TTIC 31070: Convex Optimization Spring 2026

# Lecture 2: Separation and Duality

Lecture 1 established the slogan that convexity turns local information into global information. Lecture 2 starts from convex functions, but very quickly it switches to sets. The point of the epigraph is exactly to make that switch. Once a convex function is encoded as a subset of $E \times \mathbb{R}$, the right tools are projection, separation, and supporting hyperplanes. After that, we translate the geometry back into function language through subgradients.

Lecture 2 turns the slogan of Lecture 1 that convexity turns local information into global information into one geometric picture:

$$\operatorname{cl}\operatorname{conv}(S) = \bigcap \left\{ H(\xi, b) : S \subseteq H(\xi, b) \right\}.$$

The two separation theorems to remember are the point-to-closed-convex-set theorem (Theorem 2.2) and the cone-separation theorem (Theorem 2.7). Everything else in the lecture is either a reformulation of that picture or a way of bringing it back to convex functions.

## 2.1 Extended real and epigraph

We begin by enlarging the value space from $\mathbb{R}$ to $\mathbb{R} \cup \{+\infty\}$. The point is not to do arithmetic with $+\infty$ for its own sake. The point is that $+\infty$ gives us a uniform way to mark points where the function should be treated as unavailable or forbidden. The set of points where the function stays finite will be called its effective domain. That makes the epigraph construction completely uniform.

**Definition 2.1** (Extended real line used in these notes)**.** We write

$$\mathbb{R} \cup \{+\infty\}$$

for the one-sided extended real line. Its order is the usual order on $\mathbb{R}$, extended by

$$a \le +\infty \qquad \text{for every } a \in \mathbb{R} \cup \{+\infty\}.$$

In Lecture 2 we use the conventions

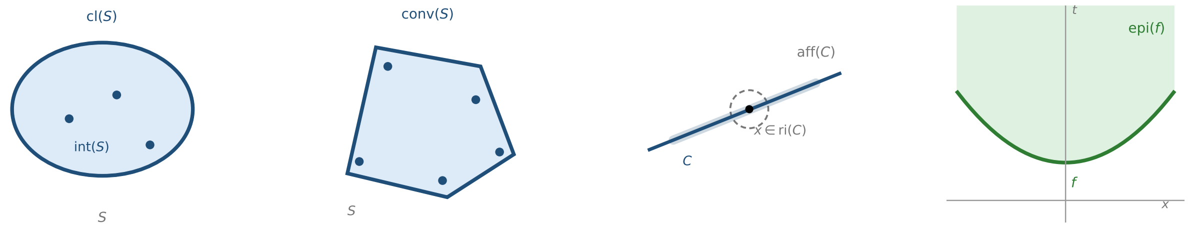
$$a + (+\infty) = +\infty, \qquad \lambda(+\infty) = +\infty \ (\lambda > 0), \qquad 0 \cdot (+\infty) := 0, \qquad \inf \varnothing = +\infty.$$

Similar properties hold for $-\infty$. The expression $+\infty - \infty$ is undefined, and we will not use it.

**Definition 2.2** (Extended-value function and effective domain)**.** Let $f : E \to \mathbb{R} \cup \{+\infty\}$. Its effective domain is

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

Thus the effective domain records exactly where the function is finite, while the value $+\infty$ encodes points outside that finite part of the model.



(a) Closure (b) Convex hull (c) Relative interior (d) Epigraph

Figure 1: A compact visual dictionary for the geometric operators that recur throughout Lectures 2 and 3: closure, interior, convex hull, relative interior, and epigraph.

**Definition 2.3** (Convex extended-value function)**.** Let $f : E \to \mathbb{R} \cup \{+\infty\}$. We say that $f$ is *convex* if

$$\forall x, y \in E,\ \forall \theta \in [0,1], \qquad f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y).$$

*Remark* 2.1 (Compatibility with the Lecture 1 notion of convexity). Let $C \subseteq E$ be convex, and let $h : C \to \mathbb{R}$. Define the ambient extension $\tilde{h} : E \to \mathbb{R} \cup \{+\infty\}$ by

$$\tilde{h}(x) := \begin{cases} h(x), & x \in C, \\ +\infty, & x \notin C. \end{cases}$$

Then $h$ is convex in the sense of Definition 1.4 if and only if $\tilde{h}$ is convex in the extended-value sense above. So Lecture 2 is not changing the notion of convexity. It is only changing the packaging.

**Definition 2.4** (Epigraph)**.** Let $f : E \to \mathbb{R} \cup \{+\infty\}$. Its epigraph is

$$\operatorname{epi}(f) := \{(x,t) \in E \times \mathbb{R} : t \ge f(x)\}.$$

The epigraph is the first place where the function-to-set translation becomes concrete (see Figure 1, panel (d)). The next lemma says that convexity of $f$ is exactly convexity of $\operatorname{epi}(f)$.

**Lemma 2.1** (Epigraph criterion)**.** *Let $f : E \to \mathbb{R} \cup \{+\infty\}$. Then the following are equivalent:*

1. *For all $x, y \in E$ and all $\theta \in [0,1]$,*

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y).$$

2. *$\operatorname{epi}(f)$ is a convex subset of $E \times \mathbb{R}$.*

After the epigraph criterion, the lecture switches to convex sets. The first geometric input is Euclidean projection. Projection is what manufactures separating hyperplanes, and later those separating hyperplanes will come back as subgradients.

## 2.2 Separation theorems for convex sets

**Definition 2.5** (Closed halfspace)**.** Let $\xi \in E^* \setminus \{0\}$ and $b \in \mathbb{R}$. The associated closed halfspace is

$$H(\xi, b) := \{x \in E : \langle \xi, x \rangle \leq b\}.$$

**Definition 2.6** (Interior and closure)**.** Let $S \subseteq E$. Its interior is

$$\operatorname{int}(S) := \{x \in S : \exists r > 0 \text{ such that } \{y \in E : \|y - x\| < r\} \subseteq S\}.$$

Its closure is

$$\operatorname{cl}(S) := \bigcap \{F \subseteq E : F \text{ is closed and } S \subseteq F\}.$$

Equivalently,

$$x \in \operatorname{cl}(S) \iff \forall r > 0,\ \{y \in E : \|y - x\| < r\} \cap S \neq \varnothing.$$

*Remark* 2.2 (Basic properties of interior and closure). For every $S \subseteq E$, the set $\operatorname{int}(S)$ is open, the set $\operatorname{cl}(S)$ is closed, and

$$\operatorname{int}(S) \subseteq S \subseteq \operatorname{cl}(S).$$

Moreover, $S$ is open if and only if $\operatorname{int}(S) = S$, and $S$ is closed if and only if $\operatorname{cl}(S) = S$.

**Theorem 2.2** (Point-to-set separation for a closed convex set)**.** *Let $C \subseteq E$ be nonempty, closed, and convex, and let $z \in E \setminus C$. Then there exist $\xi \in E^* \setminus \{0\}$ and $b \in \mathbb{R}$ such that*

$$\langle \xi, z \rangle > b \qquad \text{and} \qquad \forall y \in C,\ \langle \xi, y \rangle \leq b.$$

*Equivalently, $C \subseteq H(\xi, b)$ while $z \notin H(\xi, b)$.*

Why is the closedness assumption needed? The proof below reduces separation to Euclidean projection, and projection may fail to exist for a convex set that is not closed. The statement itself can also fail: for example, if $C = (0, 1) \subseteq \mathbb{R}$ and $z = 1$, then $z \notin C$ but no strict separating halfspace exists. So closedness is not just a proof artifact.

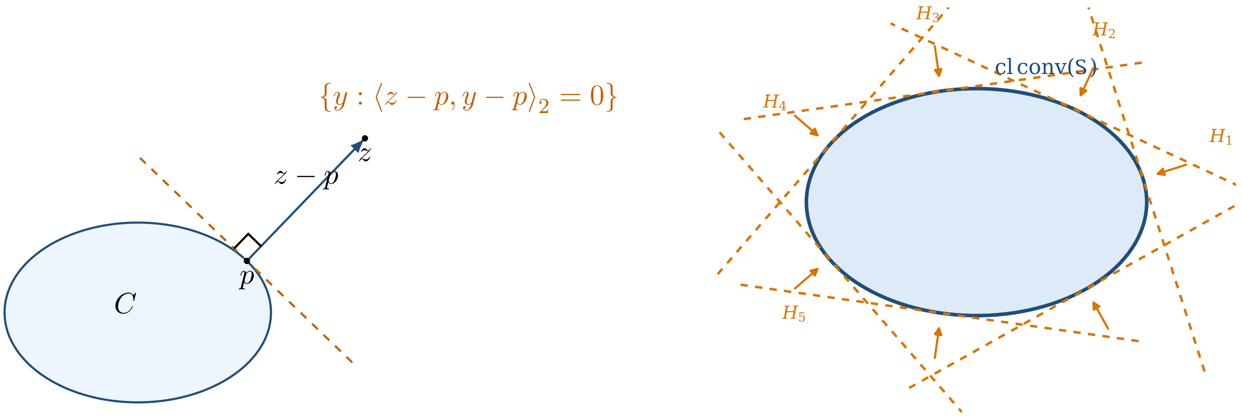
Proof idea. Project $z$ onto $C$, call the projection $p$, and use the normal vector $z - p$ to define the separating hyperplane. The next theorem provides exactly this projection step.

**Theorem 2.3** (Projection onto a nonempty closed convex set)**.** *Let $\langle \cdot, \cdot \rangle_2$ be an inner product on $E$, and let $\|\cdot\|_2$ be its induced Euclidean norm. Let $C \subseteq E$ be nonempty, closed, and convex, and let $z \in E$. Then there exists a unique point $p \in C$ such that*

$$\|z - p\|_2 = \inf_{y \in C} \|z - y\|_2.$$

*Moreover, for $p \in C$, the following are equivalent:*

1. $p \in \arg\min_{y \in C} \|z - y\|_2$.
2. $\forall y \in C,\ \langle z - p, y - p \rangle_2 \leq 0$.



**(a)** Euclidean projection and point-to-set separation.

**(b)** $\operatorname{cl}\operatorname{conv}(S)$ as the intersection of all containing halfspaces.

Figure 2: Panel **(a)** shows the geometry behind Theorems 2.2 and 2.3: the point $p$ is the Euclidean projection of $z$ onto the closed convex set $C$, and the vector $z-p$ determines a separating supporting hyperplane. Panel **(b)** shows the memory image behind Theorem 2.6: intersecting all halfspaces that contain $S$ recovers exactly $\operatorname{cl}\operatorname{conv}(S)$.

Projection has now produced the first genuine separation statement. From the certificate point of view, Theorem 2.2 does more than assert that $z \notin C$. It produces a covector $\xi$ and a threshold $b$ such that the single linear inequality $\langle \xi, y\rangle \leq b$ is valid for every $y \in C$, while the offending point $z$ violates it. In that sense, a separating hyperplane is a proof certificate for non-membership in a convex set.

The next step is to let exterior points approach a boundary point of $C$; the limiting argument gives the supporting hyperplane theorem.

**Theorem 2.4** (Supporting hyperplane theorem)**.** *Let $C \subseteq E$ be nonempty and convex, and let $x \in \operatorname{cl}(C) \setminus \operatorname{int}(C)$. Then there exist $\xi \in E^* \setminus \{0\}$ and $b \in \mathbb{R}$ such that*

$$\langle \xi, x\rangle = b \qquad \textit{and} \qquad \forall y \in C,\ \langle \xi, y\rangle \leq b.$$

## 2.3 Duality of convex sets and cones

Once supporting hyperplanes are available, the natural global question is whether a closed convex set can be reconstructed from all the halfspaces that contain it. To state this cleanly, we first introduce convex hull and closed convex hull (see Figure 1, panels (a) and (b)).

**Definition 2.7** (Convex hull)**.** Let $S \subseteq E$. The convex hull of $S$ is

$$\operatorname{conv}(S) := \bigcap \{C \subseteq E : C \text{ is convex and } S \subseteq C\}.$$

**Definition 2.8** (Closed convex hull)**.** Let $S \subseteq E$. The closed convex hull of $S$ is

$$\operatorname{cl}\operatorname{conv}(S) := \bigcap \{C \subseteq E : C \text{ is closed and convex and } S \subseteq C\}.$$

**Proposition 2.5** (Basic properties of closure, convex hull, and closed convex hull)**.** *Let $S \subseteq E$. Then the following hold:*

1. $\mathrm{cl}(S)$ *is closed,* $\mathrm{conv}(S)$ *is convex, and* $S \subseteq \mathrm{cl}(S) \cap \mathrm{conv}(S)$.
2. $\mathrm{cl}\,\mathrm{conv}(S) = \mathrm{cl}(\mathrm{conv}(S))$.
3. $\mathrm{cl}\,\mathrm{conv}(S)$ *is closed and convex, and* $S \subseteq \mathrm{cl}\,\mathrm{conv}(S)$.
4. *If $C \subseteq E$ is closed and convex, then* $\mathrm{cl}\,\mathrm{conv}(C) = C$.
5. $\mathrm{cl}\,\mathrm{conv}(\mathrm{cl}\,\mathrm{conv}(S)) = \mathrm{cl}\,\mathrm{conv}(S)$.
6. *There exists a closed set $S \subseteq \mathbb{R}^2$ such that*

$$\mathrm{conv}(\mathrm{cl}(S)) \subsetneq \mathrm{cl}\,\mathrm{conv}(S).$$

**Theorem 2.6** (Closed convex hull equals the intersection of all containing halfspaces)**.** *Let $S \subseteq E$. Then*

$$\mathrm{cl}\,\mathrm{conv}(S) = \bigcap \left\{ H(\xi, b) : \xi \in E^* \setminus \{0\},\ b \in \mathbb{R},\ S \subseteq H(\xi, b) \right\}.$$

*In particular, if $S$ is nonempty, closed, and convex, then*

$$S = \mathrm{cl}\,\mathrm{conv}(S) = \bigcap \left\{ H(\xi, b) : \xi \in E^* \setminus \{0\},\ b \in \mathbb{R},\ S \subseteq H(\xi, b) \right\}.$$

At this point the set-level duality picture is in place: $\mathrm{cl}\,\mathrm{conv}(S)$ is recovered by intersecting all containing halfspaces. The next step is to specialize that picture to convex cones. That cone version will be used again later, especially in Lecture 4 and Lecture 6.

**Certificate viewpoint.** One way to read Theorem 2.6 is as a completeness statement for linear proof certificates. To prove that a point does *not* belong to a closed convex set, we search for a valid linear inequality that the point violates. The theorem says that for closed convex sets this proof system is complete: every false membership claim can be refuted by some containing halfspace.

**Example 2.1** (The PSD cone as a proof system)**.** Work in the space $\mathbb{S}^n$ of symmetric $n \times n$ matrices, equipped with the Frobenius pairing

$$\langle Y, Z \rangle := \mathrm{tr}(YZ).$$

The cone of positive semidefinite matrices satisfies

$$\mathbb{S}^n_+ = \left\{ X \in \mathbb{S}^n : \forall u \in \mathbb{R}^n,\ u^\top X u \geq 0 \right\} = \bigcap_{u \in \mathbb{R}^n} \left\{ X \in \mathbb{S}^n : \left\langle uu^\top, X \right\rangle \geq 0 \right\}.$$

So each rank-one matrix $uu^\top$ defines a linear inequality

$$\left\langle uu^\top, X \right\rangle \geq 0$$

that is valid on $\mathbb{S}^n_+$. If $X \notin \mathbb{S}^n_+$, then there exists $u \in \mathbb{R}^n$ such that

$$\left\langle uu^\top, X \right\rangle = u^\top X u < 0.$$

Thus a rank-one matrix $uu^\top$ is a concrete proof that the symmetric matrix $X$ is not PSD. In this language, Theorem 2.6 says that this kind of linear proof system is complete: every false claim "$X \succeq 0$" can be refuted by a separating inequality.

**Definition 2.9** (Convex cone, dual cone, and polar cone)**.** A set $K \subseteq E$ is a convex cone if

$$\forall x_1, x_2 \in K,\ \forall \alpha_1, \alpha_2 \in [0, +\infty), \qquad \alpha_1 x_1 + \alpha_2 x_2 \in K.$$

Its dual cone and polar cone are

$$K^* := \{\xi \in E^* : \forall x \in K,\ \langle \xi, x\rangle \geq 0\}, \qquad K^\circ := \{\xi \in E^* : \forall x \in K,\ \langle \xi, x\rangle \leq 0\}.$$

**Theorem 2.7** (Cone separation and bipolarity)**.** *Let $K \subseteq E$ be a nonempty closed convex cone, and let $z \in E \setminus K$. Then there exists $\xi \in K^\circ$ such that*

$$\langle \xi, z\rangle > 0.$$

*In finite dimensions, the canonical map $\iota_E : E \to E^{**}$ defined by*

$$(\iota_E(x))(\xi) := \langle \xi, x\rangle \qquad \text{for every } x \in E,\ \xi \in E^*$$

*is an isomorphism. Under this identification, define the bipolar cone by*

$$K^{\circ\circ} := \{x \in E : \forall \xi \in K^\circ,\ \langle \xi, x\rangle \leq 0\}.$$

*Consequently,*

$$K^{\circ\circ} = K.$$

*With the convention $K^* = -K^\circ$, it also reads $K^{**} = K$.*

The cone theorem is the last set-level specialization in this lecture. Before returning to functions, we record affine hull and relative interior, because the final subgradient existence theorem naturally lives on $\mathrm{ri}(\mathrm{dom}\, f)$.

**Definition 2.10** (Affine sets and affine spaces)**.** Let $A \subseteq E$. We say that $A$ is *affine* if

$$\forall x, y \in A,\ \forall \theta \in \mathbb{R}, \qquad \theta x + (1-\theta) y \in A.$$

In these notes, an *affine space* means an affine subset of some ambient finite-dimensional real vector space, equipped with the induced affine operations.

**Definition 2.11** (Affine hull and relative interior)**.** Let $S \subseteq E$. Its affine hull is

$$\mathrm{aff}(S) := \bigcap \{A \subseteq E : A \text{ is affine and } S \subseteq A\}.$$

If $C \subseteq E$, then the relative interior of $C$ is

$$\mathrm{ri}(C) := \{x \in C : \exists r > 0 \text{ such that } \{y \in \mathrm{aff}(C) : \|y - x\| < r\} \subseteq C\}.$$

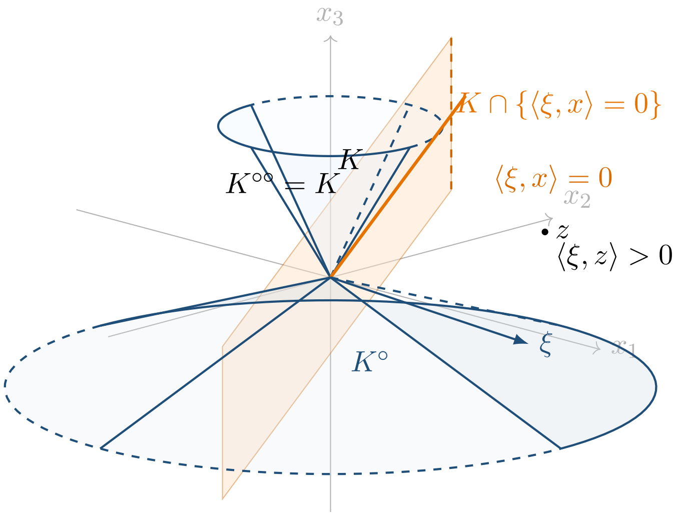


Figure 3: A three-dimensional schematic for cone separation and bipolarity. Here the upper cone is $K = \left\{(x_1, x_2, x_3) : \sqrt{x_1^2 + x_2^2} \le x_3/2,\ x_3 \ge 0\right\}$, while the lower cone is its polar $K^\circ = \left\{(\xi_1, \xi_2, \xi_3) : \sqrt{\xi_1^2 + \xi_2^2} \le -2\xi_3,\ \xi_3 \le 0\right\}$. Thus the upper cone is genuinely narrower, and its polar cone below is genuinely wider.

*Remark* 2.3 (Why affine spaces appear here). Convexity and the separation theorems above are fundamentally affine notions: they only use affine combinations and halfspaces, not a preferred origin in the ambient vector space. One could therefore formulate much of this lecture directly on affine spaces. To keep the terminology lighter, we do not globalize that viewpoint throughout the course. Instead, we work globally in a vector space $E$, and pass locally to affine subspaces when needed. The proof of Theorem 2.10 is the first place where this local affine viewpoint becomes essential. See Figure 1, panel (c) for a picture of relative interior.

## 2.4 Subgradient and its existence

We now translate the geometry back into function language. A supporting hyperplane to $\operatorname{epi}(f)$ through $(x, f(x))$ is exactly a global affine lower bound on $f$, and that is what a subgradient is. The only new point is that the relevant ambient space is not always all of $E$, but $\operatorname{aff}(\operatorname{dom} f) \times \mathbb{R}$.

**Definition 2.12** (Subgradient and subdifferential)**.** Let $f : E \to \mathbb{R} \cup \{+\infty\}$, and let $x \in \operatorname{dom} f$. A covector $g \in E^*$ is a subgradient of $f$ at $x$ if

$$\forall y \in E, \qquad f(y) \ge f(x) + \langle g, y - x\rangle\,.$$

The set of all such covectors is denoted by $\partial f(x)$.

**Definition 2.13** (Proper extended-value function)**.** Let $f : E \to \mathbb{R} \cup \{+\infty\}$. We say that $f$ is *proper* if

$$\operatorname{dom} f \ne \varnothing.$$

Equivalently, $f$ is proper if there exists $x \in E$ such that $f(x) < +\infty$.[a]

---

[a] In treatments based on the full extended real line $\mathbb{R} \cup \{\pm\infty\}$, one also requires $f(x) > -\infty$ for every $x$. We do not impose that extra condition here because our present extended-value convention has already excluded the value $-\infty$.

**Lemma 2.8** (Subgradients are supporting hyperplanes of the epigraph)**.** *Let* $f : E \to \mathbb{R} \cup \{+\infty\}$*, let* $x \in \operatorname{dom} f$*, and let* $g \in E^*$*. Then the following are equivalent:*

1. $g \in \partial f(x)$.
2. $\forall (y,t) \in \operatorname{epi}(f),\ \langle (g,-1),(y,t) - (x, f(x)) \rangle \leq 0$.

**Proposition 2.9** (Differentiable convex functions have a unique subgradient)**.** *Let* $f : E \to \mathbb{R}$ *be convex and differentiable. Then for every* $x \in E$*,*

$$\partial f(x) = \{\nabla f(x)\}.$$

**Theorem 2.10** (Existence of a subgradient on the relative interior)**.** *Let* $f : E \to \mathbb{R} \cup \{+\infty\}$ *be proper and convex. Then*

$$\forall x \in \operatorname{ri}(\operatorname{dom} f), \qquad \partial f(x) \neq \varnothing.$$

*Remark* 2.4 (Why the (relative) interior condition is needed)*.* The hypothesis $x \in \operatorname{ri}(\operatorname{dom} f)$ cannot be weakened to $x \in \operatorname{dom} f$. Consider

$$f(x) := \begin{cases} x \log x, & x > 0, \\ 0, & x = 0, \\ +\infty, & x < 0. \end{cases}$$

Then $f$ is proper, closed, and convex, with $\operatorname{dom} f = [0, \infty)$, so $0 \in \operatorname{cl}(\operatorname{dom} f) \setminus \operatorname{ri}(\operatorname{dom} f)$. However, $\partial f(0) = \varnothing$. Indeed, if $g \in \partial f(0)$, then for every $y > 0$,

$$y \log y = f(y) \geq f(0) + gy = gy,$$

so $\log y \geq g$ for every $y > 0$, which is impossible because $\log y \to -\infty$ as $y \downarrow 0$.

## Proofs

> *Proof of Lemma 2.1.* Assume first that $f$ is convex, and let $(x_1, t_1), (x_2, t_2) \in \operatorname{epi}(f)$ and $\theta \in [0,1]$. Then $f(x_i) \leq t_i$ for $i \in \{1,2\}$, hence
>
> $$\begin{aligned} f(\theta x_1 + (1-\theta)x_2) &\leq \theta f(x_1) + (1-\theta) f(x_2) \\ &\leq \theta t_1 + (1-\theta) t_2. \end{aligned}$$
>
> Therefore $(\theta x_1 + (1-\theta)x_2, \theta t_1 + (1-\theta)t_2) \in \operatorname{epi}(f)$, so $\operatorname{epi}(f)$ is convex.
> Conversely, assume that $\operatorname{epi}(f)$ is convex, and let $x_1, x_2 \in E$ and $\theta \in [0,1]$. If $\theta = 0$ or $\theta = 1$, the desired inequality is immediate. Assume therefore that $\theta \in (0,1)$. If $f(x_1) = +\infty$ or $f(x_2) = +\infty$,

then

$$\theta f(x_1) + (1-\theta) f(x_2) = +\infty,$$

so the desired inequality is trivial. Thus both $f(x_1)$ and $f(x_2)$ are finite real numbers, and therefore $(x_1, f(x_1))$ and $(x_2, f(x_2))$ belong to $\operatorname{epi}(f)$. By convexity of $\operatorname{epi}(f)$,

$$(\theta x_1 + (1-\theta)x_2,\ \theta f(x_1) + (1-\theta) f(x_2)) \in \operatorname{epi}(f),$$

which means exactly that

$$f(\theta x_1 + (1-\theta)x_2) \le \theta f(x_1) + (1-\theta) f(x_2).$$

Hence $f$ is convex. $\square$

*Proof of Theorem 2.3.* Choose any $y_0 \in C$ and define $R := \|z - y_0\|_2 + 1$. If $y \in C$ and $\|y - z\|_2 > R$, then $\|y - z\|_2 > \|y_0 - z\|_2$, so a minimizer over $C$ must lie in the nonempty set

$$K := C \cap \overline{B}_2(z, R).$$

Because the auxiliary norm $\|\cdot\|_2$ comes from an inner product on the finite-dimensional space $E$, the set $K$ is closed and bounded, hence compact. The function

$$\phi(y) := \|z - y\|_2^2$$

is continuous on $K$. Therefore, by the Weierstrass theorem (*Theorem* 1.1), there exists $p \in K \subseteq C$ such that

$$\phi(p) = \min_{y \in K} \phi(y).$$

By the choice of $R$, this point is also a minimizer over all of $C$.

To prove uniqueness, suppose that both $p, q \in C$ minimize the distance to $z$. Since $C$ is convex, $(p+q)/2 \in C$. Strict convexity of the squared auxiliary Euclidean norm gives

$$\left\| z - \tfrac{1}{2}(p+q) \right\|_2^2 < \frac{1}{2} \|z - p\|_2^2 + \frac{1}{2} \|z - q\|_2^2$$

whenever $p \ne q$, contradicting minimality. Hence the minimizer is unique.

Now assume that $p$ is the projection of $z$ onto $C$. Fix $y \in C$ and define $\phi_y(\tau) := \|z - (p + \tau(y-p))\|_2^2$ for $\tau \in [0,1]$. Since $p + \tau(y-p) \in C$ for all $\tau \in [0,1]$ and $\phi_y$ is minimized at $\tau = 0$, we have $\phi_y'(0^+) \ge 0$. A direct differentiation yields

$$\phi_y'(0^+) = -2 \langle z - p, y - p \rangle_2 ,$$

so $\langle z - p, y - p \rangle_2 \le 0$.

Conversely, suppose that $p \in C$ and that $\langle z - p, y - p \rangle_2 \le 0$ for all $y \in C$. Then for every $y \in C$,

$$\|z - y\|_2^2 = \|z - p\|_2^2 + \|y - p\|_2^2 - 2 \langle z - p, y - p \rangle_2 \ge \|z - p\|_2^2 .$$

Hence $p$ is the unique projection of $z$ onto $C$. $\square$

*Proof of Theorem 2.2.* Let $p$ be the auxiliary Euclidean projection of $z$ onto $C$, whose existence and uniqueness follow from Theorem 2.3. Because $z \notin C$, we have $z \neq p$. Define $\xi \in E^*$ by

$$\langle \xi, y \rangle := \langle z - p, y \rangle_2 \qquad \text{for every } y \in E,$$

and set $b := \langle \xi, p \rangle$. By Theorem 2.3,

$$\forall y \in C, \qquad \langle z - p, y - p \rangle_2 \leq 0.$$

Equivalently,

$$\forall y \in C, \qquad \langle \xi, y \rangle \leq \langle \xi, p \rangle = b.$$

Moreover,

$$\langle \xi, z \rangle - b = \langle z - p, z - p \rangle_2 = \|z - p\|_2^2 > 0.$$

Therefore $\langle \xi, z \rangle > b \geq \langle \xi, y \rangle$ for all $y \in C$. Since $z - p \neq 0$, also $\xi \neq 0$. ☐

**Lemma 2.11** (Closure preserves convexity and interior for convex sets)**.** *Let $C \subseteq E$ be convex. Then* $\mathrm{cl}(C)$ *is convex and*

$$\mathrm{int}(\mathrm{cl}(C)) = \mathrm{int}(C).$$

*Proof of Lemma 2.11.* To prove convexity of $\mathrm{cl}(C)$, let $u, v \in \mathrm{cl}(C)$ and $\theta \in [0, 1]$. Choose sequences $u_n, v_n \in C$ such that $u_n \to u$ and $v_n \to v$. Then

$$\theta u_n + (1 - \theta) v_n \in C$$

for every $n$, and

$$\theta u_n + (1 - \theta) v_n \to \theta u + (1 - \theta) v.$$

Hence $\theta u + (1 - \theta) v \in \mathrm{cl}(C)$.

The inclusion $\mathrm{int}(C) \subseteq \mathrm{int}(\mathrm{cl}(C))$ is immediate from $C \subseteq \mathrm{cl}(C)$. For the reverse inclusion, let $x_0 \in \mathrm{int}(\mathrm{cl}(C))$. Choose $r > 0$ such that

$$\{y \in E : \|y - x_0\| < r\} \subseteq \mathrm{cl}(C).$$

Let $d := \dim(E)$, fix a basis $e_1, \dots, e_d$ of $E$, and choose $\varepsilon > 0$ so small that the points

$$y_i := x_0 + \varepsilon e_i \qquad (1 \leq i \leq d), \qquad y_0 := x_0 - \varepsilon \sum_{i=1}^{d} e_i$$

all belong to $\{y \in E : \|y - x_0\| < r\}$. Then $y_0, \dots, y_d \in \mathrm{cl}(C)$, they are affinely independent, and

$$x_0 = \frac{1}{d+1} \sum_{i=0}^{d} y_i,$$

so $x_0 \in \mathrm{int}(\mathrm{conv}\{y_0, \dots, y_d\})$. Choose $z_0, \dots, z_d \in C$ sufficiently close to $y_0, \dots, y_d$ that $z_0, \dots, z_d$ remain affinely independent and the barycentric coordinates of $x_0$ with respect to $z_0, \dots, z_d$ remain strictly positive. Then

$$x_0 \in \mathrm{int}(\mathrm{conv}\{z_0, \dots, z_d\}) \subseteq C,$$

so $x_0 \in \mathrm{int}(C)$. Therefore $\mathrm{int}(\mathrm{cl}(C)) \subseteq \mathrm{int}(C)$, proving the claim. ☐

*Proof of Theorem 2.4.* By Lemma 2.11, the set $\mathrm{cl}(C)$ is convex and

$$\mathrm{int}(\mathrm{cl}(C)) = \mathrm{int}(C).$$

Since $x \notin \mathrm{int}(C) = \mathrm{int}(\mathrm{cl}(C))$, every neighborhood of $x$ meets $E \setminus \mathrm{cl}(C)$. Choose a sequence $(z_k)_{k\geq 1} \subseteq E \setminus \mathrm{cl}(C)$ such that $z_k \to x$. For each $k$, apply Theorem 2.2 to $\mathrm{cl}(C)$ and $z_k$. Then there exist $\xi_k \in E^* \setminus \{0\}$ and $b_k \in \mathbb{R}$ such that

$$\langle \xi_k, z_k \rangle > b_k \qquad \text{and} \qquad \forall y \in \mathrm{cl}(C),\ \langle \xi_k, y \rangle \leq b_k.$$

By the auxiliary Euclidean inner product, there exists a unique $a_k \in E$ such that

$$\langle \xi_k, y \rangle = \langle a_k, y \rangle_2 \qquad \text{for every } y \in E.$$

Rescale so that $\|a_k\|_2 = 1$ for every $k$. By compactness of the auxiliary Euclidean unit sphere, after passing to a subsequence we may assume that $a_k \to a$ for some $a \in E$ with $\|a\|_2 = 1$. Since $x \in \mathrm{cl}(C)$, we have $\langle \xi_k, x \rangle \leq b_k < \langle \xi_k, z_k \rangle$ for every $k$. Because $z_k \to x$ and $\|a_k\|_2 = 1$, the difference

$$0 \leq b_k - \langle \xi_k, x \rangle < \langle \xi_k, z_k - x \rangle = \langle a_k, z_k - x \rangle_2 \leq \|z_k - x\|_2$$

tends to 0. Hence $b_k \to \langle a, x \rangle_2$. Now fix $y \in C$. Passing to the limit in $\langle \xi_k, y \rangle \leq b_k$, equivalently $\langle a_k, y \rangle_2 \leq b_k$, gives

$$\langle a, y \rangle_2 \leq \langle a, x \rangle_2 .$$

Define $\xi \in E^*$ by $\langle \xi, y \rangle = \langle a, y \rangle_2$ for all $y \in E$, and set $b := \langle \xi, x \rangle$. Then

$$\langle \xi, x \rangle = b \qquad \text{and} \qquad \forall y \in C,\ \langle \xi, y \rangle \leq b.$$

Since $\|a\|_2 = 1$, we have $a \neq 0$, and therefore $\xi \neq 0$. □

*Proof of Theorem 2.6.* Define

$$\mathcal{H}(S) := \bigcap \{ H(\xi, b) : \xi \in E^* \setminus \{0\},\ b \in \mathbb{R},\ S \subseteq H(\xi, b) \} .$$

Every closed halfspace is closed and convex, so $\mathrm{cl}\,\mathrm{conv}(S) \subseteq \mathcal{H}(S)$.
Conversely, let $z \in E \setminus \mathrm{cl}\,\mathrm{conv}(S)$. Applying Theorem 2.2 to the closed convex set $\mathrm{cl}\,\mathrm{conv}(S)$ and the point $z$, we obtain $\xi \neq 0$ and $b \in \mathbb{R}$ such that

$$\langle \xi, z \rangle > b \qquad \text{and} \qquad \forall y \in \mathrm{cl}\,\mathrm{conv}(S),\ \langle \xi, y \rangle \leq b.$$

In particular, $S \subseteq H(\xi, b)$, whereas $z \notin H(\xi, b)$. Hence $z \notin \mathcal{H}(S)$. This proves $\mathcal{H}(S) \subseteq \mathrm{cl}\,\mathrm{conv}(S)$, and therefore $\mathcal{H}(S) = \mathrm{cl}\,\mathrm{conv}(S)$. If $S$ is additionally nonempty, closed, and convex, then $\mathrm{cl}\,\mathrm{conv}(S) = S$ by Proposition 2.5, so the final displayed identity follows immediately. □

*Proof of Theorem 2.7.* Apply Theorem 2.2 to the closed convex set $K$ and the point $z \notin K$. Then there exist $\xi \in E^* \setminus \{0\}$ and $b \in \mathbb{R}$ such that

$$\langle \xi, z \rangle > b \qquad \text{and} \qquad \forall x \in K,\ \langle \xi, x \rangle \leq b.$$

Because $0 \in K$, we have $0 \leq b$. If there existed $x_0 \in K$ with $\langle \xi, x_0 \rangle > 0$, then $\lambda x_0 \in K$ for every $\lambda \geq 0$, so

$$\lambda \langle \xi, x_0 \rangle = \langle \xi, \lambda x_0 \rangle \leq b \qquad \forall \lambda \geq 0,$$

which is impossible as $\lambda \to +\infty$. Hence

$$\forall x \in K, \qquad \langle \xi, x \rangle \leq 0,$$

that is, $\xi \in K^\circ$. Since $b \geq 0$, we conclude that

$$\langle \xi, z \rangle > b \geq 0.$$

This proves the separation statement.
Under the finite-dimensional identification $E \simeq E^{**}$, define

$$K^{\circ\circ} := \{x \in E : \forall \xi \in K^\circ,\ \langle \xi, x \rangle \leq 0\}\,.$$

The inclusion $K \subseteq K^{\circ\circ}$ is immediate from the definition of the polar. Conversely, let $z \in E \setminus K$. By the first part, there exists $\xi \in K^\circ$ such that $\langle \xi, z \rangle > 0$. Therefore $z \notin K^{\circ\circ}$, since membership in $K^{\circ\circ}$ would require $\langle \xi, z \rangle \leq 0$ for every $\xi \in K^\circ$. Hence $K^{\circ\circ} \subseteq K$, and so $K^{\circ\circ} = K$. With the convention $K^* = -K^\circ$, it also reads $K^{**} = K$. ☐

*Proof of Lemma 2.8.* We have

$$\langle (g, -1), (y, t) - (x, f(x)) \rangle = \langle g, y - x \rangle - (t - f(x)).$$

Therefore

$$\forall (y, t) \in \operatorname{epi}(f), \quad \langle (g, -1), (y, t) - (x, f(x)) \rangle \leq 0$$

if and only if

$$\forall (y, t) \in \operatorname{epi}(f), \quad t \geq f(x) + \langle g, y - x \rangle\,.$$

Choosing $t = f(y)$, which is admissible for every $y \in \operatorname{dom} f$, this becomes

$$\forall y \in \operatorname{dom} f, \qquad f(y) \geq f(x) + \langle g, y - x \rangle\,.$$

For $y \notin \operatorname{dom} f$, the same inequality is automatic because $f(y) = +\infty$. Hence this is exactly the definition of $g \in \partial f(x)$. ☐

*Proof of Proposition 2.9.* Fix $x \in E$. By Lemma 1.4, the covector $\nabla f(x)$ gives a global affine lower bound on $f$, so $\nabla f(x) \in \partial f(x)$.
Conversely, let $g \in \partial f(x)$, and define

$$h(y) := f(y) - \langle g, y \rangle\,.$$

Since $f$ is convex and subtracting a linear functional preserves convexity, the function $h$ is convex. It is also differentiable. Moreover, the subgradient inequality for $g$ is exactly

$$\forall y \in E, \qquad h(x) \leq h(y).$$

Applying Theorem 1.5 with $\Omega = E$ gives

$$\forall y \in E, \qquad \langle \nabla h(x), y - x \rangle \geq 0.$$

Now fix $v \in E$. Taking $y = x + v$ gives $\langle \nabla h(x), v \rangle \geq 0$, while taking $y = x - v$ gives $\langle \nabla h(x), -v \rangle \geq 0$, hence $\langle \nabla h(x), v \rangle \leq 0$. Therefore $\langle \nabla h(x), v \rangle = 0$ for every $v \in E$, so $\nabla h(x) = 0$. Since $\nabla h(x) = \nabla f(x) - g$, we conclude that $g = \nabla f(x)$. Thus $\partial f(x) = \{\nabla f(x)\}$. ☐

*Proof of Theorem 2.10.* Let $x \in \operatorname{ri}(\operatorname{dom} f)$, set $A := \operatorname{aff}(\operatorname{dom} f)$, and let $L := A - x$. Consider the convex set

$$\mathcal{E} := \{(y,t) \in A \times \mathbb{R} : f(y) \le t\}\,.$$

Since $\mathcal{E} = \operatorname{epi}(f) \cap (A \times \mathbb{R})$, Lemma 2.1 implies that $\mathcal{E}$ is convex. Translate $\mathcal{E}$ so that $(x, f(x))$ becomes the origin:

$$\widetilde{\mathcal{E}} := \{(y-x, t-f(x)) \in L \times \mathbb{R} : (y,t) \in \mathcal{E}\}\,.$$

Then $\widetilde{\mathcal{E}}$ is a convex subset of the finite-dimensional vector space $L \times \mathbb{R}$, and $(0,0) \in \widetilde{\mathcal{E}}$. Moreover, $(0,0) \notin \operatorname{int}(\widetilde{\mathcal{E}})$: indeed, for every $\varepsilon > 0$, we have $(0,-\varepsilon) \notin \widetilde{\mathcal{E}}$.
Apply Theorem 2.4 in the finite-dimensional vector space $L \times \mathbb{R}$. We obtain $(\eta, \alpha) \in L^* \times \mathbb{R}$, not both zero, such that

$$\forall (v,s) \in \widetilde{\mathcal{E}}, \qquad \langle \eta, v\rangle + \alpha s \le 0.$$

Since $(v,s) \in \widetilde{\mathcal{E}}$ implies $(v, s+r) \in \widetilde{\mathcal{E}}$ for every $r \ge 0$, we must have $\alpha \le 0$. We claim that $\alpha < 0$. If $\alpha = 0$, then

$$\forall y \in \operatorname{dom} f, \qquad \langle \eta, y - x\rangle \le 0.$$

Because $x \in \operatorname{ri}(\operatorname{dom} f)$, there exists $r > 0$ such that $x + v \in \operatorname{dom} f$ for every $v \in L$ with $\|v\| < r$. Fix $v \in L$, and choose $t > 0$ so small that $t\,\|v\| < r$. Then $x + tv, x - tv \in \operatorname{dom} f$, so the displayed inequality gives

$$\langle \eta, tv\rangle \le 0 \qquad \text{and} \qquad \langle \eta, -tv\rangle \le 0.$$

Hence $\langle \eta, v\rangle = 0$. Since $v \in L$ was arbitrary, $\eta = 0$, a contradiction. Thus $\alpha < 0$.
Set $g_L := -\eta/\alpha \in L^*$. For every $y \in \operatorname{dom} f$, the point $(y, f(y))$ belongs to $\mathcal{E}$, so

$$(y - x, f(y) - f(x)) \in \widetilde{\mathcal{E}}.$$

Therefore

$$f(y) \ge f(x) + \langle g_L, y - x\rangle\,.$$

By Lemma 2.12, extend $g_L$ to some covector $g \in E^*$. Using $y - x \in L$, we still have

$$\forall y \in \operatorname{dom} f, \qquad f(y) \ge f(x) + \langle g, y - x\rangle\,.$$

For $y \notin \operatorname{dom} f$, the same inequality is again automatic, so $g \in \partial f(x)$. Equivalently, by Lemma 2.8, $(g, -1)$ defines a supporting hyperplane of $\operatorname{epi}(f)$ at $(x, f(x))$. ☐

**Lemma 2.12** (Extending a covector from a subspace)**.** *Let $L \subseteq E$ be a linear subspace, and let $g_L \in L^*$. Then there exists $g \in E^*$ such that*

$$\forall v \in L, \qquad \langle g, v\rangle = \langle g_L, v\rangle\,.$$

*Proof of Lemma 2.12.* Choose a linear subspace $M \subseteq E$ such that

$$E = L \oplus M.$$

Every $x \in E$ can then be written uniquely as $x = v + w$ with $v \in L$ and $w \in M$. Define $g : E \to \mathbb{R}$ by

$$g(v + w) := g_L(v).$$

This is well defined by the uniqueness of the decomposition $x = v + w$, and it is linear because

> the projection $E \to L$ along $M$ is linear. Thus $g \in E^*$, and for every $v \in L$ we have
>
> $$g(v) = g_L(v).$$
>
> This is exactly the desired extension. $\square$

## Exercises

**Exercise 2.1.** Prove Proposition 2.5. Hint for item 6: try $T := \{(x, y) \in \mathbb{R}^2 : x > 0,\ y \geq 1/x\} \cup \{(0, 0)\}$.

**Exercise 2.2.** Fix the auxiliary Euclidean norm $\|\cdot\|_2$ from the ambient convention and let $f(x) = \|x\|_2$ on $E$. Compute $\partial f(0)$ and $\partial f(x)$ for $x \neq 0$.