

Lecture 4: Convex conjugates, and Marginal duality

Lecture 3 showed strong duality first in the polyhedral world of linear programming. Lecture 4 now passes from linear inequality certificates to the duality of convex functions and their epigraphs. The main geometric idea is that the conjugate records affine functions lying below f . For each $\xi \in E^*$, the affine function

$$x \mapsto \langle \xi, x \rangle - f^*(\xi)$$

is the tightest such affine lower bound with slope ξ . Therefore the biconjugate

$$f^{**}(x) = \sup_{\xi \in E^*} \{\langle \xi, x \rangle - f^*(\xi)\}$$

is the supremum of all affine functions lying below f . Geometrically, those affine lower bounds are exactly the affine halfspaces that contain $\text{epi}(f)$, so Fenchel–Moreau will be proved below from the fact that a closed convex set is the intersection of its supporting halfspaces.

4.1 Conjugates and Biconjugation

Definition 4.1 (Convex conjugate and biconjugate). Let $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper. Its convex conjugate is defined by

$$\forall \xi \in E^*, \quad f^*(\xi) := \sup_{x \in E} \{\langle \xi, x \rangle - f(x)\}.$$

Under the present convention, this keeps f^* valued in $\mathbb{R} \cup \{+\infty\}$. If one allowed an improper function such as $f \equiv +\infty$, then the same formula would give $f^* \equiv -\infty$, which lies outside the current extended-value convention. Let

$$\iota_E : E \rightarrow E^{**}, \quad (\iota_E(x))(\xi) := \langle \xi, x \rangle.$$

Since E is finite-dimensional, ι_E is an isomorphism. We therefore view the biconjugate back on E through this natural identification and write

$$f^{**}(x) := (f^*)^*(\iota_E(x)) = \sup_{\xi \in E^*} \{\langle \xi, x \rangle - f^*(\xi)\} \quad (x \in E).$$

1

Definition 4.2 (Closure and closed extended-value function). Let $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$. The *closure* of f , denoted $\text{cl } f$, is the function characterized by

$$\text{epi}(\text{cl } f) = \text{cl}(\text{epi}(f)).$$

¹The local Verso page for Lecture 4 contains a one-dimensional conjugate animation: [open the Lecture 4 Verso page](#) on the course website.



(a) The function f and its biconjugate f^{**} .

(b) Three representative affine lower bounds of f .

Figure 1: One-dimensional illustration of biconjugation for $f(x) = \min \{(x-1)^2 + 1, (x+1)^2 + 1\}$, whose biconjugate is $f^{**}(x) = (\max \{|x| - 1, 0\})^2 + 1$. Thus f^{**} agrees with f on the two outer branches and flattens the middle interval $[-1, 1]$. Panel (b) shows three representative affine lower bounds of f ; taking the supremum over all such affine lower bounds recovers f^{**} .

We say that f is *closed* if $\text{epi}(f)$ is a closed subset of $E \times \mathbb{R}$, equivalently if $\text{cl } f = f$.^a

^aIn finite dimensions, this is equivalent to lower semicontinuity.

Lemma 4.1 (Fenchel–Young inequality and equality case). *Let $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper. Then*

$$\forall x \in E, \forall \xi \in E^*, \quad f(x) + f^*(\xi) \geq \langle \xi, x \rangle.$$

If in addition f is closed and convex, then for every $x \in E$ and every $\xi \in E^$,*

$$f(x) + f^*(\xi) = \langle \xi, x \rangle \iff \xi \in \partial f(x).$$

Proof idea of Lemma 4.1. By definition, $f^*(\xi)$ is the smallest shift that makes the affine function $z \mapsto \langle \xi, z \rangle - f^*(\xi)$ lie below f for every z . Evaluating at $z = x$ gives the inequality, and equality means that this affine lower bound touches f at x , which is exactly the subgradient condition.

Lemma 4.2. *Let $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper. Then $\forall x \in E, f^{**}(x) \leq f(x)$.*

This can be read as a weak-duality statement for biconjugation: f^{**} always gives a lower bound on f . Unlike Theorem 4.3, this direction does not require convexity or closedness.

Theorem 4.3 (Fenchel–Moreau). *Let E be a finite-dimensional real normed space, and let $f : E \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, closed, and convex. Then*

$$\forall x \in E, \quad f^{**}(x) = f(x).$$

More generally, if f is proper and convex, then

$$(\text{cl } f)^* = f^*.$$

Applying Theorem 4.3 to $\text{cl } f$ therefore gives

$$f^{**} = (\text{cl } f)^{**} = \text{cl } f.$$

Proof idea of Theorem 4.3. An affine function lying below f , say $x \mapsto \langle y, x \rangle - a$, is the same thing as a closed halfspace

$$H_{\xi,a} := \{(x, t) \in E \times \mathbb{R} : t \geq \langle \xi, x \rangle - a\}$$

that contains $\text{epi}(f)$. Since f is closed and convex, $\text{epi}(f)$ is a closed convex set, so by [Theorem 2.6](#) it is the intersection of all closed halfspaces that contain it. For fixed x , the smallest height t allowed by that intersection is therefore the supremum of all affine lower bounds evaluated at x , namely

$$\sup_{\xi \in E^*} \{\langle \xi, x \rangle - f^*(\xi)\} = f^{**}(x).$$

This is exactly the height of $\text{epi}(f)$ above x , namely $f(x)$.

4.2 Basic Conjugate Pairs

With the structural theorem in place, it is useful to collect a few conjugate pairs that will recur later in the course.

Example 4.1 (A one-dimensional conjugate pair). Let $p, q > 1$ satisfy $\frac{1}{p} + \frac{1}{q} = 1$, and define $f(x) := \frac{|x|^p}{p}$ on \mathbb{R} . Then $f^*(\xi) = \frac{|\xi|^q}{q}$ for $\xi \in \mathbb{R}$. Thus

$$\left(\frac{|x|^p}{p}\right)^* = \frac{|\xi|^q}{q}.$$

This is the basic one-dimensional power-law conjugate pair.^a

^aApplying [Lemma 4.1](#) to this pair gives $x\xi \leq \frac{|x|^p}{p} + \frac{|\xi|^q}{q}$ for all $x, \xi \in \mathbb{R}$, and in particular $ab \leq \frac{a^p}{p} + \frac{b^q}{q}$ for $a, b \geq 0$. This is the classical Young inequality, which explains the ‘Young’ part of the name ‘Fenchel–Young inequality’.

Definition 4.3 (Indicator function). Let $C \subseteq E$. The indicator function of C is the function $\delta_C : E \rightarrow \mathbb{R} \cup \{+\infty\}$ defined by

$$\delta_C(x) := \begin{cases} 0, & x \in C, \\ +\infty, & x \notin C. \end{cases}$$

Example 4.2 (Indicator functions and support functions). Let $C \subseteq E$ be nonempty, and define its support function by

$$\sigma_C(\xi) := \sup_{x \in C} \langle \xi, x \rangle \quad (\xi \in E^*).$$

Then

$$(\delta_C)^* = \sigma_C.$$

In one dimension, if $C = [-1, 1] \subseteq \mathbb{R}$, then

$$\sigma_C(\xi) = \sup_{x \in [-1, 1]} \xi x = |\xi|.$$

Thus

$$(\delta_{[-1,1]})^* = |\cdot|.$$

Since $|\cdot|$ is proper, closed, and convex, [Theorem 4.3](#) then gives

$$|\cdot|^* = \delta_{[-1,1]}.$$

So the pair $|x|$ and $\delta_{[-1,1]}$ is the one-dimensional special case of the general indicator/support-function correspondence.

Example 4.3 (Norms and dual unit balls). Let $\|\cdot\|$ be a norm on E , and let $\|\cdot\|_*$ be its dual norm on E^* , defined by

$$\|\xi\|_* := \sup_{\|x\| \leq 1} \langle \xi, x \rangle.$$

Then

$$\|\cdot\|^* = \delta_{B_*}, \quad B_* := \{\xi \in E^* : \|\xi\|_* \leq 1\}.$$

Indeed, if $\|\xi\|_* \leq 1$, then

$$\langle \xi, x \rangle - \|x\| \leq \|\xi\|_* \|x\| - \|x\| \leq 0 \quad \forall x \in E,$$

so $\|\cdot\|^*(\xi) = 0$. If $\|\xi\|_* > 1$, choose $x_0 \in E$ such that $\langle \xi, x_0 \rangle > \|x_0\|$; then

$$\langle \xi, tx_0 \rangle - \|tx_0\| = t(\langle \xi, x_0 \rangle - \|x_0\|) \rightarrow +\infty \quad (t \rightarrow +\infty),$$

so $\|\cdot\|^*(\xi) = +\infty$.

Example 4.4 (Exponential and negative entropy). Let $f(x) = e^x$ on \mathbb{R} . Then

$$f^*(\xi) = \begin{cases} \xi \log \xi - \xi, & \xi \geq 0, \\ +\infty, & \xi < 0, \end{cases}$$

with the convention $0 \log 0 := 0$. This is the basic exponential/entropy conjugate pair.

4.3 Marginal Duality and Applications

For subsets $A, B \subseteq Y$, write $A - B := \{a - b : a \in A, b \in B\} = A + (-B)$.

Theorem 4.4 (Marginal duality / perturbation duality). *Let X and U be finite-dimensional real vector spaces, and let*

$$\Phi : X \times U \rightarrow \mathbb{R} \cup \{+\infty\}$$

be convex. Define the marginal value function of Φ as

$$p : U \rightarrow \mathbb{R} \cup \{\pm\infty\}, \quad p(u) := \inf_{x \in X} \Phi(x, u).$$

Here p may take the value $-\infty$, even though Φ itself only takes values in $\mathbb{R} \cup \{+\infty\}$. Then the following hold.

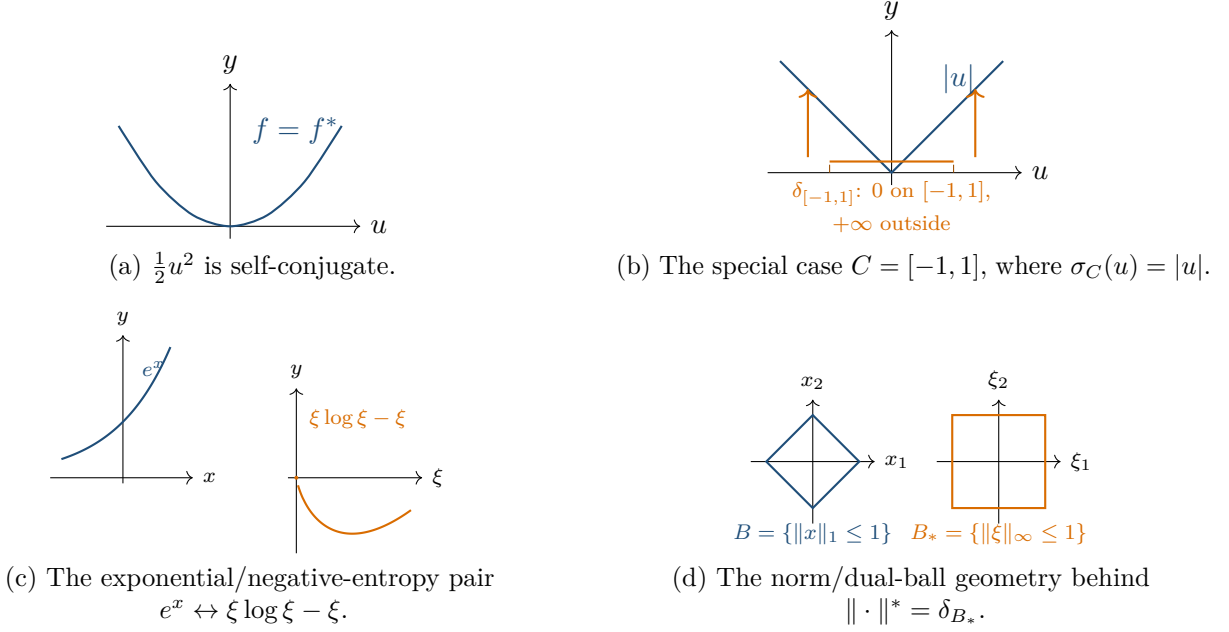


Figure 2: Four basic scalar pictures. Panel (a) shows the self-conjugate quadratic. Panel (b) is the one-dimensional special case $C = [-1, 1]$ of Example 4.2, where the support function is $|u|$. Panel (c) shows the exponential/negative-entropy conjugate pair from Example 4.4. Panel (d) shows the geometric norm/dual-ball picture behind Example 4.3.

1. For every $y \in U^*$,

$$p^*(y) = \Phi^*(0, y).$$

Consequently, for every $u \in U$ and every $y \in U^*$,

$$p(u) \geq -\Phi^*(0, y) + \langle y, u \rangle.$$

In particular,

$$\sup_{y \in U^*} \{-\Phi^*(0, y)\} \leq p(0).$$

2. If $p(0) = -\infty$, then

$$\sup_{y \in U^*} \{-\Phi^*(0, y)\} = -\infty.$$

3. If $p(0) > -\infty$ and $\partial p(0) \neq \emptyset$, then

$$p(0) = \max_{y \in U^*} \{-\Phi^*(0, y)\}.$$

In particular, strong duality and dual attainment hold.

4. Define

$$D := \{u \in U : \exists x \in X \text{ such that } \Phi(x, u) < +\infty\} = \text{dom } p.$$

If $p(0) \in \mathbb{R}$ and $0 \in \text{ri}(D)$, then

$$p(0) = \max_{y \in U^*} \{-\Phi^*(0, y)\}.$$

One especially useful specialization of [Theorem 4.4](#) is the template

$$\inf_{x \in X} \{f(x) + g(Ax)\}.$$

We postpone that reduction to [Exercise 4.1](#). The worked applications below instead compute the relevant perturbation conjugates directly from [Theorem 4.4](#).

Example 4.5 (Recovering the LP dual). Consider the linear program

$$\inf \{c^\top x : Ax \geq b, x \geq 0\},$$

with $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$. Define the perturbation

$$\Phi(x, u) := c^\top x + \delta_{\mathbb{R}_+^n}(x) + \delta_{\mathbb{R}_+^m}(Ax - b - u), \quad (x, u) \in \mathbb{R}^n \times \mathbb{R}^m.$$

Then

$$p(u) := \inf_{x \in \mathbb{R}^n} \Phi(x, u) = \inf \{c^\top x : Ax \geq b + u, x \geq 0\}.$$

For $y \in \mathbb{R}^m$,

$$\Phi^*(0, y) = \begin{cases} -b^\top y, & A^\top y \leq c, y \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Therefore [Theorem 4.4](#) recovers the dual formula

$$\sup \{b^\top y : A^\top y \leq c, y \geq 0\} \leq \inf \{c^\top x : Ax \geq b, x \geq 0\},$$

and, if $0 \in \text{ri}(\text{dom } p)$ and the primal value is finite, it gives

$$\inf \{c^\top x : Ax \geq b, x \geq 0\} = \max \{b^\top y : A^\top y \leq c, y \geq 0\}.$$

This recovers the usual LP dual from the general perturbation template. [Lecture 3](#) proves a sharper strong-duality statement using polyhedral structure, without this extra relative-interior hypothesis.

Proof of [Example 4.5](#). We compute the dual formula in four steps. The main calculation in *Step 2* does not introduce any extra variable; *Step 3* then revisits the same conjugate by introducing a slack variable, because that second route is the one that generalizes most directly to conic programs.

Step 1: compute the perturbation value function $p(u)$. By definition,

$$\Phi(x, u) = c^\top x + \delta_{\mathbb{R}_+^n}(x) + \delta_{\mathbb{R}_+^m}(Ax - b - u).$$

This is finite exactly when

$$x \geq 0 \quad \text{and} \quad Ax - b - u \in \mathbb{R}_+^m,$$

that is, exactly when

$$x \geq 0 \quad \text{and} \quad Ax \geq b + u.$$

Therefore

$$p(u) := \inf_{x \in \mathbb{R}^n} \Phi(x, u) = \inf \{c^\top x : Ax \geq b + u, x \geq 0\}.$$

In particular,

$$p(0) = \inf \{c^\top x : Ax \geq b, x \geq 0\}.$$

Step 2 (direct): compute $\Phi^*(0, y)$ without introducing a new variable. Fix $y \in \mathbb{R}^m$. Starting from the definition of the conjugate,

$$\Phi^*(0, y) = \sup_{x \in \mathbb{R}^n, u \in \mathbb{R}^m} \{y^\top u - c^\top x - \delta_{\mathbb{R}_+^n}(x) - \delta_{\mathbb{R}_+^m}(Ax - b - u)\}.$$

The indicator terms force $x \geq 0$ and $u \leq Ax - b$. Therefore

$$\Phi^*(0, y) = \sup_{x \geq 0} \left\{ \sup_{u \leq Ax - b} y^\top u - c^\top x \right\}.$$

The inner supremum is

$$\sup_{u \leq Ax - b} y^\top u = \begin{cases} y^\top (Ax - b), & y \geq 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

because if $y \geq 0$, the linear form $y^\top u$ is maximized at the largest allowed value $u = Ax - b$, while if some coordinate of y is negative, the corresponding coordinate of u can be sent to $-\infty$. Hence

$$\Phi^*(0, y) = \begin{cases} -b^\top y + \sup_{x \geq 0} \{x^\top (A^\top y - c)\}, & y \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Finally,

$$\sup_{x \geq 0} \{x^\top (A^\top y - c)\} = \begin{cases} 0, & A^\top y \leq c, \\ +\infty, & \text{otherwise,} \end{cases}$$

so

$$\Phi^*(0, y) = \begin{cases} -b^\top y, & A^\top y \leq c, y \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Alternative Step 2 (slack variable): the same computation with a slack variable. The direct calculation above is already enough, but it is helpful to see the more structured version that will reappear in conic duality. Starting again from the constraint $Ax - b - u \in \mathbb{R}_+^m$, write

$$s := Ax - b - u \in \mathbb{R}_+^m, \quad \text{so that} \quad u = Ax - b - s.$$

Substituting this into the objective gives

$$\begin{aligned} \Phi^*(0, y) &= \sup_{x \geq 0, s \geq 0} \{y^\top (Ax - b - s) - c^\top x\} \\ &= -b^\top y + \sup_{x \geq 0} \{x^\top (A^\top y - c)\} + \sup_{s \geq 0} \{(-y)^\top s\}. \end{aligned}$$

Now these two suprema can be read coordinatewise.

For the s -part:

$$\sup_{s \geq 0} \{(-y)^\top s\} = \begin{cases} 0, & y \geq 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

because if $y \geq 0$, then $(-y)^\top s \leq 0$ for every $s \geq 0$, with equality at $s = 0$; while if some coordinate of y is negative, one can send the corresponding coordinate of s to $+\infty$.

For the x -part:

$$\sup_{x \geq 0} \{x^\top (A^\top y - c)\} = \begin{cases} 0, & A^\top y \leq c, \\ +\infty, & \text{otherwise,} \end{cases}$$

because if $A^\top y \leq c$, then $x^\top (A^\top y - c) \leq 0$ for every $x \geq 0$, with equality at $x = 0$; while if some coordinate of $A^\top y - c$ is positive, one can scale the corresponding basis vector.

Combining the two pieces yields the same formula as above:

$$\Phi^*(0, y) = \begin{cases} -b^\top y, & A^\top y \leq c, y \geq 0, \\ +\infty, & \text{otherwise.} \end{cases}$$

Step 3: substitute into Theorem 4.4. Part (1) of Theorem 4.4 gives

$$\sup_{y \in \mathbb{R}^m} \{-\Phi^*(0, y)\} \leq p(0),$$

which becomes

$$\sup \{b^\top y : A^\top y \leq c, y \geq 0\} \leq \inf \{c^\top x : Ax \geq b, x \geq 0\}.$$

This is exactly the usual weak-duality inequality for the LP pair.

If in addition $0 \in \text{ri}(\text{dom } p)$ and $p(0) \in \mathbb{R}$, then part (4) of Theorem 4.4 upgrades the inequality to equality and gives dual attainment:

$$\inf \{c^\top x : Ax \geq b, x \geq 0\} = \max \{b^\top y : A^\top y \leq c, y \geq 0\}.$$

□

Example 4.6 (Norm regularization and dual-norm constraints). Let X and Y be finite-dimensional real normed spaces, let $A : X \rightarrow Y$ be linear, let $\lambda > 0$, and let $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ be proper, closed, and convex. Consider

$$\inf_{x \in X} \{f(x) + \lambda \|Ax\|\}.$$

Then

$$\inf_{x \in X} \{f(x) + \lambda \|Ax\|\} = \max_{\substack{y \in Y^* \\ \|y\|_* \leq \lambda}} \{-f^*(-A^*y)\}.$$

Proof of Example 4.6. Again, the calculation has three parts.

Step 1: choose the perturbation and identify $p(0)$. Set

$$g(u) := \lambda \|u\|, \quad \Phi(x, u) := f(x) + g(Ax + u).$$

Then the marginal value function is

$$p(u) := \inf_{x \in X} \Phi(x, u) = \inf_{x \in X} \{f(x) + \lambda \|Ax + u\|\}.$$

At $u = 0$, this is exactly the original problem:

$$p(0) = \inf_{x \in X} \{f(x) + \lambda \|Ax\|\}.$$

Because g is finite everywhere on Y , one has $\text{dom } g = Y$. Since f is proper, there exists $x_0 \in X$ with $f(x_0) < +\infty$. For every $u \in Y$,

$$p(u) \leq f(x_0) + \lambda \|Ax_0 + u\| < +\infty,$$

so

$$\text{dom } p = Y, \quad 0 \in \text{ri}(\text{dom } p).$$

Thus the qualification condition in part (4) of [Theorem 4.4](#) will be automatic as soon as $p(0) \in \mathbb{R}$.

Step 2: compute $\Phi^(0, y)$.* Fix $y \in Y^*$. By definition,

$$\Phi^*(0, y) = \sup_{x \in X, u \in Y} \{\langle y, u \rangle - f(x) - g(Ax + u)\}.$$

The standard trick is to replace $Ax + u$ by a new variable:

$$z := Ax + u.$$

Because u ranges freely over all of Y , this change of variables is bijective: for each pair $(x, z) \in X \times Y$, there is a unique $u = z - Ax \in Y$. Thus

$$\begin{aligned} \Phi^*(0, y) &= \sup_{x \in X, z \in Y} \{\langle y, z - Ax \rangle - f(x) - g(z)\} \\ &= \sup_{x \in X, z \in Y} \{\langle y, z \rangle - \langle A^*y, x \rangle - f(x) - g(z)\} \\ &= \sup_{x \in X} \{\langle -A^*y, x \rangle - f(x)\} + \sup_{z \in Y} \{\langle y, z \rangle - g(z)\} \\ &= f^*(-A^*y) + g^*(y). \end{aligned}$$

So the remaining work is to compute g^* .

Step 3: compute g^ and substitute into the marginal theorem.* For $y \in Y^*$,

$$g^*(y) = \sup_{u \in Y} \{\langle y, u \rangle - \lambda \|u\|\}.$$

If $\|y\|_* \leq \lambda$, then for every $u \in Y$,

$$\langle y, u \rangle \leq \|y\|_* \|u\| \leq \lambda \|u\|,$$

hence

$$\langle y, u \rangle - \lambda \|u\| \leq 0.$$

Taking $u = 0$ shows that the supremum is exactly 0.

If $\|y\|_* > \lambda$, then by the definition of the dual norm there exists $u_0 \in Y$ such that

$$\langle y, u_0 \rangle > \lambda \|u_0\|.$$

Scaling u_0 gives

$$\langle y, tu_0 \rangle - \lambda \|tu_0\| = t(\langle y, u_0 \rangle - \lambda \|u_0\|) \rightarrow +\infty \quad (t \rightarrow +\infty),$$

so $g^*(y) = +\infty$. Therefore

$$g^*(y) = \begin{cases} 0, & \|y\|_* \leq \lambda, \\ +\infty, & \|y\|_* > \lambda. \end{cases}$$

In other words, the conjugate of the norm penalty is the indicator of the dual norm ball of radius λ .

Now apply [Theorem 4.4](#). If $p(0) = -\infty$, then part (2) already gives

$$\sup_{y \in Y^*} \{-\Phi^*(0, y)\} = -\infty,$$

so the desired identity holds trivially. If instead $p(0) > -\infty$, then $p(0) \in \mathbb{R}$ because $p(0) < +\infty$, and part (4) gives

$$p(0) = \max_{y \in Y^*} \{-\Phi^*(0, y)\}.$$

Substituting the formula for $\Phi^*(0, y)$ and then the formula for $g^*(y)$ produces

$$\inf_{x \in X} \{f(x) + \lambda \|Ax\|\} = \max_{\substack{y \in Y^* \\ \|y\|_* \leq \lambda}} \{-f^*(-A^*y)\}.$$

This is the desired dual problem. □

Proofs

Proof of Lemma 4.1. By definition of the conjugate,

$$f^*(y) = \sup_{z \in E} \{\langle y, z \rangle - f(z)\} \geq \langle y, x \rangle - f(x),$$

which rearranges to

$$f(x) + f^*(y) \geq \langle x, y \rangle.$$

Assume now that f is proper, closed, and convex. Equality holds if and only if

$$f^*(y) = \langle y, x \rangle - f(x).$$

By the definition of f^* , this is equivalent to

$$\forall z \in E, \quad \langle y, z \rangle - f(z) \leq \langle y, x \rangle - f(x),$$

that is,

$$\forall z \in E, \quad f(z) \geq f(x) + \langle y, z - x \rangle.$$

This is exactly the statement that $y \in \partial f(x)$. □

Proof of Example 4.1. By definition,

$$f^*(\xi) = \sup_{x \in \mathbb{R}} \left\{ x\xi - \frac{|x|^p}{p} \right\}.$$

The maximizer has the same sign as ξ , so it is enough to consider $x \geq 0$ and $\xi \geq 0$. For fixed $\xi \geq 0$, set $\varphi_\xi(x) := x\xi - \frac{x^p}{p}$ on $x \geq 0$. Since $\varphi'_\xi(x) = \xi - x^{p-1}$, the unique critical point is

$x = \xi^{1/(p-1)} = \xi^{q-1}$. Substituting gives

$$f^*(\xi) = \xi \cdot \xi^{q-1} - \frac{\xi^{p(q-1)}}{p} = \xi^q - \frac{\xi^q}{p} = \frac{\xi^q}{q}.$$

By symmetry, $f^*(\xi) = \frac{|\xi|^q}{q}$ for all $\xi \in \mathbb{R}$. Applying [Lemma 4.1](#) to this pair then gives the classical Young inequality stated in the footnote. \square

Proof of Lemma 4.2. For every $x \in E$ and every $\xi \in E^*$, [Lemma 4.1](#) gives

$$\langle \xi, x \rangle - f^*(\xi) \leq f(x).$$

Taking the supremum over ξ yields

$$f^{**}(x) \leq f(x) \quad \forall x \in E.$$

\square

Proof of Theorem 4.3. By [Lemma 4.2](#), it remains to prove the reverse inequality

$$f(x) \leq f^{**}(x) \quad \forall x \in E.$$

Fix $x_0 \in E$ and $t_0 < f(x_0)$. Then $(x_0, t_0) \notin \text{epi}(f)$. Because f is proper, closed, and convex, the epigraph $\text{epi}(f)$ is a nonempty closed convex subset of $E \times \mathbb{R}$. By [Theorem 2.6](#), there exists a closed halfspace containing $\text{epi}(f)$ but not (x_0, t_0) . Since $\text{epi}(f)$ is upward closed in the t -direction, we may take that halfspace in the form

$$H_{\xi, a} := \{(x, t) \in E \times \mathbb{R} : t \geq \langle \xi, x \rangle - a\}$$

for some $\xi \in E^*$ and $a \in \mathbb{R}$. Thus

$$\text{epi}(f) \subseteq H_{\xi, a} \quad \text{and} \quad t_0 < \langle \xi, x_0 \rangle - a.$$

The containment $\text{epi}(f) \subseteq H_{\xi, a}$ is equivalent to saying that the affine function

$$\ell_{\xi, a}(x) := \langle \xi, x \rangle - a$$

is a global affine lower bound for f , i.e.

$$\ell_{\xi, a}(x) \leq f(x) \quad \forall x \in E.$$

For fixed slope ξ , the smallest admissible intercept is exactly $f^*(\xi)$, because

$$\ell_{\xi, a}(x) \leq f(x) \quad \forall x \iff a \geq \sup_{x \in E} \{\langle \xi, x \rangle - f(x)\} = f^*(\xi).$$

Hence every affine lower bound of slope ξ lies below the tight affine function of the same slope,

$$x \mapsto \langle \xi, x \rangle - f^*(\xi).$$

In particular,

$$t_0 < \langle \xi, x_0 \rangle - a \leq \langle \xi, x_0 \rangle - f^*(\xi) \leq f^{**}(x_0).$$

Since this holds for every $t_0 < f(x_0)$, we conclude that

$$f(x_0) \leq f^{**}(x_0).$$

Combined with Lemma 4.2, this gives

$$\forall x \in E, \quad f^{**}(x) = f(x).$$

□

Proof of Theorem 4.4. Because Φ is convex, its epigraph is a convex subset of $X \times U \times \mathbb{R}$. The epigraph of p is the projection

$$\text{epi}(p) = \{(u, t) \in U \times \mathbb{R} : \exists x \in X \text{ with } \Phi(x, u) \leq t\}.$$

Hence $\text{epi}(p)$ is convex, so p is convex.

Fix $y \in U^*$. By definition,

$$\begin{aligned} p^*(y) &= \sup_{u \in U} \{\langle y, u \rangle - p(u)\} \\ &= \sup_{u \in U} \sup_{x \in X} \{\langle y, u \rangle - \Phi(x, u)\} \\ &= \sup_{x \in X, u \in U} \{\langle 0, x \rangle + \langle y, u \rangle - \Phi(x, u)\} \\ &= \Phi^*(0, y). \end{aligned}$$

This proves the identity in part (1). For every $u \in U$, the definition of $p^*(y)$ gives

$$p^*(y) = \sup_{v \in U} \{\langle y, v \rangle - p(v)\} \geq \langle y, u \rangle - p(u).$$

Using $p^*(y) = \Phi^*(0, y)$, we obtain

$$p(u) + \Phi^*(0, y) \geq \langle y, u \rangle,$$

which is exactly

$$p(u) \geq -\Phi^*(0, y) + \langle y, u \rangle.$$

Assume now that $p(0) = -\infty$. For every $M > 0$, choose $x_M \in X$ such that

$$\Phi(x_M, 0) \leq -M.$$

Then, for every $y \in U^*$,

$$\Phi^*(0, y) \geq \langle 0, x_M \rangle + \langle y, 0 \rangle - \Phi(x_M, 0) \geq M.$$

Since M is arbitrary, $\Phi^*(0, y) = +\infty$ for every y , and hence

$$\sup_{y \in U^*} \{-\Phi^*(0, y)\} = -\infty.$$

This proves part (2).

Assume next that $p(0) > -\infty$ and choose $y \in \partial p(0)$. Then

$$\forall u \in U, \quad p(u) \geq p(0) + \langle y, u \rangle.$$

Equivalently,

$$\forall u \in U, \quad \langle y, u \rangle - p(u) \leq -p(0).$$

Taking the supremum over u gives

$$p^*(y) \leq -p(0).$$

On the other hand, evaluating at $u = 0$ gives

$$p^*(y) \geq \langle y, 0 \rangle - p(0) = -p(0).$$

Hence

$$p^*(y) = -p(0).$$

Using part (1) of [Theorem 4.4](#),

$$p(0) = -p^*(y) = -\Phi^*(0, y).$$

Combined with the weak-duality inequality from part (1), this yields

$$p(0) = \max_{v \in U^*} \{-\Phi^*(0, v)\},$$

with the maximum attained at the chosen y . This proves part (3).

Finally, define

$$D := \{u \in U : \exists x \in X \text{ such that } \Phi(x, u) < +\infty\}.$$

If $u \in D$, then there exists $x \in X$ such that $\Phi(x, u) < +\infty$, so

$$p(u) = \inf_{x' \in X} \Phi(x', u) \leq \Phi(x, u) < +\infty,$$

hence $u \in \text{dom } p$. Conversely, if $u \in \text{dom } p$, then $p(u) < +\infty$, so not all values $\Phi(x, u)$ can equal $+\infty$; therefore there exists $x \in X$ such that $\Phi(x, u) < +\infty$, and hence $u \in D$. Thus

$$D = \text{dom } p.$$

Assume now that $p(0) \in \mathbb{R}$ and $0 \in \text{ri}(D) = \text{ri}(\text{dom } p)$. By convexity of p , it is enough to show that p never takes the value $-\infty$ on $\text{dom } p$. Fix $u \in \text{dom } p$. Since $0 \in \text{ri}(\text{dom } p)$, there exists $\lambda \in (0, 1)$ such that $\lambda u \in \text{dom } p$. Then $p(0)$ and $p(\lambda u)$ are both finite, and convexity gives

$$p(\lambda u) \leq \lambda p(u) + (1 - \lambda)p(0).$$

Therefore $p(u) > -\infty$. Hence $p : U \rightarrow \mathbb{R} \cup \{+\infty\}$ is proper and convex, with $0 \in \text{ri}(\text{dom } p)$. By [Theorem 2.10](#), we obtain $\partial p(0) \neq \emptyset$. Part (3) now applies and proves part (4). \square

Exercises

Exercise 4.1. Fenchel–Rockafellar specialization of [Theorem 4.4](#). Let X and Y be finite-dimensional real vector spaces, let $A : X \rightarrow Y$ be linear, and let

$$f : X \rightarrow \mathbb{R} \cup \{+\infty\}, \quad g : Y \rightarrow \mathbb{R} \cup \{+\infty\}$$

be proper, closed, and convex. Do the following.

1. Define

$$\Phi(x, u) := f(x) + g(Ax + u)$$

and prove that Φ is convex on $X \times Y$.

2. Define

$$p(u) := \inf_{x \in X} \Phi(x, u) = \inf_{x \in X} \{f(x) + g(Ax + u)\}.$$

Show that

$$p(0) = \inf_{x \in X} \{f(x) + g(Ax)\}.$$

3. Prove that for every $y \in Y^*$,

$$\Phi^*(0, y) = f^*(-A^*y) + g^*(y).$$

4. Show that

$$\text{dom } p = \text{dom } g - A(\text{dom } f) = -(A(\text{dom } f) - \text{dom } g).$$

Deduce that

$$0 \in \text{ri}(A(\text{dom } f) - \text{dom } g) \iff 0 \in \text{ri}(\text{dom } p).$$

5. Use part (1) of [Theorem 4.4](#) to prove

$$\sup_{y \in Y^*} \{-f^*(-A^*y) - g^*(y)\} \leq \inf_{x \in X} \{f(x) + g(Ax)\}.$$

6. Assume additionally that

$$0 \in \text{ri}(A(\text{dom } f) - \text{dom } g)$$

and that the primal infimum is finite. Use part (4) of [Theorem 4.4](#) to prove

$$\inf_{x \in X} \{f(x) + g(Ax)\} = \max_{y \in Y^*} \{-f^*(-A^*y) - g^*(y)\}.$$

Exercise 4.2. Dual norm twice. Let $\|\cdot\|$ be a norm on E , and let $\|\cdot\|_*$ be its dual norm on E^* . Prove that, under the natural identification $E \simeq E^{**}$, the dual norm of $\|\cdot\|_*$ is the original norm:

$$\|x\|_{**} = \|x\| \quad \forall x \in E.$$