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,

$$\alpha x_1 + (1-\alpha)x_2 \in C \qquad \text{for all } x_1, x_2 \in C \text{ and all } 0 < \alpha < 1.$$

A real-valued function f defined on a convex subset C (of a vector space $\mathcal{V}$) is said to be convex if

$$f(\alpha x_1 + (1-\alpha)x_2) \le \alpha f(x_1) + (1-\alpha) f(x_2) \qquad \text{for all } x_1, x_2 \in C \text{ and } 0 < \alpha < 1.$$

Equivalently, the ***epigraph*** of the function,

$$\text{epi}(f) := \{(x,t) \in C \times \mathbb{R} : t \ge f(x)\},$$

is a convex subset of $C \times \mathbb{R}$. Some authors (such as Rockafellar 1970) define $f(x)$ to equal $+\infty$ for $x \in \mathcal{V}\backslash C$, so that the function is convex on the whole of $\mathcal{V}$, and $\text{epi}(f)$ is a convex subset of $\mathcal{V} \times \mathbb{R}$.

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 $f(x) = \sup_{i \in \mathbb{N}} (\alpha_i + \beta_i x)$ 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 $f'' \ge 0$, then f is convex.

(iii) 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.

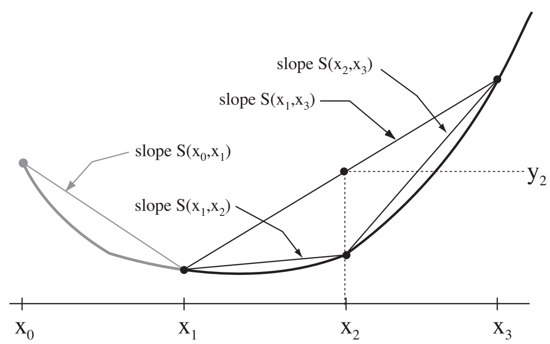
(iv) If C_1 and C_2 are disjoint convex subsets of $\mathbb{R}^n$ then there exists a nonzero ℓ in $\mathbb{R}^n$ for which $\sup_{x \in C_1} x \cdot \ell \leq \inf_{x \in C_2} x \cdot \ell$. That is, the linear functional $x \mapsto x \cdot \ell$ ***separates*** the two convex sets.

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 α 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 $S(x_i, x_j)$ for $\left(f(x_j) - f(x_i)\right)/(x_j - x_i)$, the slope of the chord joining the points $(x_i, f(x_i))$ and $(x_j, f(x_j))$. Then

$$\begin{aligned} S(x_2, x_3) &= \frac{f(x_3) - f(x_2)}{x_3 - x_2} \\ &\geq \frac{f(x_3) - y_2}{x_3 - x_2} = S(x_1, x_3) = \frac{y_2 - f(x_1)}{x_2 - x_1} \\ &\geq \frac{f(x_2) - f(x_1)}{x_2 - x_1} = S(x_1, x_2). \end{aligned}$$



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

$$D_+(x_2) \le S(x_2, x_3) \to S(x_1, x_3) \qquad \text{as } x_2 \downarrow x_1, \text{ for fixed } x_3$$
$$\to D_+(x_1) \qquad \text{as } x_3 \downarrow x_1.$$

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 $D_-(x_1) \le D_+(x_1)$. 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,

$$\frac{f(y) - f(x)}{y - x} \downarrow D_+(x) \qquad \text{as } y \downarrow x,$$

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,

$$\frac{f(x) - f(z)}{x - z} \uparrow D_-(x) \qquad \text{as } z \uparrow x,$$

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 $a \le x < y \le b$,

$$D_+(x) \le \frac{f(y) - f(x)}{y - x} \le D_-(y).$$

(iv) $D_-(x) \le D_+(x)$ for each x in (a, b), and

$$f(w) \ge f(x) + c(w - x) \qquad \text{for all } w \text{ in } [a, b],$$

for each real c with $D_-(x) \le c \le D_+(x)$.

Proof. Only the second part of assertion (iv) remains to be proved. For $w > x$ use

$$\frac{f(w) - f(x)}{w - x} = S(x, w) \ge D_+(x) \ge c;$$

for $w < x$ use

$$\frac{f(x) - f(w)}{x - w} = S(w, x) \le D_-(x) \le c,$$

□ 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 $f(x) \ge f(x_0)$ for all x in a neighborhood of x_0, then $f(w) \ge f(x_0)$ for all w in C.*

Proof. Consider first the case $n = 1$. Suppose $w \in C$ with $w > x_0$. The right-hand derivative $D_+(x_0) = \lim_{y \downarrow x_0} (f(y) - f(x_0))/(y - x_0)$ 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 $f(0+) := \lim_{w \downarrow 0} f(w)$.

<3> **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 $\mathcal{X}_0 := \{x_i : i \in \mathbb{N}\}$ 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

$$\psi(w) \geq f(x_i) + c_i(w - x_i) \geq f(w) - (x_i - w)(D_+(x_i) - D_+(w)) \qquad \text{for all } x_i.$$

Let x_i decrease to w (through $\mathcal{X}_0$) to conclude, via right-continuity of D_+ at w, that $\psi(w) \geq f(w)$.

If $D_+(a) > -\infty$ then f is continuous at a, and

$$f(a) \geq \psi(a) \geq \limsup_{x_i \downarrow a} (f(x_i) + (a - x_i)c_i) = f(a+) = f(a).$$

If $D_+(a) = -\infty$ then f must be decreasing in some neighborhood $\mathcal{N}$ of a, with $c_i < 0$ when $x_i \in \mathcal{N}$, and

$$\psi(a) \geq \sup_{x_i \in \mathcal{N}} (f(x_i) + (a - x_i)c_i) \geq \sup_{x_i \in \mathcal{N}} f(x_i) = f(a+).$$

If $\psi(a)$ were strictly greater than $f(a+)$, the open set

$$\{w : \psi(w) > f(a+)\} = \cup_i \{w : d_i + c_i w > f(a+)\}$$

would contain a neighborhood of a, which would imply existence of points w in $\mathcal{N} \backslash \{a\}$ for which $\psi(w) > f(a+) \geq f(w)$, contradicting the inequality $\psi(w) \leq f(w)$. A similar argument works at the other endpoint. □

3. Integral representations

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

<4> **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*

$$f(x) = f(a) + \int_a^x D_+(t)\,dt = f(a) + \int_a^x D_-(t)\,dt \qquad \text{for } a \le x \le b.$$

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

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

which sums to give

$$\int_{\alpha}^{x_{n-2}} D_+(t)\,dt \le f(x_{n-1}) - f(x_1) \le \int_{x_2}^{\beta} D_-(t)\,dt.$$

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

Let α 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 $f(\beta) - f(a) = \int_a^\beta D_+(t)\,dt = \int_a^\beta D_-(t)\,dt$. In particular, the negative parts of both $D_\pm$ are integrable. Then let β 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 $a \le x < b$. 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 $a \le x \le b$. (Possibly $g(b) = \infty$.) Then g is convex. For if $a \le x_0 < x_1 \le b$ and $0 < \alpha < 1$ and $x_\alpha := (1-\alpha)x_0 + \alpha x_1$, then

$$\begin{aligned}
&(1-\alpha)g(x_0) + \alpha g(x_1) - g(x_\alpha) \\
&\qquad = \int_a^b \big((1-\alpha)\{t \le x_0\} + \alpha\{t \le x_1\} - \{t \le x_\alpha\}\big)\, D(t)\,dt \\
&\qquad = \int_a^b \big(\alpha\{x_\alpha < t \le x_1\} - (1-\alpha)\{x_0 < t \le x_\alpha\}\big)\, D(t)\,dt \\
&\qquad \ge \big(\alpha(x_1 - x_\alpha) - (1-\alpha)(x_\alpha - x_0)\big)\, D(x_\alpha) = 0.
\end{aligned}$$

<5> **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

$$f(x) = x \int \{0 \le s \le 1\} f'(xs)\, ds$$
$$= x^2 \iint \{0 \le t \le s \le 1\} f''(xt)\, dt\, ds = x^2 \int_0^1 (1-t) f''(xt)\, dt,$$

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 ϕ.

Moreover, Jensen's inequality for the uniform distribution λ on the triangular region $\{0 \le t \le s \le 1\}$ implies that

$$\phi(x) = \lambda^{s,t} f''(xt) \ge f''\left(\lambda^{s,t} xt\right) = f''(x/3).$$

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

$$\Delta(x) := \begin{cases} \dfrac{e^x - 1 - x}{x^2/2} & \text{for } x \ne 0 \\ 1 & \text{for } x = 0 \end{cases}$$

is nonnegative and increasing over the whole real line. The choice $f(x) := (1+x)\log(1+x) - x$, for $x \ge -1$, with $f'(x) = \log(1+x)$ and $f''(x) = (1+x)^{-1}$, leads to the conclusion that the function

$$\psi(x) := \begin{cases} \dfrac{(1+x)\log(1+x) - x}{x^2/2} & \text{for } x \ge -1 \text{ and } x \ne 0 \\ 1 & \text{for } x = 0. \end{cases}$$

is nonnegative, convex, and decreasing. Also $x\psi(x)$ is increasing on $\mathbb{R}^+$, and $\psi(x) \ge (1 + x/3)^{-1}$. □

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 $\text{int}(C) \ne \emptyset$, there exists a convex, nonnegative function ρ defined on $\mathbb{R}^n$ for which $\text{int}(C) = \{x : \rho(x) < 1\} \subseteq C \subseteq \{x : \rho(x) \le 1\} = \overline{\text{int}(C)}$.

Proof. With no loss of generality, suppose $0 \in C$. Let $x_1, \dots, x_k$ 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.

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

If $\text{int}(C) \ne \emptyset$, 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 $\rho(z_1 + z_2) \le \rho(z_1) + \rho(z_2)$ for all z_i: if $z_i/t_i \in C$ 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, ρ is a convex function. Also ρ satisfies a Lipschitz condition: if $y = \sum_i y_i e_i$ and $z = \sum_i z_i e_i$ then

$$\begin{aligned}\rho(y) - \rho(z) \le \rho(y - z) &= \rho\left(\textstyle\sum_i (y_i - z_i)e_i\right) \\ &\le \textstyle\sum_i \left((y_i - z_i)^+\rho(e_i) + (y_i - z_i)^-\rho(-e_i)\right) \\ &\le |y - z| \left(\textstyle\sum_i \rho(e_i)^2 \vee \rho(-e_i)^2\right)^{1/2}.\end{aligned}$$

Thus $\{\rho < 1\}$ is open and $\{\rho \le 1\}$ is closed.

Clearly $\rho(x) \le 1$ 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 $\{z : \rho(z) < 1\} \subseteq C \subseteq \{z : \rho(z) \le 1\}$. 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 $C \subseteq x_0 \oplus \mathcal{V} \subseteq \mathbb{R}^n$, with $\dim(\mathcal{V}) = k < n$, we can identify $\mathcal{V}$ 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 $x_0 \oplus \mathcal{V}$. That is, there exist points x of C with open neighborhoods (in $\mathbb{R}^n$) for which $\mathcal{N} \cap (x_0 \oplus \mathcal{V}) \subseteq C$. The set of all such points is called the ***relative interior*** of C, and is denoted by 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 ρ defined only on $x_0 \oplus \mathcal{V}$.

5. Separation of convex sets by linear functionals

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

<7> **Lemma.** *Let f be a real-valued convex function, defined on a vector space $\mathcal{V}$. Let T_0 be a linear functional defined on a vector subspace $\mathcal{V}_0$, on which $T_0(x) \le f(x)$ for all $x \in \mathcal{V}_0$. Let y_1 be a point of $\mathcal{V}$ not in $\mathcal{V}_0$. There exists an extension of T_0 to a linear functional T_1 on the subspace $\mathcal{V}_1$ spanned by $\mathcal{V}_0 \cup \{y_1\}$ for which $T_1(z) \le f(z)$ on $\mathcal{V}_1$.*

Proof. Each point z in $\mathcal{V}_1$ has a unique representation $z := x + ry_1$, for some $x \in \mathcal{V}_0$ and some $r \in \mathbb{R}$. We need to find a value for $T_1(y_1)$ for which $f(x + ry_1) \geq T_0(x) + rT_1(y_1)$ for all $r \in \mathbb{R}$. Equivalently we need a real number c such that

$$\inf_{x_0 \in \mathcal{V}_0,\, t>0} \frac{f(x_0 + ty_1) - T_0(x_0)}{t} \geq c \geq \sup_{x_1 \in \mathcal{V}_0,\, s>0} \frac{T_0(x_1) - f(x_1 - sy_1)}{s},$$

for then $T_1(y_1) := c$ will give the desired extension.

For given x_0, x_1 in $\mathcal{V}_0$ and $s, t > 0$, define $\alpha := s/(s+t)$ and $x_\alpha := \alpha x_0 + (1-\alpha)x_1$. Then, by convexity of f on $\mathcal{V}_1$ and linearity of T_0 on $\mathcal{V}_0$,

$$\frac{s}{s+t} f(x_0 + ty_1) + \frac{t}{s+t} f(x_1 - sy_1) \geq f(x_\alpha) \geq T_0(x_\alpha) = \frac{s}{s+t} T_0(x_0) + \frac{t}{s+t} T_0(x_1),$$

which implies

$$\infty > \frac{f(x_0 + ty_1) - T_0(x_0)}{t} \geq \frac{T_0(x_1) - f(x_1 - sy_1)}{s} > -\infty.$$

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 $\mathcal{V}$ need not be finite dimensional. We can order extensions of T_0, bounded above by f, by defining $(T_\alpha, \mathcal{V}_\alpha) \succeq (T_\beta, \mathcal{V}_\mathcal{B})$ to mean that $\mathcal{V}_\beta$ is a subspace of $\mathcal{V}_\alpha$, and T_α is an extension of T_β. Zorn's lemma gives a maximal element of the set of extensions $(T_\gamma, \mathcal{V}_\gamma) \succeq (T_0, \mathcal{V}_0)$. Lemma <7> shows that $\mathcal{V}_\gamma$ must equal the whole of $\mathcal{V}$, otherwise there would be a further extension. That is, T_0 has an extension to a linear functional T defined on $\mathcal{V}$ with $T(x) \leq f(x)$ for every x in $\mathcal{V}$. This result is a minor variation on the ***Hahn-Banach theorem*** from functional analysis (compare with page 62 of Dunford & Schwartz 1958).

<8> **Theorem.** *Let C be a convex subset of $\mathbb{R}^n$ and y_0 be a point not in* rel-int(C).

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

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

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

If $y_0 \in \mathcal{V}$, the problem reduces to construction of a suitable linear functional T on $\mathcal{V}$: we then have only to define $T(z) := 0$ for $z \notin \mathcal{V}$ to complete the proof. Equivalently, we may suppose that $\mathcal{V} = \mathbb{R}^n$. Define T_0 on $\mathcal{V}_0 := \{rx_0 : r \in \mathbb{R}\}$ by $T(ry_0) := r\rho(y_0)$, for the ρ defined in Theorem <6>. Note that $T_0(y_0) = \rho(y_0) \geq 1$, because $y_0 \notin$ rel-int$(C) = \{\rho < 1\}$. Clearly $T_0(x) \leq \rho(x)$ for all $x \in \mathcal{V}_0$. 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,

$$T(y_0) \geq 1 \geq \rho(x) \geq T(x) \qquad \text{for all } x \in \overline{C} = \{\rho \leq 1\}.$$

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

<9> **Corollary.** *Let C_1 and C_2 be disjoint convex subsets of $\mathbb{R}^n$. Then there is a nonzero linear functional for which* $\inf_{x \in \overline{C}_1} T(x) \geq \sup_{x \in \overline{C}_2} T(x)$.

Proof. Define C as the convex set $\{x_1 - x_2 : x_i \in C_i\}$. The origin does not belong to C. Thus there is a nonzero linear functional for which $0 = T(0) \geq T(x_1 - x_2)$ for all $x_i \in C_i$. □

<10> **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 ℓ_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 $x \notin H_i$. □

<11> **Corollary.** *Let f be a convex (real-valued) function defined on a convex subset C of $\mathbb{R}^n$, such that* 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 epi(f), there exist $\ell_i \in \mathbb{R}^n$ and constants $\alpha_i, k_i \in \mathbb{R}$ such that

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

The ith inequality can hold for arbitrarily large t only if $\alpha_i \geq 0$. Define $\psi(x) := \sup_{\alpha_i > 0} (\ell_i \cdot x - k_i)/\alpha_i$. 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 function 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, $M(t) := \mathbb{P}\exp(tX)$, 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 $e^{tX}/M(t)$ with respect to $\mathbb{P}$, for $t \in J$, with corresponding variance var$_t(\cdot)$. Define $\Lambda(t) := \log M(t)$.

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

$$\Lambda'(t) = M'(t)/M(t) = \mathbb{P}\big(Xe^{tX}/M(t)\big) = \mathbb{P}_t X,$$
$$\Lambda''(t) = \big(M(t)M''(t) - M'(t)^2\big)/M(t)^2 = \text{var}_t(X).$$

(ii) Deduce that Λ is a convex function on J.

(iii) Show that Λ achieves its minimum at $t = 0$ if $\mathbb{P}X = 0$.

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

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

(ii) Suppose also that $Q^x(x) = 0$. Define $\Lambda(t) := \log\left(Q^x e^{xt}\right)$, 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

$$\mathbb{P}\{X_1 + \ldots + X_n \geq \epsilon\} \leq \inf_{t>0} e^{-\epsilon t} \textstyle\prod_i \mathbb{P} e^{tX_i} \leq \exp\left(-2\epsilon^2 \big/ \sum_i (b_i - a_i)^2\right).$$

[4] Let P be a probability measure on $\mathbb{R}^k$. Define $M(t); = P^x\left(e^{x \cdot t}\right)$ 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

Dunford, N. & Schwartz, J. T. (1958), *Linear Operators, Part I: General Theory*, Wiley.

Hoeffding, W. (1963), 'Probability inequalities for sums of bounded random variables', *Journal of the American Statistical Association* **58**, 13–30.

Rockafellar, R. T. (1970), *Convex Analysis*, Princeton Univ. Press, Princeton, New Jersey.