Appendix C

Convexity

SECTION 1 defines convex sets and functions.
SECTION 2 shows that convex functions defined on subintervals of the real line have left- and right-hand derivatives everywhere.
SECTION 3 shows that convex functions on the real line can be recovered as integrals of their one-sided derivatives.
SECTION 4 shows that convex subsets of Euclidean spaces have nonempty relative interiors.
SECTION 5 derives various facts about separation of convex sets by linear functions.

1. Convex sets and functions

A subset \( C \) of a vector space is said to be convex if it contains all the line segments joining pairs of its points, that is,
\[
αx_1 + (1 - α)x_2 \in C \quad \text{for all } x_1, x_2 \in C \text{ and all } 0 < α < 1.
\]
A real-valued function \( f \) defined on a convex subset \( C \) (of a vector space \( V \)) is said to be convex if
\[
f(αx_1 + (1 - α)x_2) \leq αf(x_1) + (1 - α)f(x_2) \quad \text{for all } x_1, x_2 \in C \text{ and } 0 < α < 1.
\]
Equivalently, the **epigraph** of the function,
\[
\text{epi}(f) := \{(x, t) \in C \times \mathbb{R} : t \geq f(x)\},
\]
is a convex subset of \( C \times \mathbb{R} \). Some authors (such as Rockafellar 1970) define \( f(x) \) to equal \(+\infty\) for \( x \in V \setminus C \), so that the function is convex on the whole of \( V \), and \( \text{epi}(f) \) is a convex subset of \( V \times \mathbb{R} \).

This Appendix will establish several facts about convex functions and sets, mostly for Euclidean spaces. In particular, the facts include the following results as special cases.

(i) For a convex function \( f \) defined at least on an open interval of the real line (possibly the whole real line), there exists a countable collection of linear functions for which \( f(x) = \sup_{i\in\mathbb{N}} (α_i + β_i x) \) on that interval.

(ii) If a real-valued function \( f \) has an increasing, real-valued right-hand derivative at each point of an open interval, then \( f \) is convex on that interval. In particular, if \( f \) is twice differentiable, with \( f'' \geq 0 \), then \( f \) is convex.
2. One-sided derivatives

Let \( f \) be a convex function, defined and real-valued at least on an interval \( J \) of the real line.

Consider any three points \( x_1 < x_2 < x_3 \), all in \( J \). (For the moment, ignore the point \( x_0 \) shown in the picture.) Write \( \alpha \) for \( (x_2 - x_1)/(x_3 - x_1) \), so that \( x_2 = \alpha x_3 + (1 - \alpha)x_1 \). By convexity, \( y_2 := \alpha f(x_3) + (1 - \alpha)f(x_1) \geq f(x_2) \). Write \( S(x_1, x_2) \) for \((f(x_2) - f(x_1))/(x_2 - x_1)\), the slope of the chord joining the points \((x_1, f(x_1))\) and \((x_2, f(x_2))\). Then

\[
S(x_2, x_3) = \frac{f(x_3) - f(x_2)}{x_3 - x_2} \\
\geq \frac{f(x_3) - y_2}{x_3 - x_2} = S(x_1, x_3) = \frac{y_2 - f(x_1)}{x_2 - x_1} \\
\geq \frac{f(x_2) - f(x_1)}{x_2 - x_1} = S(x_1, x_2).
\]

From the second inequality it follows that \( S(x_1, x) \) decreases as \( x \) decreases to \( x_1 \). That is, \( f \) has right-hand derivative \( D^+(x_1) \) at \( x_1 \), if there are points of \( J \) that are larger than \( x_1 \). The limit might equal \(-\infty\), as in the case of the function \( f(x) = -\sqrt{x} \) defined on \( \mathbb{R}^+ \), with \( x_1 = 0 \). However, if there is at least one point \( x_0 \) of \( J \) for which \( x_0 < x_1 \) then the limit \( D^+(x_1) \) must be finite: Replacing \( \{x_1, x_2, x_3\} \) in the argument just made by \( \{x_0, x_1, x_2\} \), we have \( S(x_0, x_1) \leq S(x_1, x_2) \), implying that \(-\infty < S(x_0, x_1) \leq D^+(x_1)\).

The inequality \( S(x_1, x) \leq S(x_1, x_2) \leq S(x_2, x') \) if \( x_1 < x < x_2 < x' \), leads to the conclusion that \( D^+ \) is an increasing function. Moreover, it is continuous from the
right, because
\[
D_+(x_2) \leq S(x_2, x_3) \rightarrow S(x_1, x_3) \quad \text{as } x_2 \downarrow x_1, \text{ for fixed } x_3
\]
\[
\rightarrow D_+(x_1) \quad \text{as } x_3 \downarrow x_1.
\]

Analogous arguments show that \(S(x_0, x_1)\) increases to a limit \(D_-(x_1)\) as \(x_0\) increases to \(x_1\). That is, \(f\) has left-hand derivative \(D_1(x_1)\) at \(x_1\), if there are points of \(J\) that are smaller than \(x_1\).

If \(x_1\) is an interior point of \(J\) then both left-hand and right-hand derivatives exist, and \(D_-(x_1) \leq D_+(x_1)\). The inequality may be strict, as in the case where \(f(x) = |x|\) with \(x_1 = 0\). The left-hand derivative has properties analogous to those of the right-hand derivative. The following Theorem summarizes.

<1> **Theorem.** Let \(f\) be a convex, real-valued function defined (at least) on a bounded interval \([a, b]\) of the real line. The following properties hold.

(i) The right-hand derivative \(D_+(x)\) exists,
\[
\frac{f(y) - f(x)}{y - x} \downarrow D_+(x) \quad \text{as } y \downarrow x,
\]
for each \(x\) in \([a, b]\). The function \(D_+(x)\) is increasing and right-continuous on \([a, b]\). It is finite for \(a < x < b\), but \(D_+(a)\) might possibly equal \(-\infty\).

(ii) The left-hand derivative \(D_-(x)\) exists,
\[
\frac{f(x) - f(z)}{x - z} \uparrow D_-(x) \quad \text{as } z \uparrow x,
\]
for each \(x\) in \((a, b]\). The function \(D_-(x)\) is increasing and left-continuous function on \((a, b]\). It is finite for \(a < x < b\), but \(D_-(b)\) might possibly equal \(+\infty\).

(iii) For \(a \leq x < y \leq b\),
\[
D_+(x) \leq \frac{f(y) - f(x)}{y - x} \leq D_-(y).
\]

(iv) \(D_-(x) \leq D_+(x)\) for each \(x\) in \((a, b]\), and
\[
f(w) \geq f(x) + c(w - x) \quad \text{for all } w \text{ in } [a, b],
\]
for each real \(c\) with \(D_-(x) \leq c \leq D_+(x)\).

**Proof.** Only the second part of assertion (iv) remains to be proved. For \(w > x\) use
\[
\frac{f(w) - f(x)}{w - x} = S(x, w) \geq D_+(x) \geq c;
\]
for \(w < x\) use
\[
\frac{f(x) - f(w)}{x - w} = S(w, x) \leq D_-(x) \leq c,
\]

\(\square\) where \(S(\cdot, \cdot)\) denotes the slope function, as above.

<2> **Corollary.** If a convex function \(f\) on a convex subset \(C \subseteq \mathbb{R}^n\) has a local minimum at a point \(x_0\), that is, if \(f(x) \geq f(x_0)\) for all \(x\) in a neighborhood of \(x_0\), then \(f(w) \geq f(x_0)\) for all \(w\) in \(C\).
3. Integral representations

Convex functions on the real line are expressible as integrals of one-sided derivatives.

Proof. Consider first the case $n = 1$. Suppose $w \in C$ with $w > x_0$. The right-hand derivative $D_+(x_0) = \lim_{y \to 0^+} (f(y) - f(x_0)) / (y - x_0)$ must be nonnegative, because $f(y) \geq f(x_0)$ for $y$ near $x_0$. Assertion (iv) of the Theorem then gives

$$f(w) \geq f(x_0) + (w - x_0)D_+(x_0) \geq f(x_0).$$

The argument for $w < x_0$ is similar.

For general $\mathbb{R}^n$, apply the result for $\mathbb{R}$ along each straight line through $x_0$.

Existence of finite left-hand and right-hand derivatives ensures that $f$ is continuous at each point of the open interval $(a, b)$. It might not be continuous at the endpoints, as shown by the example

$$f(x) = \begin{cases} -\sqrt{x} & \text{for } x > 0 \\ 1 & \text{for } x = 0. \end{cases}$$

Of course, we could recover continuity by redefining $f(0)$ to equal 0, the value of the limit $\lim_{w \to 0^+} f(w)$.

Corollary. Let $f$ be a convex, real-valued function on an interval $[a, b]$. There exists a countable collection of linear functions $d_i + c_i w$, for which the convex function $\psi(w) := \sup_{i \in \mathbb{N}}(d_i + c_i w)$ is everywhere $\leq f(w)$, with equality except possibly at the endpoints $w = a$ or $w = b$, where $\psi(a) = f(a^+)$ and $\psi(b) = f(b^-)$.

Proof. Let $X_0 := \{x_i : i \in \mathbb{N}\}$ be a countable dense subset of $(a, b)$. Define $c_i := D_+(x_i)$ and $d_i := f(x_i) - c_i x_i$. By assertion (iv) of the Theorem, $f(w) \geq d_i + c_i w$ for $a \leq w \leq b$ for each $i$, and hence $f(w) \geq \psi(w)$.

If $a < w < b$ then (iv) also implies that $f(x_i) \geq f(w) + (x_i - w)D_+(w)$, and hence

$$\psi(w) \geq f(x_i) + c_i(w - x_i) \geq f(w) - (x_i - w)(D_+(x_i) - D_+(w)) \quad \text{for all } x_i.$$
C.3 Integral representations

Theorem. If \( f \) is real-valued and convex on \([a, b]\), with \( f(a) = f(a^+) \) and \( f(b) = f(b^-) \), then both \( D_+(x) \) and \( D_-(x) \) are integrable with respect to Lebesgue measure on \([a, b]\), and

\[
f(x) = f(a) + \int_a^x D_+(t) \, dt = f(a) + \int_a^x D_-(t) \, dt \quad \text{for } a \leq x \leq b.
\]

Proof. Choose \( \alpha \) and \( \beta \) with \( a < \alpha < \beta < x \). For a positive integer \( n \), define \( \delta := (\beta - \alpha)/n \) and \( x_i := \alpha + i\delta \) for \( i = 0, 1, \ldots, n \). Both \( D_+ \) and \( D_- \) are bounded on \([\alpha, \beta]\). For \( i = 2, \ldots, n-1 \), part (iii) of Theorem <1> and monotonicity of both one-sided derivatives gives

\[
\int_{x_{i-2}}^{x_{i-1}} D_+(t) \, dt \leq \delta D_+(x_{i-1}) \leq f(x_i) - f(x_{i-1}) \leq \delta D_-(x_i) \leq \int_{x_i}^{x_{i+1}} D_-(t) \, dt,
\]

which sums to give

\[
\int_a^{x_{n-2}} D_+(t) \, dt \leq f(x_{n-1}) - f(x_1) \leq \int_a^b D_-(t) \, dt.
\]

Let \( n \) tend to infinity, invoking Dominated Convergence and continuity of \( f \), to deduce that \( \int_a^b D_+(t) \, dt \leq f(\beta) - f(\alpha) \leq \int_a^b D_-(t) \, dt \). Both inequalities must actually be equalities, because \( D_-(t) \leq D_+(t) \) for all \( t \in (a, b) \).

Let \( \alpha \) decrease to \( a \). Monotone Convergence—the functions \( D_\pm \) are bounded above by \( D_+(\beta) \) on \([a, \beta]\)—and continuity of \( f \) at \( a \) give \( f(\beta) - f(\alpha) = \int_a^\beta D_+(t) \, dt = \int_a^\beta D_-(t) \, dt \). In particular, the negative parts of both \( D_\pm \) are integrable. Then let \( \beta \) increase to \( x \) to deduce, via a similar argument, the asserted integral expressions for \( f(x) - f(a) \), and the integrability of \( D_\pm \) on \([a, b]\).

Conversely, suppose \( f \) is a continuous function defined on an interval \([a, b]\), with an increasing, real-valued right-hand derivative \( D_+(t) \) existing at each point of \([a, b]\). On each closed proper subinterval \([a, x]\), the function \( D_+ \) is bounded, and hence Lebesgue integrable. From Section 3.4, \( f(x) = \int_a^x D_+(t) \, dt \) for all \( a \leq x < b \). Equality for \( x = b \) also follows, by continuity and Monotone Convergence. A simple argument will show that \( f \) is then convex on \([a, b]\).

More generally, suppose \( D \) is an increasing, real-valued function defined (at least) on \([a, b]\). Define \( g(x) := \int_a^x D(t) \, dt \), for \( a \leq x \leq b \). (Possibly \( g(b) = \infty \).) Then \( g \) is convex. For \( a \leq x_0 < x_1 \leq b \) and \( 0 < \alpha < 1 \) and \( x_\alpha := (1-\alpha)x_0 + \alpha x_1 \), then

\[
(1-\alpha)g(x_0) + \alpha g(x_1) - g(x_\alpha)
= \int_a^b \left( (1-\alpha)[t \leq x_0] + \alpha[t \leq x_1] - [t \leq x_\alpha] \right) D(t) \, dt
= \int_a^b \left( \alpha[x_\alpha < t \leq x_1] - (1-\alpha)[x_0 < t \leq x_\alpha] \right) D(t) \, dt
\geq (\alpha(x_1 - x_\alpha) - (1-\alpha)(x_\alpha - x_0)) D(x_\alpha) = 0.
\]

Example. Let \( f \) be a twice continuously differentiable (actually, absolute continuity of \( f' \) would suffice) convex function, defined on a convex interval \( J \subseteq \mathbb{R} \)
that contains the origin. Suppose \( f(0) = f'(0) = 0 \). The representations
\[
f(x) = x \int_{0 \leq s \leq 1} f'(xs) \, ds = x^2 \int_{0 \leq t \leq s \leq 1} f''(xt) \, dt \, ds = x^2 \int_{0}^{1} (1-t) f''(xt) \, dt,
\]
establish the following facts.

(i) The function \( f(x)/x \) is increasing.
(ii) The function \( \phi(x) := 2f(x)/x^2 \) is nonnegative and convex.
(iii) If \( f'' \) is increasing then so is \( \phi \).

Moreover, Jensen's inequality for the uniform distribution \( \lambda \) on the triangular region \( [0 \leq t \leq s \leq 1] \) implies that
\[
\phi(x) = \lambda^{s,t} f''(xt) \geq f''(\lambda^{s,t} xt) = f''(x/3).
\]

Two special cases of these results were needed in Chapter 10, to establish the Bennett inequality and to establish Kolmogorov's exponential lower bound. The choice \( f(x) := e^x - 1 - x \), with \( f''(x) = e^x \), leads to the conclusion that the function
\[
\Delta(x) := \begin{cases} 
  e^x - 1 - x & \text{for } x \neq 0 \\
  1 & \text{for } x = 0
\end{cases}
\]
is nonnegative and increasing over the whole real line. The choice \( f(x) := (1 + x) \log(1 + x) - x \), for \( x \geq -1 \), with \( f'(x) = \log(1 + x) \) and \( f''(x) = (1 + x)^{-1} \), leads to the conclusion that the function
\[
\psi(x) := \begin{cases} 
  (1 + x) \log(1 + x) - x & \text{for } x \geq -1 \text{ and } x \neq 0 \\
  1 & \text{for } x = 0.
\end{cases}
\]
is nonnegative, convex, and decreasing. Also \( x\psi(x) \) is increasing on \( \mathbb{R}^+ \), and
\[
\psi(x) \geq (1 + x/3)^{-1}.
\]

4. Relative interior of a convex set

Convex subsets of Euclidean spaces either have interior points, or they can be regarded as embedded in lower dimensional subspaces within which they have interior points.

<6> Theorem. Let \( C \) be a convex subset of \( \mathbb{R}^n \).

(i) There exists a smallest subspace \( \mathcal{V} \) for which \( C \subseteq x_0 \oplus \mathcal{V} := \{x_0 + x : x \in \mathcal{V}\} \), for each \( x_0 \in C \).

(ii) \( \dim(\mathcal{V}) = n \) if and only if \( C \) has a nonempty interior.

(iii) If \( \text{int}(C) \neq \emptyset \), there exists a convex, nonnegative function \( \rho \) defined on \( \mathbb{R}^n \) for which \( \text{int}(C) = \{x : \rho(x) < 1\} \subseteq C \subseteq \{x : \rho(x) \leq 1\} = \text{int}(C) \).

Proof. With no loss of generality, suppose \( 0 \in C \). Let \( x_1, \ldots, x_k \) be a maximal set of linearly independent vectors from \( C \), and let \( \mathcal{V} \) be the subspace spanned by those vectors. Clearly \( C \subseteq \mathcal{V} \). If \( k < n \), there exists a unit vector \( w \) orthogonal to \( \mathcal{V} \), and every point \( x \) of \( \mathcal{V} \) is a limit of points \( x + tw \) not in \( \mathcal{V} \). Thus \( C \) has an empty interior.
5. Separation of convex sets by linear functionals

If \( k = n \), write \( \bar{x} \) for \( \sum x_i/n \). Each member of the usual orthonormal basis has a representation as a linear combination, \( e_i = \sum a_i j x_j \). Choose an \( \epsilon > 0 \) for which 
\[
2n\epsilon \left( \sum_l a_i^2 \right)^{1/2} < 1 \text{ for every } j.
\]
For every \( y := \sum y_i e_i \) in \( \mathbb{R}^n \) with \( |y| < \epsilon \), the coefficients \( \beta_j := (2n)^{-1} + \sum a_i j y_i \) are positive, summing to a quantity \( 1 - \beta_0 \leq 1 \), and \( \bar{x}/2 + y = \beta_0 0 + \sum j \beta_j x_j \in C \). Thus \( \bar{x}/2 \) is an interior point of \( C \).

If \( \text{int}(C) \neq \emptyset \), we may, with no loss of generality, suppose 0 is an interior point. Define a map \( \rho : \mathbb{R}^n \to \mathbb{R}^+ \) by \( \rho(z) := \inf\{t > 0 : z/t \in C \} \). It is easy to see that \( \rho(0) = 0 \), and \( \rho(\alpha y) = \alpha \rho(y) \) for \( \alpha > 0 \). Convexity of \( C \) implies that \( \rho(z_1 + z_2) \leq \rho(z_1) + \rho(z_2) \) for all \( z_j \): if \( z_j/t_j \in C \) then
\[
\frac{z_1 + z_2}{t_1 + t_2} = \frac{t_1}{t_1 + t_2} \left( \frac{z_1}{t_1} \right) + \frac{t_2}{t_1 + t_2} \left( \frac{z_2}{t_2} \right) \in C.
\]
In particular, \( \rho \) is a convex function. Also \( \rho \) satisfies a Lipschitz condition: if \( y = \sum y_i e_i \) and \( z = \sum z_i e_i \) then
\[
\rho(y) - \rho(z) \leq \rho(y - z) = \rho \left( \sum_i (y_i - z_i) e_i \right) \leq \sum_i \left( (y_i - z_i)^+ \rho(e_i) + (y_i - z_i)^- \rho(-e_i) \right) \leq |y - z| \left( \sum_i \rho(e_i)^2 + \rho(-e_i)^2 \right)^{1/2}.
\]
Thus \( \{ \rho < 1 \} \) is open and \( \{ \rho \leq 1 \} \) is closed.

Clearly \( \rho(x) \leq 1 \) for every \( x \) in \( C \); and if \( \rho(x) < 1 \) then \( x_0 := x/t \in C \) for some \( t < 1 \), implying \( x = (1 - t)0 + tx_0 \in C \). Thus \( \{ z : \rho(z) < 1 \} \subseteq C \subseteq \{ z : \rho(z) \leq 1 \} \). Every point \( x \) with \( \rho(x) = 1 \) lies on the boundary, being a limit of points \( x(1 \pm n^{-1}) \) from \( C \) and \( C^c \). Assertion (iii) follows.

If \( C \subseteq x_0 \oplus \mathbb{V} \subseteq \mathbb{R}^n \), with \( \dim(\mathbb{V}) = k < n \), we can identify \( \mathbb{V} \) with \( \mathbb{R}^k \) and \( C \) with a subset of \( \mathbb{R}^k \). By part (ii) of the Theorem, \( C \) has a nonempty interior, as a subset of \( x_0 \oplus \mathbb{V} \). That is, there exist points \( x \) of \( C \) with open neighborhoods (in \( \mathbb{R}^n \)) for which \( \mathbb{N} \cap (x_0 \oplus \mathbb{V}) \subseteq C \). The set of all such points is called the **relative interior** of \( C \), and is denoted by \( \text{rel-int}(C) \). Part (iii) of the Theorem has an immediate extension,
\[
\text{rel-int}(C) \subseteq C \subseteq \overline{\text{rel-int}(C)},
\]
with a corresponding representation via a convex function \( \rho \) defined only on \( x_0 \oplus \mathbb{V} \).

5. Separation of convex sets by linear functionals

The theorems asserting existence on separating linear functionals depend on the following simple extension result.

**Lemma.** Let \( f \) be a real-valued convex function, defined on a vector space \( \mathbb{V} \). Let \( T_0 \) be a linear functional defined on a vector subspace \( \mathbb{V}_0 \), on which \( T_0(x) \leq f(x) \) for all \( x \in \mathbb{V}_0 \). Let \( y_1 \) be a point of \( \mathbb{V} \) not in \( \mathbb{V}_0 \). There exists an extension of \( T_0 \) to a linear functional \( T_1 \) on the subspace \( \mathbb{V}_1 \) spanned by \( \mathbb{V}_0 \cup \{ y_1 \} \) for which \( T_1(z) \leq f(z) \) on \( \mathbb{V}_1 \).
Corollary. Let \( C_1 \) and \( C_2 \) be disjoint convex subsets of \( \mathbb{R}^n \). Then there is a nonzero linear functional for which \( \inf_{x \in C_1} T(x) \geq \sup_{x \in C_2} T(x) \).

Proof. Each point \( z \) in \( V_1 \) has a unique representation \( z := x + ry_1 \), for some \( x \in V_0 \) and some \( r \in \mathbb{R} \). We need to find a value for \( T_1(y_1) \) which for \( f(x + ry_1) \geq T_0(x) + rT_1(y_1) \) for all \( r \in \mathbb{R} \). Equivalently we need a real number \( c \) such that
\[
\inf_{x \in V_0, t > 0} \frac{f(x + ty_1) - T_0(x)}{t} \geq c \geq \sup_{x \in V_0, x > 0} \frac{T_0(x) - f(x - sy_1)}{s},
\]
for then \( T_1(y_1) := c \) will give the desired extension.

For given \( x_0, x_1 \) in \( V_0 \) and \( s, t > 0 \), define \( \alpha := s/(s+t) \) and \( x_a := \alpha x_0 + (1-\alpha)x_1 \). Then, by convexity of \( f \) on \( V_1 \) and linearity of \( T_0 \) on \( V_0 \),
\[
\frac{s}{s+t} f(x_0 + ty_1) + \frac{t}{s+t} f(x - sy_1) \geq f(x_a) \geq T_0(x_a) = \frac{s}{s+t} T_0(x_0) + \frac{t}{s+t} T_0(x_1),
\]
which implies
\[
-\infty > \frac{f(x_0 + ty_1) - T_0(x_0)}{t} \geq \frac{T_0(x_1) - f(x - sy_1)}{s} > -\infty.
\]
The infimum over \( x_0 \) and \( t > 0 \) on the left-hand side must be greater than or equal to the supremum over \( x_1 \) and \( s > 0 \) on the right-hand side, and both bounds must be finite. Existence of the desired real \( c \) follows.

Remark. The vector space \( V \) need not be finite dimensional. We can order extensions of \( T_0 \), bounded above by \( f \), by defining \( (T_0, V_0) \succeq (T_1, V_1) \) to mean that \( V_0 \) is a subspace of \( V_1 \), and \( T_0 \) is an extension of \( T_1 \). Zorn’s lemma gives a maximal element of the set of extensions \( (T_n, V_n) \succeq (T_0, V_0) \). Lemma <7> shows that \( V_n \) must equal the whole of \( V \), otherwise there would be a further extension. That is, \( T_0 \) has an extension to a linear functional \( T \) defined on \( V \) with \( T(x) \leq f(x) \) for every \( x \) in \( V \). This result is a minor variation on the Hahn-Banach theorem from functional analysis (compare with page 62 of Dunford & Schwartz 1958).

Theorem. Let \( C \) be a convex subset of \( \mathbb{R}^n \) and \( y_0 \) be a point not in \( \text{rel-int}(C) \).

(i) There exists a linear functional \( T \) on \( \mathbb{R}^k \) for which \( 0 \neq T(y_0) \geq \sup_{x \in C} T(x) \).

(ii) If \( y_0 \notin \overline{C} \), then we may choose \( T \) so that \( T(y_0) > \sup_{x \in C} T(x) \).

Proof. With no loss of generality, suppose \( 0 \in C \). Let \( V \) denote the subspace spanned by \( C \), as in Theorem <6>. If \( y_0 \notin V \), let \( \ell \) be its component orthogonal to \( V \). Then \( y_0 \cdot \ell > 0 = x \cdot \ell \) for all \( x \) in \( C \).

If \( y_0 \in V \), the problem reduces to construction of a suitable linear functional \( T \) on \( V \): we then have only to define \( T(z) := 0 \) for \( z \notin V \) to complete the proof. Equivalently, we may suppose that \( V = \mathbb{R}^n \). Define \( T_0 \) on \( V_0 := \{r x_0 : r \in \mathbb{R}\} \) by \( T(r y_0) := r f(y_0) \), for the \( f \) defined in Theorem <6>. Note that \( T_0(y_0) = f(y_0) \geq 1 \), because \( y_0 \notin \text{rel-int}(C) = \{\rho < 1\} \). Clearly \( T_0(x) \leq \rho(x) \) for all \( x \in V_0 \). Invoke Lemma <7> repeatedly to extend \( T_0 \) to a linear functional \( T \) on \( \mathbb{R}^n \), with \( T(x) \leq \rho(x) \) for all \( x \in \mathbb{R}^n \). In particular,
\[
T(y_0) \geq 1 \geq \rho(x) \geq T(x) \quad \text{for all } x \in \overline{C} = \{\rho \leq 1\}.
\]

For (ii), note that \( T(y_0) > 1 \) if \( y_0 \notin \overline{C} \).

Corollary. Let \( C_1 \) and \( C_2 \) be disjoint convex subsets of \( \mathbb{R}^n \). Then there is a nonzero linear functional for which \( \inf_{x \in C_1} T(x) \geq \sup_{x \in C_2} T(x) \).
Corollary. For each closed convex subset $F$ of $\mathbb{R}^n$ there exists a countable family of closed halfspaces $\{H_i : i \in \mathbb{N}\}$ for which $F = \cap_{i \in \mathbb{N}} H_i$.

Proof. Let $\{x_i : i \in \mathbb{N}\}$ be a countable dense subset of $F^c$. Define $r_i$ as the distance from $x_i$ to $F$, which is strictly positive for every $i$, because $F^c$ is open. The open ball $B(x_i, r_i)$ with radius $r_i$ and center $x_i$ is convex and disjoint from $F$. From the previous Corollary, there exists a unit vector $\ell_i$ and a constant $k_i$ for which $\ell_i \cdot y \geq k_i \geq \ell_i \cdot x$ for all $y \in B(x_i, r_i)$ and all $x \in F$. Define $H_i := \{x \in \mathbb{R}^n : \ell_i \cdot x \leq k_i\}$.

Each $x$ in $F^c$ is the center of some open ball $B(x, 3\epsilon)$ disjoint from $F$. There is an $x_i$ with $|x - x_i| < \epsilon$. We then have $r_i \geq 2\epsilon$, because $B(x, 3\epsilon) \supseteq B(x_i, 2\epsilon)$, and hence $x - \epsilon \ell_i \in B(x_i, r_i)$. The separation inequality $\ell_i \cdot (x - \epsilon \ell_i) \geq k_i$ then implies $\ell_i \cdot x > k_i$, that is $x \notin H_i$.

Corollary. Let $f$ be a convex (real-valued) function defined on a convex subset $C$ of $\mathbb{R}^n$, such that $\text{epi}(f)$ is a closed subset of $\mathbb{R}^{n+1}$. Then there exist $(d_i : i \in \mathbb{N}) \subseteq \mathbb{R}^n$ and $(c_i : i \in \mathbb{N}) \subseteq \mathbb{R}$ such that $f(x) = \sup_{i \in \mathbb{N}} (c_i + d_i \cdot x)$ for every $x$ in $C$.

Proof. From the previous Corollary, and the definition of $\text{epi}(f)$, there exist $\ell_i \in \mathbb{R}^n$ and constants $\alpha_i, k_i \in \mathbb{R}$ such that

$$\infty > t \geq f(x)$$
if and only if $k_i \geq \ell_i \cdot x - t\alpha_i$ for all $i \in \mathbb{N}$.

The $i$th inequality can hold for arbitrarily large $t$ only if $\alpha_i \geq 0$. Define $\psi(x) := \sup_{t > 0} (\ell_i \cdot x - k_i) / \alpha_i$. Clearly $f(x) \geq \psi(x)$ for $x \in C$. If $s < f(x)$ for an $x$ in $C$ then there must exist an $i$ for which $\ell_i \cdot x - f(x)\alpha_i \leq k_i < \ell_i \cdot x - s\alpha_i$, thereby forcing $\alpha_i > 0$ and $s < \psi(x)$.

6. Problems

[1] Let $f$ be the convex function, taking values in $\mathbb{R} \cup \{\infty\}$, defined by

$$f(x, y) = \begin{cases} -y^{1/2} & \text{for } 0 \leq 1 \text{ and } x \in \mathbb{R} \\ \infty & \text{otherwise.} \end{cases}$$

Let $T_0$ denote the linear functional defined on the $x$-axis by $T_0(x, 0) := 0$ for all $x \in \mathbb{R}$. Show that $T_0$ has no extension to a linear functional on $\mathbb{R}^2$ for which $T(x, y) \leq f(x, y)$ everywhere, even though $T_0 \leq f$ along the $x$-axis.

[2] Suppose $X$ is a random variable for which the moment generating function, $M(t) := \mathbb{E}\exp(tX)$, exists (and is finite) for $t$ in an open interval $J$ about the origin of the real line. Write $\mathbb{P}_t$ for the probability measure with density $\exp^{X}/M(t)$ with respect to $\mathbb{P}$, for $t \in J$, with corresponding variance $\text{var}_t(\cdot)$. Define $\Lambda(t) := \log M(t)$.

(i) Use Dominated Convergence to justify the operations needed to show that

$$\Lambda'(t) = M'(t)/M(t) = \mathbb{P}(X\exp^{X}/M(t)) = \mathbb{P}_t X,$$

$$\Lambda''(t) = (M(t)M''(t) - M'(t)^2)/M(t)^2 = \text{var}_t(X).$$
Appendix C: Convexity

(ii) Deduce that $\Lambda$ is a convex function on $J$.

(iii) Show that $\Lambda$ achieves its minimum at $t = 0$ if $PX = 0$.

[3] Let $Q$ be a probability measure defined on a finite interval $[a, b]$. Write $\sigma^2_Q$ for its variance.

(i) Show that $\sigma^2_Q \leq (b - a)^2/4$. Hint: Reduce to the case $b = -a$, noting that $\sigma^2_Q \leq Q^x(x^2)$.

(ii) Suppose also that $Q^x(x) = 0$. Define $\Lambda(t) := \log (Q^x e^x t)$, for $t \in \mathbb{R}$. Show that $\Lambda''(t) \leq (b - a)^2/4$, and hence $\Lambda(t) \leq t^2(b - a)^2/8$ for all $t \in \mathbb{R}$.

(iii) (Hoeffding 1963) Let $X_1, \ldots, X_n$ be independent random variables with zero expected values, and with $X_i$ taking values only in a finite interval $[a_i, b_i]$. For $\epsilon > 0$, show that

$$P\{X_1 + \cdots + X_n \geq \epsilon\} \leq \inf_{t > 0} e^{-\epsilon t} \prod_i P e^{t X_i} \leq \exp \left( -\frac{2\epsilon^2}{\sum_i (b_i - a_i)^2} \right) .$$

[4] Let $P$ be a probability measure on $\mathbb{R}^k$. Define $M(t) := P^x e^{xt}$ for $t \in \mathbb{R}^k$.

(i) Show that the set $C := \{ t \in \mathbb{R}^k : M(t) < \infty \}$ is convex.

(ii) Show that log $M(t)$ is convex on rel-int($C$).

[5] Let $f$ be a convex increasing function on $\mathbb{R}^+$. Show that there exists an increasing sequence of convex, increasing functions $f_n$, with each $f_n''$ bounded and continuous, such that $0 \leq f_n(x) \leq f_{n+1}(x) \uparrow f(x)$ for each $x$. Hint: Approximate the right-hand derivative of $f$ from below by smooth, increasing functions.

7. Notes

Most of the material described in this Appendix can be found, often in much greater generality, in the very thorough monograph by Rockafellar (1970).

References

