = 0,\ v_1, \dots, v_k \in S. \tag{B.1}$$

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 $u_{k+1} = x_0$. Similarly, consider a linear combination of the form (B.1). We can without loss of generality suppose that $x_0 \in \{v_1, \dots, v_k\}$, since if this is not true we can simply add $0x_0$ to the sum. Thus we suppose, without loss of generality, that $v_k = x_0$. We then have

$$u = \Big(\sum_{j=1}^{k-1} \lambda_j v_j + (\lambda_k + 1)x_0\Big) - x_0.$$

Since the term in the parenthesis is clearly an element of $A(S)$, 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 V. Moreover, it is immediate from Proposition B.2.7 that $A(S)$ is then an affine subspace. Since $\operatorname{aff}(S)$ is the smallest affine subspace containing S it follows that $\operatorname{aff}(S) \subseteq A(S)$. To show that $A(S) \subseteq \operatorname{aff}(S)$ we use induction on the number of elements in an affine combination in $A(S)$. For an affine combination with one term this is obvious. So suppose that every affine combination of the form

$$\sum_{j=1}^{k} \lambda_j v_j, \qquad k \in \{1, \dots, m\},$$

is in $\operatorname{aff}(S)$ and consider an affine combination of the form

$$x = \sum_{j=1}^{m+1} \lambda_j v_j = \sum_{j=1}^{m} \lambda_j v_j + \lambda_{m+1} v_{m+1}.$$

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

$$\sum_{j=1}^{m} \lambda'_j v_j \in \operatorname{aff}(S)$$

by the induction hypothesis. It then holds that

$$x = (1 - \lambda_{m+1}) \sum_{j=1}^{m} \lambda'_j v_j + \lambda_{m+1} v_{m+1}.$$

This is then in $\mathrm{aff}(S)$. ■

B.2.3 Topology of convex sets

Let us now say a few words about the topology of convex sets. In this section we restrict our attention to finite-dimensional $\mathbb{R}$-vector spaces with their standard topology.

Note that every convex set is a subset of its affine hull. Moreover, as a subset of its affine hull, a convex set has an interior.

B.2.8 Definition (Relative interior and relative boundary) If V is a finite-dimensional $\mathbb{R}$-vector space and if $C \subseteq \mathsf{V}$ is a convex set, the set

$$\mathrm{rel\,int}(C) = \{x \in C \mid x \in \mathrm{int}_{\mathrm{aff}(C)}(C)\}$$

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.

B.2.9 Definition (Dimension of a convex set) Let V be a finite-dimensional $\mathbb{R}$-vector space and let $C \subseteq \mathsf{V}$ be convex and let $\mathsf{U} \subseteq \mathsf{V}$ be the subspace for which $\mathrm{aff}(C) = \{x_0 + u \mid u \in \mathsf{U}\}$ for some $x_0 \in \mathsf{V}$. The ***dimension*** of C, denoted by $\dim(C)$, is the dimension of the subspace U. •

The following result will be used in our development.

B.2.10 Proposition (Closures and relative interiors of convex sets are convex sets) *Let* V *be a finite-dimensional* $\mathbb{R}$*-vector space and let* $\mathrm{C} \subseteq \mathsf{V}$ *be convex. Then*

(i) $\mathrm{cl}(\mathrm{C})$ *is convex and*

(ii) $\mathrm{rel\,int}(\mathrm{C})$ *is convex.*

Moreover, $\mathrm{aff}(\mathrm{C}) = \mathrm{aff}(\mathrm{cl}(\mathrm{C}))$ *and* $\mathrm{aff}(\mathrm{K}) = \mathrm{aff}(\mathrm{cl}(\mathrm{K}))$.

Proof For the purposes of the proof we put a norm $\|\cdot\|$ on V; the result and the proof are independent of the choice of this norm.

(i) Let $x, y \in \mathrm{cl}(\mathrm{C})$ and let $s \in [0, 1]$. Suppose that $(x_j)_{j\in\mathbb{Z}_{>0}}$ and $(y_j)_{j\in\mathbb{Z}_{>0}}$ are sequences in C converging to x and y, respectively. Note that $sx_j + (1 - s)y_j \in C$ for each $j \in \mathbb{Z}_{>0}$. Moreover, if $\epsilon > 0$ then

$$\|sx + (1 - s)y - sx_j - (1 - s)y_j\| \leq s\|x - x_j\| + (1 - s)\|y - y_j\| < \epsilon,$$

provided that j is sufficiently large that $s\|x - x_j\| < \frac{\epsilon}{2}$ and $(1-s)\|y - y_j\| < \frac{\epsilon}{2}$. Thus the sequence $(sx_j + (1-s)y_j)_{j\in\mathbb{Z}_{>0}}$ converges to $sx + (1-s)y$ and so $sx + (1-s)y \in \mathrm{cl}(C)$. 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" that for a vector space, we shall assume that $\mathrm{aff}(C) = \mathsf{V}$ and show that $\mathrm{int}(C)$ is convex.

We first prove a lemma.

1 Lemma *If* V *is a finite-dimensional* $\mathbb{R}$*-vector space, if* C *is a convex set, if* $\mathrm{x} \in \mathrm{rel\,int}(\mathrm{C})$*, and if* $\mathrm{y} \in \mathrm{cl}(\mathrm{C})$ *then*

$$[\mathrm{x}, \mathrm{y}) \triangleq \{\mathrm{sx} + (1-\mathrm{s})\mathrm{y} \mid \mathrm{s} \in [0, 1)\}$$

is contained in $\mathrm{rel\,int}(\mathrm{C})$.

Proof As above, let us assume, without loss of generality, that $\mathrm{aff}(C) = \mathsf{V}$. Let us also equip V with a norm. Since $x \in \mathrm{int}(C)$ there exists $r > 0$ such that $\mathsf{B}(x, r) \subseteq C$. Since $y \in \mathrm{cl}(C)$, for every $\epsilon > 0$ there exists $y_\epsilon \in C \cap \mathsf{B}(y, \epsilon)$. 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 $z' \in \mathsf{B}(z, \delta)$ we have

$$\begin{aligned}
& \|z' - z\| < \delta \\
\implies\ & \|z' - (\alpha x + (1-\alpha)y_\epsilon + (1-\alpha)(y - y_\epsilon))\| < \delta \\
\implies\ & \|z' - (\alpha x + (1-\alpha)y_\epsilon)\| \le \delta + (1-\alpha)\epsilon = \alpha r \\
\implies\ & z' \in \{\alpha x' + (1-\alpha)y_\epsilon \mid x' \in \mathsf{B}(x, r)\}.
\end{aligned}$$

Since $y_\epsilon \in C$ and $\mathsf{B}(x, r) \subseteq C$ it follows that $z' \in C$ and so $\mathsf{B}(z, \delta) \subseteq C$. 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 $\mathrm{int}(C)$. ■

The following result will also come up in our constructions.

B.2.11 Proposition (The closure of the relative interior) *If* V *if a finite-dimensional* $\mathbb{R}$*-vector space and if* $\mathrm{C} \subseteq \mathsf{V}$ *is a convex set then* $\mathrm{cl}(\mathrm{rel\,int}(\mathrm{C})) = \mathrm{cl}(\mathrm{C})$.

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 B.2.10 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 $x \in \mathrm{cl}(\mathrm{rel\,int}(C))$. ■

B.2.4 Separation theorems for convex sets

One of the most important properties of convex sets in convex analysis, and indeed for us in our proof of the Maximum Principle, is the notion of certain types of convex sets being separated by hyperplanes. We shall only examine those parts of the theory that we will use; we refer to [Rockafellar 1970] for further discussion.

In this section we again consider subsets of a finite-dimensional $\mathbb{R}$-vector space V. In order to make things clear, let us define all of our terminology precisely.

B.2.12 Definition (Hyperplane, half-space, support hyperplane) Let V be a finite-dimensional $\mathbb{R}$-vector space.

(i) A ***hyperplane*** in V is a subset of the form

$$\{x \in \mathsf{V} \mid \langle \lambda; x \rangle = a\}$$

for some $\lambda \in \mathsf{V}^* \setminus \{0\}$ and $a \in \mathbb{R}$. Such a hyperplane is denoted by $P_{\lambda,a}$.

(ii) A ***half-space*** in V is a subset of the form

$$\{x \in \mathsf{V} \mid \langle \lambda; x \rangle > a\}$$

for some $\lambda \in \mathsf{V}^* \setminus \{0\}$ and $a \in \mathbb{R}$. We shall denote

$$H^-_{\lambda,a} = \{x \in \mathsf{V} \mid \langle \lambda; x \rangle < a\}, \quad H^+_{\lambda,a} = \{x \in \mathsf{V} \mid \langle \lambda; x \rangle > a\}.$$

(iii) If $A \subseteq \mathsf{V}$, a ***support hyperplane*** for A is a hyperplane $P_{\lambda,a}$ such that $A \subseteq H^+_{\lambda,a} \cup P_{\lambda,a}$.

(iv) For subsets $A, B \subseteq \mathsf{V}$, a ***separating hyperplane*** is a hyperplane $P_{\lambda,a}$ for which

$$A \subseteq H^+_{\lambda,a} \cup P_{\lambda,a}, \; B \subseteq H^-_{\lambda,a} \cup P_{\lambda,a}. \qquad \bullet$$

The following result is a basis for many separation theorems for convex sets.

B.2.13 Theorem (Convex sets possess supporting hyperplanes) *If V is a finite-dimensional $\mathbb{R}$-vector space and if $\mathsf{C} \subseteq \mathsf{V}$ is a convex set not equal to V, then C possesses a supporting hyperplane.*

Proof For convenience in the proof we suppose that V is equipped with a norm $\|\cdot\|$ arising from an inner product $\langle \cdot, \cdot \rangle$; the statement of the result and the character of the proof is independent of this choice. We note that the inner product identifies V^* naturally with V, and we make this identification without mention in the proof.

Let $x_0 \notin \mathrm{cl}(C)$, let $z \in C$, and define $r = \|x_0 - z\|$. Define $A = \mathrm{cl}(C) \cap \mathsf{B}(x_0, r)$ noting that A is a nonempty compact set. Define $f \colon A \to \mathbb{R}_{>0}$ by $f(y) = \|x_0 - y\|$. 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 $a = \langle y_0, y_0 - x_0 \rangle$. We will show that $P_{\lambda,a}$ is a support hyperplane for C.

First let us show that $P_{\lambda,a}$ separates $\{x_0\}$ and $\mathrm{cl}(C)$. A direct computation shows that $\langle \lambda, x_0 \rangle = -\|x_0 - y_0\|^2 + a < a$. To show that $\langle \lambda, x \rangle \geq a$ for all $x \in \mathrm{cl}(C)$, suppose otherwise. Thus let $x \in C$ be such that $\langle \lambda, x \rangle < a$. By Lemma 1 in the proof of Proposition B.2.10 the line segment from y to y_0 is contained in $\mathrm{cl}(C)$. Define $g \colon [0, 1] \to \mathbb{R}$ by $g(s) = \|(1-s)y_0 + sy - x_0\|^2$.

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

$$g(s) = (1-s)^2\|y_0 - x_0\|^2 + 2s(1-s)\langle y - x_0, y_0 - x_0\rangle + s^2\|y - x_0\|^2$$

and another computation gives $g'(0) = 2(\langle \lambda, y\rangle - a)$ 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 $\langle \lambda, y\rangle \geq a$ for all $y \in \mathrm{cl}(C)$. ■

During the course of the proof of the theorem we almost proved the following result.

B.2.14 Corollary (Separation of convex sets and points) *If* V *is a finite-dimensional* $\mathbb{R}$*-vector space, if* $\mathsf{C} \subseteq \mathsf{V}$ *is convex, and if* $\mathsf{x}_0 \notin \mathrm{int}(\mathsf{C})$ *then there exists a separating hyperplane for* $\{\mathsf{x}_0\}$ *and* C*.*

Proof If $x_0 \notin \mathrm{cl}(C)$ then the result follows immediately from the proof of Theorem B.2.13. If $x_0 \in \mathrm{bd}(C)$ then let $(x_j)_{j\in\mathbb{Z}_{>0}}$ be a sequence in $\mathsf{V} \setminus \mathrm{cl}(C)$ converging to x_0. For each $j \in \mathbb{Z}_{>0}$ let $\lambda_j \in \mathsf{V}^* \setminus \{0\}$ and $a_j \in \mathbb{R}$ have the property that

$$\begin{aligned} \langle \lambda_j; x_j\rangle &\leq a_j, & & j \in \mathbb{Z}_{>0}, \\ \langle \lambda_j; y\rangle &> a_j, & & y \in C,\ j \in \mathbb{Z}_{>0}. \end{aligned}$$

Let us without loss of generality take $a_j = \langle \lambda_j; x_j\rangle$; this corresponds to choosing the hyperplane separating C from x_j to pass through x_j. Let $\alpha_j = \frac{\lambda_j}{\|\lambda_j\|}$, $j \in \mathbb{Z}_{>0}$. The sequence $(\alpha_j)_{j\in\mathbb{Z}_{>0}}$ is a sequence in the unit sphere in V^* which is compact. Thus we can choose a convergent subsequence which we also denote, by an abuse of notation, by $(\alpha_j)_{j\in\mathbb{Z}_{>0}}$. Let $\alpha \in \mathsf{V}^*$ denote the limit of this sequence. Defining $c_j = \langle \alpha_j; x_j\rangle$ we then have

$$\begin{aligned} \langle \alpha_j; x_j\rangle &= c_j, & & j \in \mathbb{Z}_{>0}, \\ \langle \alpha_j; y\rangle &> c_j, & & y \in C,\ j \in \mathbb{Z}_{>0}. \end{aligned}$$

Let $c = \lim_{j\to\infty} c_j$. For $y \in C$ this gives

$$\begin{aligned} \langle \alpha; x_0\rangle &= \lim_{j\to\infty} \langle \alpha_j; x_j\rangle = c, \\ \langle \alpha; y\rangle &= \lim_{j\to\infty} \langle \alpha_j; y\rangle \geq c, \end{aligned}$$

as desired. ■

The following consequence of Theorem B.2.13 is also of independent interest.

B.2.15 Corollary (Disjoint convex sets are separated) *If* V *is a finite-dimensional* $\mathbb{R}$*-vector space and of* $\mathsf{C}_1, \mathsf{C}_2 \subseteq \mathsf{V}$ *are disjoint convex sets, then there exists a hyperplane separating* C_1 *and* C_2*.*

Proof Define

$$C_1 - C_2 = \{x_1 - x_2 \mid x_1 \in C_1,\ x_2 \in C_2\}.$$

One checks directly that $C_1 - C_2$ is convex. Since C_1 and C_2 are disjoint it follows that $0 \notin C_1 - C_2$. By Theorem B.2.13 there exists a hyperplane P, passing through 0, separating

$C_1 - C_2$ from 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 $\lambda \in \mathsf{V}^* \setminus \{0\}$ such that

$$\langle \lambda; x_1 - x_2 \rangle \geq 0, \qquad x_1 \in C_1,\ x_2 \in C_2.$$

Let

$$a_1 = \inf\{\langle \lambda; x_1 \rangle \mid x_1 \in C_1\}, \quad a_2 = \sup\{\langle \lambda; x_2 \rangle \mid x_2 \in C_2\}$$

so that $a_1 - a_2 \geq 0$. For any $a \in [a_2, a_1]$ we have

$$\begin{aligned} \langle \lambda; x_1 \rangle &\geq a, \qquad x_1 \in C_1, \\ \langle \lambda; x_2 \rangle &\leq a, \qquad x_2 \in C_2, \end{aligned}$$

giving the separation of C_1 and C_2, as desired. ■

We shall require the following quite general result concerning separation of convex sets by hyperplanes.

B.2.16 Theorem (A general separation theorem) *If V is a finite-dimensional $\mathbb{R}$-vector space and if $\mathrm{C}_1, \mathrm{C}_2 \subseteq \mathsf{V}$ 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 $\mathrm{C}_1, \mathrm{C}_2 \subseteq \mathrm{P}$;

(ii) $\mathrm{rel\,int}(\mathrm{C}_1) \cap \mathrm{rel\,int}(\mathrm{C}_2) = \emptyset$.

Proof Suppose that C_1 and C_2 possess a separating hyperplane P. Therefore, there exists $\lambda \in \mathsf{V}^* \setminus \{0\}$ and $a \in \mathbb{R}$ such that

$$\begin{aligned} \langle \lambda; x_1 \rangle &\geq a, \qquad x_1 \in C_1, \\ \langle \lambda; x_2 \rangle &\leq a, \qquad x_2 \in C_2. \end{aligned}$$

If $\langle \lambda; x \rangle = a$ for all $x \in C_1 \cup C_2$ then (i) holds. Now suppose that $\langle \lambda; x_1 \rangle > a$ for some $x_1 \in C_1$ (a similar argument will obviously apply if this holds for some $x_2 \in C_2$) and let $x_0 \in \mathrm{rel\,int}(C_1)$. Since P is a support hyperplane for C_1 and since $C_1 \not\subseteq P$, it follows that the relative interior, and so x_0, lies in the appropriate half-space defined by P. Since P separates C_1 and C_2 this precludes 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 B.2.10 and Corollary B.2.15 it holds that $\mathrm{rel\,int}(C_1)$ and $\mathrm{rel\,int}(C_2)$ possess a separating hyperplane. Thus there exists $\lambda \in \mathsf{V}^* \setminus \{0\}$ and $a \in \mathbb{R}$ such that

$$\begin{aligned} \langle \lambda; x_1 \rangle &\leq a, \qquad x_1 \in \mathrm{rel\,int}(C_1), \\ \langle \lambda; x_2 \rangle &\geq a, \qquad x_2 \in \mathrm{rel\,int}(C_2). \end{aligned}$$

Therefore, by Proposition B.2.11 we also have

$$\begin{aligned} \langle \lambda; x_1 \rangle &\leq a, \qquad x_1 \in \mathrm{cl}(C_1), \\ \langle \lambda; x_2 \rangle &\geq a, \qquad x_2 \in \mathrm{cl}(C_2), \end{aligned}$$

which implies this part of the theorem. ■

Bibliography

Abraham, R., Marsden, J. E., and Ratiu, T. S. [1988] *Manifolds, Tensor Analysis, and Applications*, second edition, number 75 in Applied Mathematical Sciences, Springer-Verlag, ISBN 0-387-96790-7.

Rockafellar, R. T. [1970] *Convex Analysis*, Princeton Mathematical Series, Princeton University Press, Princeton, NJ.