Chapter 2

Convex sets and functions

The class of convex sets plays a central role in functional analysis. The reader may already know that a subset C of the vector space X is said to be **convex** if the following implication holds:

$$t \in (0,1),\ x,y \in C \implies (1-t)x+ty \in C.$$

Thus, a convex set is one that always contains the segment between any two of its points. When X is a normed space, as we assume in this chapter, the triangle inequality implies that open or closed balls are convex; the complement of a ball fails to be convex. A moment's thought confirms as well that convex sets share an important property of closed sets: *an arbitrary intersection of convex sets is convex.*

As we shall see, the possibility of *separating* two disjoint convex sets by a hyperplane is a fundamental issue. The celebrated Hahn-Banach theorem, that bears on this point, is perhaps the single most useful tool in the classical theory. We shall also meet convex *functions* in this chapter. These counterparts of convex sets turn out to be equally important to us later on.

2.1 Properties of convex sets

A **convex combination** of finitely many points $x_1, x_2, \ldots, x_m$ in X ($m \geqslant 2$) means any element x of the form

$$x = \sum_{i=1}^{m} t_i x_i\,,$$

where the *coefficients* t_i of the convex combination are nonnegative numbers summing to one: $t_i \geqslant 0$ and $\sum_{i=1}^{m} t_i = 1$.

2.1 Exercise. The set C is convex if and only if any convex combination of points in C belongs to C. □

F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*, Graduate Texts in Mathematics 264, DOI 10.1007/978-1-4471-4820-3_2,

The following facts are essential, not surprising, yet tricky to prove if one doesn't go about it just right.

2.2 Theorem. *Let C be a convex subset of the normed space X. Then*

(a) $\overline{C}$ *is convex*;

(b) $x \in \overline{C}$, $y \in C^{\circ} \implies (x,y] \subset C^{\circ}$;

(c) C° *is convex*;

(d) $C^{\circ} \neq \emptyset \implies \overline{C} = \overline{C^{\circ}}$ *and* $C^{\circ} = \big(\overline{C}\big)^{\circ}$.

Proof. Let $x, y \in \overline{C}$ and $0 < t < 1$. In order to prove that the point $(1-t)x+ty$ lies in $\overline{C}$ (thus proving (a)), we must show that, given any neighborhood U of 0, the set $(1-t)x+ty+U$ meets C (that is, has nonempty intersection with C). Let V be a neighborhood of 0 such that $(1-t)V+tV \subset U$. Then we have

$$(1-t)x+ty+U \supset (1-t)(x+V)+t(y+V).$$

But $x+V$ and $y+V$ both meet C, since x and y belong to $\overline{C}$. Because C is convex, we obtain the desired conclusion.

We turn now to (b). Let $0 < t < 1$; we wish to show that $(1-t)x+ty \in C^{\circ}$. There is a neighborhood V of 0 satisfying $y+V+V \subset C$. Furthermore, $x-tV/(1-t)$ meets C, so there exist $v \in V$, $c \in C$ such that $x = c+tv/(1-t)$. We then find

$$(1-t)x+ty+tV = (1-t)c+t(y+v+V) \subset (1-t)c+tC \subset C,$$

in light of the convexity of C. Since tV is a neighborhood of 0, the conclusion follows.

The reader will observe that part (c) of the theorem follows immediately from (b). We turn then to (d). Let y be a point in C°, and $x \in \overline{C}$. By part (b), we have the containment $(x,y] \subset C^{\circ}$; this implies $x \in \overline{C^{\circ}}$. The inclusion $\overline{C} \supset \overline{C^{\circ}}$ being evident, the first assertion follows. To prove the other, it suffices to show that

$$x \in \big(\overline{C}\big)^{\circ} \implies x \in C^{\circ}.$$

If x belongs to the set on the left, there is a neighborhood V of 0 such that $x+V \subset \overline{C}$, whence, by part (b), we have

$$(1-t)(x+V)+ty \subset C^{\circ} \;\; \forall t \in (0,1].$$

Now for $t > 0$ sufficiently small, the point $t(x-y)/(1-t)$ belongs to V. For such a value of t, we deduce

$$x = (1-t)\Big(x+\frac{t(x-y)}{1-t}\Big)+ty \in (1-t)(x+V)+ty \subset C^{\circ}.$$

□

2.3 Exercise. Show that the hypothesis $C^{\circ} \neq \emptyset$ in the last assertion of Theorem 2.2 is required for the conclusion, as well as the convexity of C. $\square$

Convex envelopes. Let S be a subset of X. The *convex envelope* of S, denoted $\operatorname{co} S$, is the smallest convex subset of X containing S. This definition is meaningful, since there is at least one convex set containing S (the space X itself), and since the intersection of convex sets is convex; thus, $\operatorname{co} S$ is the intersection of all convex sets containing S. The convex envelope of S can also be described as the set of all convex combinations generated by S:

2.4 Exercise. Show that

$$\operatorname{co} S = \Big\{ \sum_{i=1}^{m} t_i x_i : m \geqslant 1,\, x_i \in S,\ t_i \geqslant 0,\ \sum_{i=1}^{m} t_i = 1 \Big\},$$

and deduce that $\operatorname{co}(S_1 + S_2) \subset \operatorname{co} S_1 + \operatorname{co} S_2$. $\square$

The *closed convex envelope* of S is the smallest closed convex set containing S. It is denoted by $\overline{\operatorname{co}}\, S$. Clearly, it corresponds to the intersection of all closed convex sets containing S.

2.5 Exercise. Prove that $\overline{\operatorname{co}}\, S = \operatorname{cl}(\operatorname{co} S)$. $\square$

The characterization of $\operatorname{co} S$ given in Exer. 2.4 involves arbitrarily large integers m. In finite dimensions, however, this can be improved upon:

2.6 Proposition. (Carathéodory's theorem) *Let S be a subset of a normed space X of finite dimension n. Let $x \in \operatorname{co} S$. Then there is a subset A of S containing at most $n+1$ points such that x is a convex combination of the points of A.*

Proof. Let $x = \sum_0^k t_i x_i$ be a convex combination of $k+1$ elements of S for $k > n$. We proceed to show that x is, in fact, the convex combination of k of these elements, which implies the result.

We may suppose $t_i > 0$, $0 \leqslant i \leqslant k$, for otherwise there is nothing to prove. The vectors $x_i - x_0$ $(1 \leqslant i \leqslant k)$ are linearly dependent in X, since $k > n$. There exist, therefore, scalars r_i $(1 \leqslant i \leqslant k)$, not all zero, such that $\sum_1^k r_i(x_i - x_0) = 0$. Now define $r_0 = -\sum_1^k r_i$. Then we have $\sum_0^k r_i = 0$, $\sum_0^k r_i x_i = 0$. We pick an index j for which r_i/t_i is maximized:

$$r_i/t_i \leqslant r_j/t_j, \quad i = 0, 1, \ldots, k.$$

Then $r_j > 0$ (since the r_i are not all zero and sum to zero). We proceed to set

$$c_i = t_i - r_i t_j / r_j, \ 0 \leqslant i \leqslant k.$$

We then find $c_i \geqslant 0$, $\sum_0^k c_i = 1$, $x = \sum_0^k c_i x_i$ as well as $c_j = 0$, which expresses x in the required way. $\square$

2.7 Exercise. In $\mathbb{R}^2$, let S consist of the points (x,y) on the unit circle that lie in the first quadrant, together with the points $(-1,0)$ and $(0,-1)$. Certain points in $\operatorname{co} S$ can be expressed as a convex combination of two points in S; others require three. Which points require three? □

2.8 Exercise. Let S be a compact subset of $\mathbb{R}^n$. Prove that $\operatorname{co} S$ is compact. □

When the underlying set is convex, the tangents and normals that we met in §1.4 admit alternate characterizations, as we now see. The reader may wish to ponder these in connection with the two of the four sets in Figures 1.1 and 1.2 (pp. 24–25) that are convex.

2.9 Proposition. *Let S be a convex set in X, and let $x \in S$. Then $T_S(x)$ is convex, $S \subset x + T_S(x)$, and we have*

$$T_S(x) = \operatorname{cl}\Big\{ \frac{u-x}{t} : t > 0,\, u \in S \Big\},\ N_S(x) = \big\{ \zeta \in X^* : \langle \zeta, u-x \rangle \leqslant 0 \ \forall u \in S \big\}.$$

Proof. First, we call upon the reader to show (with the help of Theorem 2.2) that the following set W is convex:

$$\operatorname{cl}\Big\{ \frac{u-x}{t} : t > 0,\ u \in S \Big\}.$$

It is clear from the definition of tangent vector that $T_S(x) \subset W$. To prove the opposite inclusion, it suffices to show that any vector of the form $v = (u-x)/t$, where u is in S and $t > 0$, belongs to $T_S(x)$, since the latter is closed. We do this now.

Let ε_i be a positive sequence decreasing to 0. Then, for i sufficiently large, the point $x_i = x + \varepsilon_i(u-x)$ belongs to S, since S is convex. For such i, we have v equal to $(x_i - x)/(t\varepsilon_i)$. This (constant) sequence converges to v, which makes it clear that $v \in T_S(x)$ by definition of the tangent cone. The characterization of $T_S(x)$ is therefore proved; clearly, it implies $S \subset x + T_S(x)$. Finally, since the normal cone is defined by polarity with respect to the tangent cone, the stated expression for $N_S(x)$ is a direct consequence of that characterization as well. □

2.2 Extended-valued functions, semicontinuity

It will be very useful for later purposes to consider functions with values in the **extended reals**; that is, functions $f : X \to \mathbb{R} \cup \{+\infty\}$. The reader is not to suppose by this that we are slipping into informality; the idea is to accommodate such extended-valued functions without lowering our customary standards of rigor.

Notation and terminology: we denote $\mathbb{R} \cup \{+\infty\}$ by $\mathbb{R}_\infty$. The **effective domain** of an extended-valued function f, denoted $\operatorname{dom} f$, is the set

$$\operatorname{dom} f = \{x \in X : f(x) < \infty\}.$$

The function f is called **proper** when $\operatorname{dom} f \neq \emptyset$. The **epigraph** of f is the set of points on or above the graph:

$$\operatorname{epi} f = \{(x,r) \in X \times \mathbb{R} : f(x) \leqslant r\}.$$

The following two types of extended-valued functions play an important role.

2.10 Definition. *Let Σ be a nonempty subset of X^*. The* **support function** *of Σ is the mapping $H_\Sigma : X \to \mathbb{R}_\infty$ defined by*

$$H_\Sigma(x) = \sup_{\sigma \in \Sigma} \langle \sigma, x \rangle, \; x \in X.$$

Let S be a subset of X. Its **indicator function** *$I_S : X \to \mathbb{R}_\infty$ is the function which has value* 0 *on S and $+\infty$ elsewhere.*

It is unreasonable to ask of such functions that they be continuous. A more appropriate regularity property for these and other extended-valued functions that we shall encounter, is the following. We state it in an arbitrary topological space.

2.11 Definition. *Let E be a set endowed with a topology. A function $f : E \to \mathbb{R}_\infty$ is said to be* **lower semicontinuous** (*abbreviated* lsc) *if, for all $c \in \mathbb{R}$, the sublevel set $\{u \in E : f(u) \leqslant c\}$ is closed.*

It is clear that the product of an lsc function f by a positive scalar is lsc. A lower semicontinuous function is locally bounded below, as follows:

2.12 Proposition. *Let $f : E \to \mathbb{R}_\infty$ be lsc. If $x \in \operatorname{dom} f$, then for any $\varepsilon > 0$, there is a neighborhood V of x such that $f(u) > f(x) - \varepsilon \;\; \forall u \in V$. If $f(x) = \infty$, then for any $M \in \mathbb{R}$, there is a neighborhood V of x such that $f(u) > M \;\; \forall u \in V$.*

We omit the elementary proof of this, as well as of the following.

2.13 Proposition.

(a) *A positive linear combination of lsc functions is lsc.*

(b) *A function $f : E \to \mathbb{R}_\infty$ is lsc if and only if* $\operatorname{epi} f$ *is closed in $E \times \mathbb{R}$.*

(c) *The upper envelope of a family of lsc functions is lsc: if f_α is lsc for each index α, then the function f defined by $f(x) = \sup_\alpha f_\alpha(x)$ is lsc.*

A function f such that $-f$ is lower semicontinuous is called *upper semicontinuous*. Because we choose to emphasize minimization and convexity (rather than maximization and concavity), this property will play a lesser role.

Lower semicontinuity is a suitable replacement for continuity in Weierstrass's celebrated result concerning the existence of a minimum:

2.14 Exercise. Let E be a compact topological space and let $f : E \to \mathbb{R}_\infty$ be a proper lsc function. Prove that $\inf_E f$ is finite, and that f attains its minimum on E. $\square$

If f is lsc, and if x_i is a sequence in E converging to x, then it follows easily that $f(x) \leqslant \liminf_{i\to\infty} f(x_i)$. In a metric setting, lower semicontinuity can be characterized in such sequential terms:

2.15 Proposition. *Let E be a metric space, and $f : E \to \mathbb{R}_\infty$. Then f is lsc if and only if for every $x \in E$ and $\ell \in \mathbb{R}$ we have*

$$\lim_{i\to\infty} x_i = x, \quad \lim_{i\to\infty} f(x_i) \leqslant \ell \implies f(x) \leqslant \ell.$$

2.3 Convex functions

The convexity of functions is destined to play an important role in later developments. The reader may as well see the definition immediately. Let $f : X \to \mathbb{R}_\infty$ be a given extended-valued function. We say that f is *convex* if it satisfies

$$f\big((1-t)x+ty\big) \leqslant (1-t)f(x)+tf(y), \quad x,y \in X,\, t \in (0,1).$$

In the inequality (and always in future), we interpret $t \times \infty$ as ∞ for $t > 0$. We seek to avoid the indeterminate expression $0 \times \infty$; this is why $t = 0$ and $t = 1$ were excluded above. With this convention in mind, and by iterating, it follows that f is convex if and only if

$$f\Big(\textstyle\sum_{i=1}^m t_i x_i\Big) \leqslant \textstyle\sum_{i=1}^m t_i f(x_i)$$

for every convex combination of points $x_i \in X$, $m \geqslant 2$, $t_i \geqslant 0$, $\sum_{i=1}^m t_i = 1$.

A function g is called *concave* when the function $-g$ is convex.[1]

2.16 Exercise. Let Y be a vector space, $\Lambda \in L(X,Y)$, and y a point in Y. If the function $g : Y \to \mathbb{R}_\infty$ is convex, then the function $f(x) = g(\Lambda x + y)$ is convex. $\square$

[1] Although X is a normed space throughout this chapter, it is clear that these basic definitions require only that X be a vector space.

2.17 Exercise. Let $f : X \to \mathbb{R}_\infty$ be positively homogeneous and subadditive. Prove that f is convex.[2] $\square$

2.18 Exercise. Let Σ be a nonempty subset of X^*, and S a subset of X.

(a) The support function H_Σ (see Def. 2.10) is positively homogeneous and subadditive (and therefore, convex), as well as proper and lsc.

(b) The indicator function I_S is convex if and only if S is convex, lsc if and only if S is closed, and proper if and only if S is nonempty. $\square$

On occasion it is useful to restrict attention to the values of f on a specified convex subset U of X. We say that f is convex on U if

$$f\big((1-t)x+ty\big) \leqslant (1-t)f(x)+tf(y), \quad x,y \in U,\ t \in (0,1).$$

It is clear that f is convex on U if and only if the function $f+I_U$ (that is, the function which coincides with f on U and which equals ∞ elsewhere) is convex.

We leave as an exercise the proof of the following.

2.19 Proposition. *Let $f : X \to \mathbb{R}_\infty$ be an extended-valued function. Then f is convex if and only if, for every segment $[x,y]$ in X, the function g defined by $g(t) = f((1-t)x+ty)$ is convex on $(0,1)$.*

The class of convex functions is closed under certain operations, in rather similar fashion to lower semicontinuous ones (see Prop. 2.13):

2.20 Proposition.

(a) *A positive linear combination of convex functions is convex.*

(b) *A function $f : E \to \mathbb{R}_\infty$ is convex if and only if* $\operatorname{epi} f$ *is a convex subset of $E \times \mathbb{R}$.*

(c) *The upper envelope of a family of convex functions is convex.*

Proof. We prove only the last assertion. Let the function f be defined as

$$f(x) = \sup_\alpha f_\alpha(x),$$

where, for each α, the function $f_\alpha : X \to \mathbb{R}_\infty$ is convex. Let $x,y \in X$, $t \in (0,1)$ be given. Then

$$\begin{aligned} f\big((1-t)x+ty\big) &= \sup_\alpha f_\alpha\big((1-t)x+ty\big) \leqslant \sup_\alpha \big\{(1-t)f_\alpha(x)+tf_\alpha(y)\big\} \\ &\leqslant (1-t)\sup_\alpha f_\alpha(x)+t\sup_\alpha f_\alpha(y) = (1-t)f(x)+tf(y). \end{aligned}$$ $\square$

[2] We say that f is positively homogeneous if $f(tx) = tf(x)$ whenever t is a positive scalar. Subadditivity is the property $f(x+y) \leqslant f(x)+f(y)$.

2.21 Exercise. If $h\colon \mathbb{R}^n \to \mathbb{R}$ is convex, and if $\theta : \mathbb{R} \to \mathbb{R}$ is convex and increasing, prove that the function $f(x) = \theta(h(x))$ is convex. □

In the course of events, the reader will come to recognize the following recurrent theme: a convex function automatically benefits from a certain regularity, just because it is convex. Here is a first result of this type.

2.22 Proposition. *Let $f : X \to \mathbb{R}_\infty$ be convex, and $x \in \operatorname{dom} f$. Then the directional derivative $f'(x;v)$ exists for every $v \in X$, with values in $[-\infty,+\infty]$, and we have*

$$f'(x;v) = \inf_{t>0} \frac{f(x+tv)-f(x)}{t}.$$

Proof. It suffices to show that the function $g(t) = (f(x+tv)-f(x))/t$ is nondecreasing on the domain $t > 0$. Simply regrouping terms shows that, for $0 < s < t$, we have

$$g(s) \leqslant g(t) \iff f(x+sv) \leqslant (s/t)f(x+tv) + \big(1-(s/t)\big)f(x).$$

But this last inequality holds because f is convex. □

2.23 Example. The function $f : \mathbb{R} \to \mathbb{R}_\infty$ given by

$$f(x) = \begin{cases} -\sqrt{1-x^2} & \text{if } |x| \leqslant 1 \\ +\infty & \text{if } |x| > 1 \end{cases}$$

is convex, with $\operatorname{dom} f = [-1,1]$. We find $f'(1;-1) = -\infty$ and $f'(1;1) = +\infty$. Note that the derivative of f exists in the interior of $\operatorname{dom} f$, but becomes unbounded as one approaches the boundary of $\operatorname{dom} f$. □

2.24 Exercise. Let $f : X \to \mathbb{R}_\infty$ be convex. Prove the following assertions.

(a) f attains a minimum at $x \in \operatorname{dom} f$ if and only if $f'(x;v) \geqslant 0 \ \forall v \in X$.

(b) A finite local minimum of f is a global minimum. □

The necessary condition for a minimum expressed in Prop. 1.39 becomes a *sufficient* condition for optimality when the data are convex, as we now see.

2.25 Proposition. *Let $f : X \to \mathbb{R}$ be convex and differentiable, and let $A \subset X$ be convex. The point $x \in A$ minimizes f over A if and only if $-f'(x) \in N_A(x)$.*

Proof. We know the necessity of the condition $-f'(x) \in N_A(x)$ from Prop. 1.39; there remains to prove that this is a sufficient condition for x to be a solution of the optimization problem $\min_A f$ (when f and A are convex).

Let u be any point in A. Then $v := u - x$ belongs to $T_A(x)$, by Prop. 2.9, and we have (by Prop. 2.22)

$$f(u) - f(x) = f(x+v) - f(x) \geqslant \langle f'(x), v \rangle.$$

This last term is nonnegative, since $-f'(x) \in N_A(x)$, and since the normal cone is the polar of the tangent cone; it follows that $f(u) \geqslant f(x)$. □

Criteria for convexity. The following first and second order conditions given in terms of derivatives are useful for recognizing the convexity of a function.

2.26 Theorem. *Let U be an open convex set in X, and let $f : U \to \mathbb{R}$ be a function which is differentiable at each point of U.*

(a) *f is convex on U if and only if*

$$f(y) - f(x) \geqslant \langle f'(x), y - x \rangle, \quad x, y \in U. \qquad (*)$$

(b) *If in addition f is twice continuously differentiable in U, then f is convex on U if and only if $f''(x)$ is positive semidefinite for every $x \in U$.*

Proof.

(a) Fix any two points $x, y \in U$. If f is convex, then

$$\langle f'(x), y - x \rangle = f'(x; y - x) \leqslant \frac{f\big(x + (y-x)\big) - f(x)}{1},$$

in light of Prop. 2.22, whence $(*)$. Conversely, let us posit $(*)$. For $0 < t < 1$, set $z = (1-t)x + ty$. Then $z \in U$, since U is convex. Invoking $(*)$ reveals

$$f(y) - f(z) \geqslant \langle f'(z), y - z \rangle, \quad f(x) - f(z) \geqslant \langle f'(z), x - z \rangle.$$

We multiply these inequalities by t and $1-t$ respectively, and then add in order to obtain

$$f\big((1-t)x + ty\big) \leqslant (1-t)f(x) + t f(y),$$

which confirms the convexity of f.

(b) Let x, y be distinct points in U. Restricting f to an open segment containing x and y, and applying Lagrange's celebrated theorem (also known as the Taylor expansion with the Lagrange form of the remainder) on the real line, we obtain

$$f(y) - f(x) - \langle f'(x), y - x \rangle = (1/2)\langle f''(z)(y-x), y - x \rangle \text{ for some } z \in (x, y).$$

(Note that $f''(z)$ lies in $L_C(X, X^*)$.) If $f''(\cdot)$ is known to be positive semidefinite everywhere on U, the right side above is nonnegative, and the convexity of f follows from part (a) of the theorem.

For the converse, let us assume that f is convex. Then the left side above is nonnegative, by part (a). For $v \in X$ fixed, set $y = x + tv$, which lies in U when $t > 0$ is sufficiently small. We deduce

$$t^2 \langle f''(z)v, v \rangle \geqslant 0 \text{ for some } z \in (x, x+tv).$$

Dividing by t^2 and letting $t \downarrow 0$, we arrive at $\langle f''(x)v, v \rangle \geqslant 0$. Since $x \in U$ and $v \in X$ are arbitrary, it follows that $f''(\cdot)$ is positive semidefinite on U. ☐

Recall that for a C^2 function $f : \mathbb{R}^n \to \mathbb{R}$, the **Hessian matrix** refers to the $n \times n$ symmetric matrix $\nabla^2 f(x)$ defined by

$$\nabla^2 f(x) = \left[\frac{\partial^2 f}{\partial x_i \, \partial x_j}(x) \right] \quad \big(i, j = 1, 2, \ldots, n\big).$$

It induces the quadratic form corresponding to $f''(x)$. The well-known characterization of positive semidefinite matrices by means of eigenvalues leads to:

2.27 Corollary. *Let U be an open convex subset of $\mathbb{R}^n$, and let $f : U \to \mathbb{R}$ be C^2. Then f is convex on U if and only if, for every $x \in U$, all the eigenvalues of the Hessian matrix $\nabla^2 f(x)$ are nonnegative.*

A consequence of the corollary is that the convexity of a C^2 function f on an interval (a,b) in $\mathbb{R}$ is equivalent to the condition $f''(t) \geqslant 0 \ \forall t \in (a,b)$. This fact immediately implies the inequality used in the proof of Prop. 1.7, as the reader may show. We must emphasize, however, that the convexity of a function of several variables cannot be verified "one variable at a time."

2.28 Exercise. Prove that each of the following three functions is convex separately as a function of x (for each y) and as a function of y (for each x):

$$\exp(x+y), \ \exp(xy), \ \exp x + \exp y.$$

However, at least one of them fails to be convex on $\mathbb{R}^2$ (that is, jointly in (x,y)). Which ones are convex on $\mathbb{R}^2$? ☐

2.29 Exercise. Show that $x \mapsto \ln x$ is concave on the set $x > 0$. Deduce from this the inequality between the geometric and arithmetic means:

$$\sqrt[n]{a_1 \cdot a_2 \cdot \ldots \cdot a_n} \leqslant \frac{a_1 + a_2 + \cdots + a_n}{n} \quad (a_i > 0).$$ ☐

2.30 Example. Integral functionals, which are very important to us, are naturally extended-valued in many cases. Let $\Lambda : [0,1] \times \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be continuous and bounded below. For $x \in X = \mathrm{AC}[0,1]$, we set

$$f(x) = \int_0^1 \Lambda\big(t, x(t), x'(t)\big)\, dt.$$

Under the given hypotheses, the composite function $t \mapsto \Lambda(t, x(t), x'(t))$ is measurable and bounded below, so that its (Lebesgue) integral is well defined, possibly as $+\infty$. When x is identically zero, or more generally, when x is a Lipschitz function, $x'(t)$ is bounded, and it follows that $f(x)$ is finite. Thus $f : X \to \mathbb{R}_\infty$ is proper. However, it is easy to construct an example in which x' is summable but unbounded ($x(t) = t^{1/2}$, say) and for which $f(x) = +\infty$ (take $\Lambda(t,x,v) = v^2$).

We claim that f is lsc. To see this, let x_i be a sequence in X converging to x, with $\lim_{i\to\infty} f(x_i) \leqslant \ell$. (We intend to use the criterion of Prop. 2.15.) Then x_i' converges in $L^1(0,1)$ to x', and $x_i(a) \to x(a)$. It follows easily from this that $x_i(t) \to x(t)$ for each t. Furthermore, there is a subsequence of x_i' (we do not relabel) that converges almost everywhere to x'. Then, by Fatou's lemma, we calculate

$$\begin{aligned} f(x) &= \int_0^1 \Lambda\big(t, x(t), x'(t)\big)\,dt = \int_0^1 \liminf_{i\to\infty} \Lambda\big(t, x_i(t), x_i'(t)\big)\,dt \\ &\leqslant \liminf_{i\to\infty} \int_0^1 \Lambda\big(t, x_i(t), x_i'(t)\big)\,dt = \lim_{i\to\infty} f(x_i) \leqslant \ell, \end{aligned}$$

whence the lower semicontinuity.

We ask the reader to show that if, for each $t \in [0,1]$, the function $(x,v) \mapsto \Lambda(t,x,v)$ is convex, then f is convex.

The integral functional f may be restricted to the subspace $\mathrm{AC}^p[0,1]$ of $\mathrm{AC}[0,1]$, an observation that has some importance when we consider calculating its directional derivatives. Suppose that Λ is C^1, and consider the case $p = \infty$ (thus, only Lipschitz functions x are involved). We suggest that the reader justify the formula

$$f'(x;y) = \int_0^1 \big\{ \Lambda_x(t)y(t) + \Lambda_v(t)y'(t) \big\}\,dt,$$

where $\Lambda_x(t)$ and $\Lambda_v(t)$ denote the partial derivatives of the function Λ evaluated at $(t, x(t), x'(t))$. (The proof involves switching a limit and an integral, a step for which Lebesgue's dominated convergence theorem can be invoked.) As we shall see later, deriving the formula is considerably more delicate when $1 \leqslant p < \infty$. □

The following functional property plays an important role in things to come.

2.31 Definition. (The Lipschitz property) *Let S be a subset of X, and let Y be a normed space. The function $g : S \to Y$ is said to be Lipschitz (of rank K, on S) if*

$$\| g(x) - g(y) \|_Y \leqslant K \| x - y \|_X \quad \forall x, y \in S.$$

It is said to be Lipschitz near x if, for some neighborhood V_x of x and some constant K_x, g is defined and Lipschitz on V_x of rank K_x. Finally, g is called **locally Lipschitz** *on an open set $U \subset X$ if it is Lipschitz near x for every $x \in U$.*

It is easy to see that a linear combination of functions that are Lipschitz on S is Lipschitz on S. Another useful fact is the following: if $f : X \to \mathbb{R}$ is continuously differentiable in a neighborhood of a point x, then f is Lipschitz near x; this is a consequence of the mean value theorem.

2.32 Exercise.

(a) Let $f : X \to \mathbb{R}$ be Lipschitz near each point x of a compact set C. Prove that f is Lipschitz on C.

(b) Let A be a nonempty subset of X. Show that the distance function d_A is Lipschitz of rank 1 on X.

(c) Let $\{ f_\alpha : S \to \mathbb{R} \}_\alpha$ be a family of functions, each of which is Lipschitz of rank K on S, such that the upper envelope $f(x) = \sup_\alpha f_\alpha(x)$ is finite-valued on S. Prove that f is Lipschitz of rank K on S. □

2.33 Proposition. *Let S be a nonempty subset of X, and let $f : S \to \mathbb{R}$ be Lipschitz of rank K. Then there exists $F : X \to \mathbb{R}$ extending f, and which is Lipschitz of rank K on X.*

Proof. The reader is asked to verify (with the help of part (c) of the preceding exercise) that $F(x) := \sup_{y \in S} \{ f(y) - K\|x-y\| \}$ does the trick. □

It turns out that convex functions have a natural predisposition to be continuous, even Lipschitz. This can only happen in the interior of the effective domain, of course. But even there, something more must be postulated. This can be seen by considering a discontinuous linear functional: it is convex, and its effective domain is the whole space, yet it is continuous at no point. The following shows that if a convex function f is "reasonable" at least at *one* point, then it is locally Lipschitz in the interior of $\operatorname{dom} f$.

2.34 Theorem. *Let $f : X \to \mathbb{R}_\infty$ be a convex function which admits a nonempty open set upon which f is bounded above. Then f is locally Lipschitz in the set $\operatorname{int}\operatorname{dom} f$.*

Proof. We require the following:

Lemma 1. *Let $f : X \to \mathbb{R}_\infty$ be convex, and let C be a convex set such that, for certain positive constants δ and N, we have $|f(x)| \leqslant N \;\; \forall x \in C + \delta B$. Then f is Lipschitz on C of rank $2N/\delta$.*

To prove the lemma, let us fix two distinct points x and y in C. The point z defined by $z = y + \delta(y-x)/\|y-x\|$ belongs to $C + \delta B$, and satisfies

$$y = \frac{\delta}{\delta + \|y-x\|}\, x + \frac{\|y-x\|}{\delta + \|y-x\|}\, z .$$

The convexity of f yields

$$f(y) \leqslant \frac{\delta}{\delta + \|y-x\|} f(x) + \frac{\|y-x\|}{\delta + \|y-x\|} f(z),$$

which implies

$$f(y) - f(x) \leqslant \frac{\big[f(z) - f(x)\big]\,\|y-x\|}{\delta + \|y-x\|} \leqslant \frac{2N}{\delta}\|y-x\|.$$

Since x and y are arbitrary points in C, this proves the lemma.

In view of Lemma 1, the theorem is now seen to follow from:

Lemma 2. *Let x_0 be a point such that, for certain numbers M and $\varepsilon > 0$, we have $f(x) \leqslant M \ \forall x \in B(x_0, \varepsilon)$. Then, for any $x \in \operatorname{int}\operatorname{dom} f$, there exists a neighborhood V of x and $N \geqslant 0$ such that $|f(y)| \leqslant N \ \forall y \in V$.*

Without loss of generality, we prove the lemma for $x_0 = 0$. Let $x \in \operatorname{int}\operatorname{dom} f$. There exists $r \in (0,1)$ such that $x/r \in \operatorname{dom} f$. Then

$$V := x + (1-r)B(0,\varepsilon) = B\big(x, (1-r)\varepsilon\big)$$

is a neighborhood of x. Every point u in this neighborhood can be expressed in the form $r(x/r) + (1-r)y$ for some $y \in B(0,\varepsilon)$, whence (by the convexity of f)

$$f(u) \leqslant rf(x/r) + (1-r)M =: M'.$$

Thus f is bounded above by M' on V. Now let $y \in V$. There exists $u \in V$ such that $(y+u)/2 = x$. Then we have

$$f(x) \leqslant (1/2)f(y) + (1/2)f(u) \leqslant (1/2)f(y) + M'/2,$$

which reveals that, on V, f is bounded below by $2f(x) - M'$. Since f is bounded both below and above on V, the required conclusion follows. □

We remark that Theorem 2.34 is false if "bounded above" is replaced by "bounded below." (Consider $f(x) = |\Lambda(x)|$, where Λ is a discontinuous linear functional.)

2.35 Corollary. *If X is finite dimensional, then any convex function $f : X \to \mathbb{R}_\infty$ is locally Lipschitz in the set $\operatorname{int}\operatorname{dom} f$.*

Proof. With no loss of generality, we may take $X = \mathbb{R}^n$. Let x_0 be any point in $\operatorname{int}\operatorname{dom} f$. By the theorem, it suffices to prove that f is bounded above in a neighborhood V of x_0. To see this, observe that, for some $r > 0$, we have

$$V := \operatorname{co}\big\{x_0 \pm re_i\big\}_i \subset \operatorname{dom} f,$$

where the $e_i\,(i=1,2,\ldots,n)$ are the canonical basis vectors in $\mathbb{R}^n$. Then, by the convexity of f, we deduce

$$f(y) \leqslant M := \max_i \big(\,|f(x_0+re_i)|+|f(x_0-re_i)|\,\big) \;\; \forall\, y \in V. \qquad \square$$

The gauge function. A convex set C for which $\operatorname{int} C \neq \emptyset$ is referred to as a *convex body*. For such a set, when $0 \in \operatorname{int} C$, the (Minkowski) **gauge** of C is the function g defined on X as follows:

$$g(x) = \inf\{\,\lambda > 0 : x \in \lambda C\,\}.$$

It is clear that $g(x) \leqslant 1$ if $x \in C$. We claim that $g(x) \geqslant 1$ if $x \notin C$. For suppose the contrary: then there exists $\lambda \in (0,1)$ such that $x/\lambda \in C$. But then

$$x = (1-\lambda)0+\lambda(x/\lambda)$$

expresses x as a convex combination of two points in the convex set C, whence $x \in C$, a contradiction. When $x \notin C$, then, roughly speaking, $g(x)$ is the factor by which the set C must be dilated in order to include the point x.

It is easy to see that the gauge of the unit ball is precisely the norm. The next result may be viewed as a generalization of this fact.

2.36 Theorem. *Let C be a convex subset of the normed space X for which $0 \in \operatorname{int} C$, and let g be its gauge. Then*

(a) *g has values in $[0,\infty)$.*

(b) $g(tx) = t\,g(x) \;\; \forall\, x \in X,\, t \geqslant 0$.

(c) $g(x+y) \leqslant g(x)+g(y) \;\; \forall\, x,y \in X$.

(d) *g is locally Lipschitz (and hence, continuous).*

(e) $\operatorname{int} C = \{\,x : g(x) < 1\,\} \subset C \subset \{\,x : g(x) \leqslant 1\,\} = \operatorname{cl} C$.

Proof. The first two assertions follow easily. If x/λ and y/μ belong to C, then the identity

$$\frac{x+y}{\lambda+\mu} = \frac{\lambda}{\lambda+\mu}\,\frac{x}{\lambda}+\frac{\mu}{\lambda+\mu}\,\frac{y}{\mu}$$

shows that $(x+y)/(\lambda+\mu)$ belongs to C. This observation yields the third assertion (subadditivity). A function which is positively homogeneous and subadditive (as is g) is convex. Further, we have $g(x) \leqslant 1$ on C. It follows from Theorem 2.34 that g is locally Lipschitz. The final assertion is left as an exercise. $\square$

2.4 Separation of convex sets

A set of the form $\{x \in X : \langle \zeta, x \rangle = c\}$, where $0 \neq \zeta \in X^*$ and c is a scalar, is referred to as a *hyperplane*. The sets

$$\{x \in X : \langle \zeta, x \rangle \leqslant c\} \text{ and } \{x \in X : \langle \zeta, x \rangle \geqslant c\}$$

are the associated *halfspaces*. Roughly speaking, we speak of two sets K_1 and K_2 as being *separated* if there is a hyperplane such that K_1 is contained in one of the associated halfspaces, and K_2 in the other.

The reader may be interested to know that the next result, which is known as the *separation theorem*, has often been nominated as the most important theorem in functional analysis.

2.37 Theorem. (Hahn-Banach separation) *Let K_1 and K_2 be nonempty, disjoint convex subsets of the normed space X. They can be separated in the two following cases*:

(a) *If K_1 is open, there exist $\zeta \in X^*$ and $\gamma \in \mathbb{R}$ such that*

$$\langle \zeta, x \rangle < \gamma \leqslant \langle \zeta, y \rangle \;\; \forall x \in K_1,\; y \in K_2\,.$$

(b) *If K_1 is compact and K_2 is closed, there exist $\zeta \in X^*$ and $\gamma_1, \gamma_2 \in \mathbb{R}$ such that*

$$\langle \zeta, x \rangle < \gamma_1 < \gamma_2 < \langle \zeta, y \rangle \;\; \forall x \in K_1,\; y \in K_2\,.$$

The second type of separation above is called *strict*.

Proof.

(a) Fix $\bar{x} \in K_1$ and $\bar{y} \in K_2$, and set $z = \bar{y} - \bar{x}$ and $C = K_1 - K_2 + z$. Then C is an open convex set containing 0; let p be its gauge function. Since $z \notin C$ (because K_1 and K_2 are disjoint), we have $p(z) \geqslant 1$. We prepare an appeal to Theorem 1.32, by defining $L = \mathbb{R}z$ and $\lambda(tz) = t$. We proceed to verify the hypotheses.

If $t \geqslant 0$, then $\lambda(tz) = t \leqslant t\,p(z) = p(tz)$. Consider now the case $t < 0$. Then we evidently have $\lambda(tz) = t \leqslant 0 \leqslant p(tz)$. Thus, we have $\lambda \leqslant p$ on L. Invoking the theorem, we deduce the existence of a linear functional ζ defined on X which extends λ (thus, ζ is nonzero) and which satisfies $\zeta \leqslant p$ on X. In particular, we have $\zeta \leqslant 1$ on C (a neighborhood of 0), which implies that ζ is continuous.

Now let $x \in K_1$, $y \in K_2$. Then $x - y + z \in C$. Bearing in mind that $\langle \zeta, z \rangle$ and $\langle \lambda, z \rangle$ are equal to 1, we calculate

$$\langle \zeta, x \rangle - \langle \zeta, y \rangle + 1 = \langle \zeta, x - y + z \rangle \leqslant p(x - y + z) < 1,$$

whence $\langle \zeta, x \rangle < \langle \zeta, y \rangle$. It follows that $\zeta(K_1)$ and $\zeta(K_2)$ are disjoint convex sets in $\mathbb{R}$ (that is, intervals), with $\zeta(K_1)$ lying to the left of $\zeta(K_2)$. Furthermore, $\zeta(K_1)$ is an *open* interval, by the lemma below. We set $\gamma = \sup \zeta(K_1)$ to obtain the desired conclusion.

Lemma. *Let ζ be a nonzero linear functional on X, and let V be an open subset of X. Then $\zeta(V)$ is an open subset of $\mathbb{R}$.*

To prove the lemma, take any point x such that $\zeta(x) = 1$; we may assume that V is nonempty. Let $\zeta(v)$ (for $v \in V$) be a point in $\zeta(V)$. Since V is open, there exists $\varepsilon > 0$ such that $v + tx \in V$ whenever $|t| < \varepsilon$. Then $\zeta(V)$ contains a neighborhood $(\zeta(v) - \varepsilon, \zeta(v) + \varepsilon)$ of $\zeta(v)$, proving the lemma.

(b) We now examine the second case of the theorem. A routine argument uses the compactness of K_1 in order to derive the existence of an open convex neighborhood V of 0 such that $K_1 + V$ and K_2 are disjoint.[3] We may now apply the first case of the theorem: there exists $\zeta \in X^*$ such that $\zeta(K_1 + V)$ is an interval lying to the left of $\zeta(K_2)$. But $\zeta(K_1)$ is a compact subset of the open interval $\zeta(K_1 + V)$, so that

$$\max \zeta(K_1) < \sup \zeta\big(K_1 + V\big) \leqslant \inf \zeta(K_2).$$

This implies the existence of γ_1, γ_2 as required. □

The conclusion of the separation theorem may fail if the sets K_1 and K_2 do not satisfy the extra hypotheses of either the first or the second case:

2.38 Exercise. Let $X = \ell^2$, and set

$$K_1 = \big\{ x = (x_1, x_2, \dots) \in X : x_i > 0 \;\; \forall i \big\}, \quad K_2 = \ell_c^\infty$$

(see Example 1.6). Show that these sets are disjoint convex subsets of X, but that there is no $\zeta \in X^*$ that satisfies $\langle \zeta, x \rangle < \langle \zeta, y \rangle \;\; \forall x \in K_1,\, y \in K_2$. □

The rest of this section derives some consequences of the separation theorem.

2.39 Theorem. *Let X be a normed space.*

(a) *X^* separates points in X:* $x, y \in X,\, x \neq y \implies \exists\, \zeta \in X^* : \langle \zeta, x \rangle \neq \langle \zeta, y \rangle$.

(b) *Let L be a subspace of X. If $x \notin \overline{L}$, then there exists $\zeta \in X^*$ such that $\langle \zeta, x \rangle = 1$ and $\zeta \equiv 0$ on L. Consequently, if the following implication holds:*

$$\zeta \in X^*,\; \langle \zeta, L \rangle = 0 \implies \zeta = 0,$$

then L is dense in X.

Proof. Left as an exercise. □

[3] Each $x \in K_1$ admits $r(x) > 0$ such that $B(x, 2r(x)) \subset X \setminus K_2$. Let $\{B(x_i, r(x_i))\}$ be a finite subcovering of K_1. Then we may take $V = \cap_i B^\circ(0, r(x_i))$.

Positive linear independence. A given set of vectors $\{\zeta_i : i = 1,2,\dots,k\}$ in X^* is said to be *positively linearly independent* if the following implication holds:

$$\sum_{i=1}^k \lambda_i \zeta_i = 0\,,\ \lambda_i \geqslant 0 \implies \lambda_i = 0\ \forall i \in \{1,2,\dots,k\}.$$

This property is related to the existence of a *decrease direction* v for the given set: an element v satisfying $\langle \zeta_i, v\rangle < 0\ \forall i$. This concept plays an important role in constrained optimization. The *nonexistence* of such a direction is equivalent to positive linear dependence, as we now see.

2.40 Exercise. The goal is to prove the following:

Proposition. *Let $\{\zeta_i : i = 1,2,\dots,k\}$ be a finite subset in X^*. The following are equivalent*:

(a) *There is no $v \in X$ such that $\langle \zeta_i, v\rangle < 0\ \forall i \in \{1,2,\dots,k\}$;*

(b) *The set $\{\zeta_i : i = 1,2,\dots,k\}$ is positively linearly dependent: there exists a nonzero nonnegative vector $\gamma \in \mathbb{R}^k$ such that $\sum_1^k \gamma_i \zeta_i = 0$.*

Show first that (b) $\Longrightarrow$ (a). Now suppose that (a) holds. Why does Theorem 2.37 apply to the sets

$$K_1 = \big\{ y \in \mathbb{R}^k : y_i < 0\ \forall i \in \{1,2,\dots,k\}\big\},$$
$$K_2 = \big\{ \big(\langle \zeta_1, v\rangle, \langle \zeta_2, v\rangle, \dots, \langle \zeta_k, v\rangle\big) : v \in X \big\}?$$

Use separation to deduce (b). □

2.41 Exercise. Let E be a vector space, and let $f_0, f_1, \dots, f_n$ be linear functionals on E. Use the separation theorem to prove that the following are equivalent:

(a) There exists $(\lambda_1,\dots,\lambda_n) \in \mathbb{R}^n$ such that $f_0 = \sum_{i=1}^n \lambda_i f_i$;

(b) There exists $M \geqslant 0$ such that $|f_0(x)| \leqslant M \max_{1\leqslant i\leqslant n} |f_i(x)|\ \forall x \in E$;

(c) $x \in E,\ f_i(x) = 0\ \forall i \in \{1,\dots,n\} \implies f_0(x) = 0$. □

Support functions redux. We defined earlier the support function H_Σ of a subset Σ of X^* (see Def. 2.10), a function that is defined on X. We now consider the support function of a nonempty subset S of X. This refers to the function $H_S : X^* \to \mathbb{R}_\infty$ defined on X^* by

$$H_S(\zeta) = \sup_{x \in S} \langle \zeta, x\rangle,\ \ \zeta \in X^*.$$

The support function transforms certain inclusions into functional inequalities:

2.42 Proposition. *Let C and D be nonempty subsets of X, with D closed and convex. Then $C \subset D$ if and only if $H_C \leqslant H_D$.*

Proof. It is clear that the inclusion implies the inequality. Let us prove the converse by contradiction, supposing therefore that the inequality holds, but that there is a point $\alpha \in C \setminus D$. We may separate the sets $\{\alpha\}$ and D according to the second case of Theorem 2.37: there exists $\zeta \in X^*$ and scalars γ_i such that

$$\langle \zeta, x \rangle < \gamma_1 < \gamma_2 < \langle \zeta, \alpha \rangle \;\; \forall x \in D.$$

(The order of the separation has been reversed, which simply corresponds to replacing ζ by $-\zeta$.) But this implies $H_D(\zeta) < H_C(\zeta)$, a contradiction. □

2.43 Corollary. *Closed convex subsets of X are characterized by their support functions: two closed convex sets coincide if and only if their support functions are equal.*

2.44 Exercise. Let D be a compact convex subset of $\mathbb{R}^n$ and $f : [a,b] \to D$ a measurable function. Prove that

$$\frac{1}{b-a} \int_a^b f(t)\,dt \;\in D.$$

□

2.45 Exercise. Let C and D be nonempty subsets of X, with D closed and convex. Let S be a nonempty bounded subset of X. Prove that

$$C \subset D \iff C+S \subset D+S.$$

Proceed to show that this equivalence is false in general, even in one dimension, if S is not bounded, or if D is not closed, or if D is not convex. □

Separation in finite dimensions. When the underlying space X is finite dimensional, the separation theorem can be refined somewhat, as we now show. Let D be a convex subset of $\mathbb{R}^n$. The geometric content of the following is that there is a hyperplane that passes through any boundary point of D in such a way that D lies entirely in one of the associated halfspaces. The reader is invited to observe that this does *not* correspond to either of the cases treated by Theorem 2.37.

2.46 Proposition. *Let D be a convex subset of $\mathbb{R}^n$, and let α be a point in its boundary: $\alpha \in \partial D$. Then $\alpha \in \partial \overline{D}$, and there exists a nonzero vector $\zeta \in \mathbb{R}^n$ such that*

$$\langle \zeta, x - \alpha \rangle \leqslant 0 \;\; \forall x \in \overline{D}.$$

Proof. By translating, we may do the proof *en français…non, pardon*, we may reduce to the case $\alpha = 0$. We proceed to prove that $0 \in \operatorname{int} \overline{D} \implies 0 \in \operatorname{int} D$; this will establish the first assertion of the proposition.

There exists $r > 0$ such that the points $\pm r e_i$ $(i = 1, 2, \ldots, n)$ lie in $\overline{D}$, where the e_i are the canonical basis vectors of $\mathbb{R}^n$. Let these $2n$ points be denoted by

$x_1, x_2, \ldots, x_{2n}$. Then, for some $\delta > 0$, we have $2\delta B \subset \operatorname{co}\{x_j\}_j$. For each j, let y_j be a point in D satisfying $|y_j - x_j| < \delta$. Then (see Exer. 2.4)

$$\delta B + \delta B = 2\delta B \subset \operatorname{co}\big\{ y_j + (x_j - y_j) \big\}_j \subset \operatorname{co}\{y_j\}_j + \delta B.$$

Since $\operatorname{co}\{y_j\}_j$ is compact and convex (Exer. 2.8), we deduce from this

$$\delta B \subset \operatorname{co}\{y_j\}_j \subset D$$

(see Exer. 2.45). It follows that $0 \in \operatorname{int} D$.

We now prove the second assertion of the proposition. We have $0 \in \partial \overline{D}$, whence there is a sequence x_i of points in $\mathbb{R}^n \backslash \overline{D}$ converging to 0. Let y_i be the closest point in $\overline{D}$ to x_i. It is clear that $y_i \to 0$.

Now set $\zeta_i = x_i - y_i \neq 0$, and extract a subsequence (if necessary) so that $\zeta_i / |\zeta_i|$ converges to a limit ζ. Fix any $x \in \overline{D}$. For $0 < t < 1$, the point $(1-t)y_i + tx$ belongs to $\overline{D}$, since $\overline{D}$ is convex by Theorem 2.2. Since y_i is closest to x_i in $\overline{D}$, we deduce

$$|(1-t)y_i + tx - x_i| \geqslant |y_i - x_i|.$$

Squaring both sides and simplifying, we find $2t\langle \zeta_i, x - y_i \rangle - t^2 |x - y_i|^2 \leqslant 0$. This leads to $\langle \zeta_i, x - y_i \rangle \leqslant 0$, and, in the limit, to $\langle \zeta, x \rangle \leqslant 0$, as required. □

As a corollary, we obtain a third case that can be added to the two of the separation theorem. Note that no openness or compactness hypotheses are made here concerning the sets to be separated; it is the finite dimensionality that compensates.

2.47 Corollary. *If X is finite dimensional, and if K_1 and K_2 are disjoint convex subsets of X, then there exists $\zeta \in X^*$ different from 0 such that*

$$\langle \zeta, x \rangle \leqslant \langle \zeta, y \rangle \;\; \forall x \in K_1,\; y \in K_2.$$

Proof. We may take $X = \mathbb{R}^n$ without loss of generality. Let $D = K_1 - K_2$, a set not containing 0. If 0 is not in the boundary of D, then, for $\varepsilon > 0$ sufficiently small, the sets $K_1 + \varepsilon B$ and K_2 are disjoint; the first case of Theorem 2.37 applies to this situation, and yields the result. If, to the contrary, we have $0 \in \partial D$, then the required conclusion is a direct consequence of Prop. 2.46. □

Existence of nonzero normals. The existence of nonzero normals is closely related to separation. The reader will recall (Prop. 2.9) that when C is convex, and when $\alpha \in C$, the normal cone to C at α is described by

$$N_C(\alpha) = \big\{ \zeta \in X^* : \langle \zeta, x - \alpha \rangle \leqslant 0 \;\; \forall x \in C \big\}.$$

It follows that this normal cone is trivial (that is, reduces to $\{0\}$) when $\alpha \in \operatorname{int} C$. In certain applications, the question of whether $N_C(\alpha)$ is nontrivial for a point α in the boundary of C is crucial. The following result summarizes the two principal cases in which this can be asserted directly.

2.48 Corollary. *Let C be a convex subset of X, and α a point in the boundary of C. Suppose that one of the two following conditions holds*: $\operatorname{int} C \neq \emptyset$, *or X is of finite dimension. Then $N_C(\alpha) \neq \{0\}$.*

Proof. Consider first the case in which X is finite dimensional. Then (for a suitable equivalent norm) it is isometric to $\mathbb{R}^n$ for some positive integer n (Theorem 1.22), via some isometry $T : X \to \mathbb{R}^n$. It follows that $T\alpha \in \partial(TC)$, so that, by Prop. 2.46, there is a nonzero $\zeta \in N_{TC}(T\alpha)$. The following lemma, whose simple proof is omitted, then yields the required assertion.

Lemma. *If $\zeta \in N_{TC}(T\alpha)$, then the formula $\Lambda x = \langle \zeta, Tx \rangle \ \ \forall x \in X$ defines an element $\Lambda \in N_C(\alpha)$.*

(In fact, we have $\Lambda = T^*\zeta$, where T^* is the adjoint of T; see §1.4.)

Consider now the case $\operatorname{int} C \neq \emptyset$. We may then separate the open set $\operatorname{int} C$ (which is convex by Theorem 2.2) from the set $\{\alpha\}$, according to the first case of Theorem 2.37. There results an element ζ of X^* such that

$$\langle \zeta, x \rangle < \langle \zeta, \alpha \rangle \ \forall x \in \operatorname{int} C.$$

Note that ζ is necessarily nonzero. Since $\overline{C} = \overline{\operatorname{int} C}$ (by Theorem 2.2), the preceding inequality implies $\langle \zeta, x \rangle \leqslant \langle \zeta, \alpha \rangle \ \forall x \in \overline{C}$. Thus, ζ is a nonzero element of $N_C(\alpha)$. $\square$

2.49 Exercise. The reader may feel a need to see an example of a closed convex set admitting a boundary point at which the normal cone is trivial; we give one now. Let $X = \ell^2$, and consider

$$C = \big\{ x \in X : |\, x_i | \leqslant 1/i \ \forall i \big\}.$$

Show that C is closed and convex, that $0 \in \partial C$, and that $N_C(0) = \{0\}$. $\square$

http://www.springer.com/978-1-4471-4819-7

Functional Analysis, Calculus of Variations and Optimal Control
Clarke, F.
2013, XIV, 591 p. 24 illus., 8 illus. in color., Hardcover
ISBN: 978-1-4471-4819-7