Chapter 4

Conjugate Functions

4.1 Definition and Basic Properties

We begin with the definition of the conjugate function.

Definition 4.1 (conjugate functions). *Let $f : \mathbb{E} \to [-\infty, \infty]$ be an extended real-valued function. The function $f^* : \mathbb{E}^* \to [-\infty, \infty]$, defined by*

$$f^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x})\}, \quad \mathbf{y} \in \mathbb{E}^*,$$

is called the **conjugate function of f**.

Example 4.2 (conjugate of indicator functions). Let $f = \delta_C$, where $C \subseteq \mathbb{E}$ is nonempty. Then for any $\mathbf{y} \in \mathbb{E}^*$

$$f^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} \rangle - \delta_C(\mathbf{x})\} = \max_{\mathbf{x} \in C} \langle \mathbf{y}, \mathbf{x} \rangle = \sigma_C(\mathbf{y}).$$

That is, the conjugate of the indicator function is the support function of the same underlying set:

$$\boxed{\delta_C^* = \sigma_C.}$$

∎

Two fundamental properties of conjugate functions are their convexity and closedness (regardless of whether the original function is closed or convex).

Theorem 4.3 (convexity and closedness of conjugate functions). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be an extended real-valued function. Then the conjugate function f^* is closed and convex.*

Proof. Note that f^* is the pointwise maximum of affine functions, which are convex and closed, and thus, invoking Theorems 2.16(c) and 2.7(c), it follows that f^* is closed and convex. ∎

Example 4.4 (conjugate of $\frac{1}{2}\|\cdot\|^2 + \delta_C$). Suppose that $\mathbb{E}$ is Euclidean and that $C \subseteq \mathbb{E}$ is nonempty. Define $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 + \delta_C(\mathbf{x})$. Then by Example 2.17 (specifically, (2.6)), it follows that

$$f^*(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|^2 - \frac{1}{2}d_C^2(\mathbf{y}).$$

Note that while f is convex only if C is convex, the convexity of f^* is guaranteed regardless of whether C is convex or not. ∎

The next result states that the conjugate function of a proper convex function is also proper.

Theorem 4.5 (properness of conjugate functions). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper convex function. Then f^* is proper.*

Proof. Since f is proper, it follows that there exists $\hat{\mathbf{x}} \in \mathbb{E}$ such that $f(\hat{\mathbf{x}}) < \infty$. By the definition of the conjugate function, for any $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) \geq \langle \mathbf{y}, \hat{\mathbf{x}} \rangle - f(\hat{\mathbf{x}}),$$

and hence $f^*(\mathbf{y}) > -\infty$. What remains in order to establish the properness of f^* is to show that there exists $\mathbf{g} \in \mathbb{E}^*$ such that $f^*(\mathbf{g}) < \infty$. By Corollary 3.19, there exists $\mathbf{x} \in \mathrm{dom}(f)$ such that $\partial f(\mathbf{x}) \neq \emptyset$. Take $\mathbf{g} \in \partial f(\mathbf{x})$. Then by the definition of the subgradient, for any $\mathbf{z} \in \mathbb{E}$,

$$f(\mathbf{z}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{z} - \mathbf{x} \rangle.$$

Hence,

$$f^*(\mathbf{g}) = \max_{\mathbf{z} \in \mathbb{E}} \{\langle \mathbf{g}, \mathbf{z} \rangle - f(\mathbf{z})\} \leq \langle \mathbf{g}, \mathbf{x} \rangle - f(\mathbf{x}) < \infty,$$

concluding that f^* is a proper function. ∎

The following result, called *Fenchel's inequality*, is a trivial implication of the definition of conjugacy.

Theorem 4.6 (Fenchel's inequality). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be an extended real-valued proper function. Then for any $\mathbf{x} \in \mathbb{E}$ and $\mathbf{y} \in \mathbb{E}^*$,*

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

Proof. By the definition of the conjugate function we have that for any $\mathbf{x} \in \mathbb{E}$ and $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) \geq \langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}). \tag{4.1}$$

Since f is proper, it follows that $f(\mathbf{x}), f^*(\mathbf{y}) > -\infty$. We can thus add $f(\mathbf{x})$ to both sides of (4.1) and obtain the desired result. ∎

4.2 The Biconjugate

The conjugacy operation can be invoked twice resulting in the biconjugate operation. Specifically, for a function $f : \mathbb{E} \to [-\infty, \infty]$ we define (recall that in this book $\mathbb{E}$ and $\mathbb{E}^{**}$ are considered to be identical)

$$f^{**}(\mathbf{x}) = \max_{\mathbf{y} \in \mathbb{E}^*} \left\{ \langle \mathbf{x}, \mathbf{y} \rangle - f^*(\mathbf{y}) \right\}, \quad \mathbf{x} \in \mathbb{E}.$$

The biconjugate function is always a lower bound on the original function, as the following result states.

Lemma 4.7 ($f^{**} \leq f$). *Let $f : \mathbb{E} \to [-\infty, \infty]$ be an extended real-valued function. Then $f(\mathbf{x}) \geq f^{**}(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{E}$.*

Proof. By the definition of the conjugate function we have for any $\mathbf{x} \in \mathbb{E}$ and $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) \geq \langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}).$$

Thus,

$$f(\mathbf{x}) \geq \langle \mathbf{y}, \mathbf{x} \rangle - f^*(\mathbf{y}),$$

implying that

$$f(\mathbf{x}) \geq \max_{\mathbf{y} \in \mathbb{E}^*} \{ \langle \mathbf{y}, \mathbf{x} \rangle - f^*(\mathbf{y}) \} = f^{**}(\mathbf{x}). \quad \square$$

If we assume that f is proper closed and convex, then the biconjugate is not just a lower bound on f—it is equal to f.

Theorem 4.8 ($f = f^{**}$ for proper closed convex functions). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed and convex function. Then $f^{**} = f$.*

Proof. By Lemma 4.7, $f^{**} \leq f$. We thus need to show that $f^{**} \geq f$. Suppose by contradiction that for some $\mathbf{x} \in \mathbb{E}$ we have $f^{**}(\mathbf{x}) < f(\mathbf{x})$. This means that $(\mathbf{x}, f^{**}(\mathbf{x})) \notin \mathrm{epi}(f) \subseteq \mathbb{E} \times \mathbb{R}$. We assume as usual that the product space $\mathbb{V} = \mathbb{E} \times \mathbb{R}$ is endowed with the inner product $\langle (\mathbf{u}, s), (\mathbf{v}, t) \rangle_{\mathbb{V}} = \langle \mathbf{u}, \mathbf{v} \rangle + st$, where $\langle \cdot, \cdot \rangle$ is the inner product associated with $\mathbb{E}$ (see Section 1.9). Since f is proper closed and convex, the set $\mathrm{epi}(f)$ is nonempty closed and convex, and hence, by the strict separation theorem (Theorem 2.33), there exist $\mathbf{a} \in \mathbb{E}^*, b, c_1, c_2 \in \mathbb{R}$ such that

$$\langle \mathbf{a}, \mathbf{z} \rangle + bs \leq c_1 < c_2 \leq \langle \mathbf{a}, \mathbf{x} \rangle + b f^{**}(\mathbf{x}) \text{ for all } (\mathbf{z}, s) \in \mathrm{epi}(f).$$

We can thus conclude that

$$\langle \mathbf{a}, \mathbf{z} - \mathbf{x} \rangle + b(s - f^{**}(\mathbf{x})) \leq c_1 - c_2 \equiv c < 0 \text{ for all } (\mathbf{z}, s) \in \mathrm{epi}(f). \tag{4.2}$$

The scalar b must be nonpositive, since otherwise, if it was positive, the inequality would have been violated by taking a fixed $\mathbf{z}$ and large enough s. We will now consider two cases.

- If $b < 0$, then dividing (4.2) by $-b$ and taking $\mathbf{y} = -\frac{\mathbf{a}}{b}$, we get

$$\langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle - s + f^{**}(\mathbf{x}) \leq \frac{c}{-b} < 0 \text{ for all } (\mathbf{z}, s) \in \mathrm{epi}(f).$$

In particular, taking $s = f(\mathbf{z})$ (which is possible since $(\mathbf{z}, f(\mathbf{z})) \in \mathrm{epi}(f)$), we obtain that

$$\langle \mathbf{y}, \mathbf{z} \rangle - f(\mathbf{z}) - \langle \mathbf{y}, \mathbf{x} \rangle + f^{**}(\mathbf{x}) \leq \frac{c}{-b} < 0 \text{ for all } \mathbf{z} \in \mathbb{E}.$$

Taking the maximum over $\mathbf{z}$ yields the inequality

$$f^*(\mathbf{y}) - \langle \mathbf{y}, \mathbf{x} \rangle + f^{**}(\mathbf{x}) \leq \frac{c}{-b} < 0,$$

which is a contradiction of Fenchel's inequality (Theorem 4.6).

- If $b = 0$, then take some $\hat{\mathbf{y}} \in \mathrm{dom}(f^*)$. Such a vector exists since f^* is proper by the properness and convexity of f (Theorem 4.5). Let $\varepsilon > 0$ and define $\hat{\mathbf{a}} = \mathbf{a} + \varepsilon \hat{\mathbf{y}}$ and $\hat{b} = -\varepsilon$. Then for any $\mathbf{z} \in \mathrm{dom}(f)$,

$$\begin{aligned} \langle \hat{\mathbf{a}}, \mathbf{z} - \mathbf{x} \rangle + \hat{b}(f(\mathbf{z}) - f^{**}(\mathbf{x})) &= \langle \mathbf{a}, \mathbf{z} - \mathbf{x} \rangle + \varepsilon[\langle \hat{\mathbf{y}}, \mathbf{z} \rangle - f(\mathbf{z}) + f^{**}(\mathbf{x}) - \langle \hat{\mathbf{y}}, \mathbf{x} \rangle] \\ &\leq c + \varepsilon[\langle \hat{\mathbf{y}}, \mathbf{z} \rangle - f(\mathbf{z}) + f^{**}(\mathbf{x}) - \langle \hat{\mathbf{y}}, \mathbf{x} \rangle] \\ &\leq c + \varepsilon[f^*(\hat{\mathbf{y}}) - \langle \hat{\mathbf{y}}, \mathbf{x} \rangle + f^{**}(\mathbf{x})], \end{aligned}$$

where the first inequality is due to (4.2) and the second by the definition of $f^*(\hat{\mathbf{y}})$ as the maximum of $\langle \hat{\mathbf{y}}, \mathbf{z} \rangle - f(\mathbf{z})$ over all possible $\mathbf{z} \in \mathbb{E}$. We thus obtained the inequality

$$\langle \hat{\mathbf{a}}, \mathbf{z} - \mathbf{x} \rangle + \hat{b}(f(\mathbf{z}) - f^{**}(\mathbf{x})) \leq \hat{c}, \tag{4.3}$$

where $\hat{c} \equiv c + \varepsilon[f^*(\hat{\mathbf{y}}) - \langle \hat{\mathbf{y}}, \mathbf{x} \rangle + f^{**}(\mathbf{x})]$. Since $c < 0$, we can pick $\varepsilon > 0$ small enough to ensure that $\hat{c} < 0$. At this point we employ exactly the same argument used in the first case. Dividing (4.3) by $-\hat{b}$ and denoting $\tilde{\mathbf{y}} = -\frac{1}{\hat{b}}\hat{\mathbf{a}}$ yields the inequality

$$\langle \tilde{\mathbf{y}}, \mathbf{z} \rangle - f(\mathbf{z}) - \langle \tilde{\mathbf{y}}, \mathbf{x} \rangle + f^{**}(\mathbf{x}) \leq -\frac{\hat{c}}{\hat{b}} < 0 \text{ for any } \mathbf{z} \in \mathrm{dom}(f).$$

Taking the maximum over $\mathbf{z}$ results in

$$f^*(\tilde{\mathbf{y}}) - \langle \tilde{\mathbf{y}}, \mathbf{x} \rangle + f^{**}(\mathbf{x}) \leq \frac{\hat{c}}{-\hat{b}} < 0,$$

which, again, is a contradiction of Fenchel's inequality.   □

Example 4.9 (conjugate of support functions). We will now show how to exploit Theorem 4.8 in order to compute the conjugate of support functions. Suppose that $C \subseteq \mathbb{E}$ is a given nonempty set. Since $\mathrm{cl}(\mathrm{conv}(C))$ is closed and convex, it follows that $\delta_{\mathrm{cl}(\mathrm{conv}(C))}$ is closed and convex, and hence, by Example 4.2 and Theorem 4.8,

$$\sigma^*_{\mathrm{cl}(\mathrm{conv}(C))} = (\delta^*_{\mathrm{cl}(\mathrm{conv}(C))})^* = \delta^{**}_{\mathrm{cl}(\mathrm{conv}(C))} = \delta_{\mathrm{cl}(\mathrm{conv}(C))}. \tag{4.4}$$

Finally, by Lemma 2.35,

$$\sigma_C = \sigma_{\mathrm{cl}(\mathrm{conv}(C))},$$

which, combined with (4.4), establishes the result

$$\sigma_C^* = \delta_{\mathrm{cl}(\mathrm{conv}(C))}.$$

∎

Example 4.10 (conjugate of the max function). Consider the function $f :$ $\mathbb{R}^n \to \mathbb{R}$ given by $f(\mathbf{x}) = \max\{x_1, x_2, \ldots, x_n\}$. Note that the following elementary identity holds for any $\mathbf{x} \in \mathbb{R}^n$:

$$\max\{x_1, x_2, \ldots, x_n\} = \max_{\mathbf{y} \in \Delta_n} \mathbf{y}^T \mathbf{x} = \sigma_{\Delta_n}(\mathbf{x}).$$

Therefore, using Example 4.9, we can conclude, exploiting the convexity and closedness of Δ_n, that

$$f^* = \delta_{\Delta_n}.$$

∎

Example 4.11 (conjugate of $\frac{1}{2}\|\cdot\|^2 - d_C^2$). Let $\mathbb{E}$ be Euclidean, and let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Define $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2 - \frac{1}{2}d_C^2(\mathbf{x})$. By Example 4.4, $f = g^*$, where $g(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|^2 + \delta_C(\mathbf{y})$. By the nonemptiness, closedness, and convexity of C, it follows that g is proper closed and convex, and hence, by Theorem 4.8,

$$f^*(\mathbf{y}) = g^{**}(\mathbf{y}) = g(\mathbf{y}) = \frac{1}{2}\|\mathbf{y}\|^2 + \delta_C(\mathbf{y}).$$

∎

4.3 Conjugate Calculus Rules

In this section we present the basic calculus rules for computing conjugate functions. We begin with a very simple rule for separable functions.

Theorem 4.12 (conjugate of separable functions). *Let $g : \mathbb{E}_1 \times \mathbb{E}_2 \times \cdots \times \mathbb{E}_p \to (-\infty, \infty]$ be given by $g(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p) = \sum_{i=1}^p f_i(\mathbf{x}_i)$, where $f_i : \mathbb{E}_i \to (-\infty, \infty]$ is a proper function for any $i = 1, 2, \ldots, p$. Then*

$$g^*(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p) = \sum_{i=1}^p f_i^*(\mathbf{y}_i) \text{ for any } \mathbf{y}_i \in \mathbb{E}_i^*, \quad i = 1, 2, \ldots, p.$$

Proof. For any $(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p) \in \mathbb{E}_1^* \times \mathbb{E}_2^* \times \cdots \times \mathbb{E}_p^*$, it holds that

$$g^*(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p) = \max_{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p} \{\langle (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_p), (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p) \rangle - g(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p)\}$$

$$= \max_{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p} \left\{ \sum_{i=1}^p \langle \mathbf{y}_i, \mathbf{x}_i \rangle - \sum_{i=1}^p f_i(\mathbf{x}_i) \right\}$$

$$= \sum_{i=1}^p \max_{\mathbf{x}_i} \{\langle \mathbf{y}_i, \mathbf{x}_i \rangle - f_i(\mathbf{x}_i)\}$$

$$= \sum_{i=1}^p f_i^*(\mathbf{y}_i). \quad \square$$

The next result shows how the conjugate operation is affected by invertible affine change of variables as well as by addition of an affine function.

Theorem 4.13 (conjugate of $f(\mathcal{A}(\mathbf{x} - \mathbf{a})) + \langle \mathbf{b}, \mathbf{x} \rangle + c$). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be an extended real-valued function, and let $\mathcal{A} : \mathbb{V} \to \mathbb{E}$ be an invertible linear transformation, $\mathbf{a} \in \mathbb{V}$, $\mathbf{b} \in \mathbb{V}^*$, and $c \in \mathbb{R}$. Then the conjugate of the function $g(\mathbf{x}) = f(\mathcal{A}(\mathbf{x} - \mathbf{a})) + \langle \mathbf{b}, \mathbf{x} \rangle + c$ is given by*

$$g^*(\mathbf{y}) = f^*\left((\mathcal{A}^T)^{-1}(\mathbf{y} - \mathbf{b})\right) + \langle \mathbf{a}, \mathbf{y} \rangle - c - \langle \mathbf{a}, \mathbf{b} \rangle, \quad \mathbf{y} \in \mathbb{V}^*.$$

Proof. Making the change of variables $\mathbf{z} = \mathcal{A}(\mathbf{x} - \mathbf{a})$, which is equivalent to $\mathbf{x} = \mathcal{A}^{-1}(\mathbf{z}) + \mathbf{a}$, we can write for any $\mathbf{y} \in \mathbb{V}^*$,

$$
\begin{aligned}
g^*(\mathbf{y}) &= \max_{\mathbf{x}}\{\langle \mathbf{y}, \mathbf{x} \rangle - g(\mathbf{x})\} \\
&= \max_{\mathbf{x}}\{\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathcal{A}(\mathbf{x} - \mathbf{a})) - \langle \mathbf{b}, \mathbf{x} \rangle - c\} \\
&= \max_{\mathbf{z}}\{\langle \mathbf{y}, \mathcal{A}^{-1}(\mathbf{z}) + \mathbf{a} \rangle - f(\mathbf{z}) - \langle \mathbf{b}, \mathcal{A}^{-1}(\mathbf{z}) + \mathbf{a} \rangle - c\} \\
&= \max_{\mathbf{z}}\left\{\langle \mathbf{y} - \mathbf{b}, \mathcal{A}^{-1}(\mathbf{z}) \rangle - f(\mathbf{z}) + \langle \mathbf{a}, \mathbf{y} \rangle - \langle \mathbf{a}, \mathbf{b} \rangle - c\right\} \\
&= \max_{\mathbf{z}}\left\{\langle (\mathcal{A}^{-1})^T(\mathbf{y} - \mathbf{b}), \mathbf{z} \rangle - f(\mathbf{z}) + \langle \mathbf{a}, \mathbf{y} \rangle - \langle \mathbf{a}, \mathbf{b} \rangle - c\right\} \\
&= f^*\left((\mathcal{A}^T)^{-1}(\mathbf{y} - \mathbf{b})\right) + \langle \mathbf{a}, \mathbf{y} \rangle - c - \langle \mathbf{a}, \mathbf{b} \rangle,
\end{aligned}
$$

where in the last equality we also used the fact that $(\mathcal{A}^{-1})^T = (\mathcal{A}^T)^{-1}$. $\square$

Theorem 4.14 (conjugate of $\alpha f(\cdot)$ and $\alpha f(\cdot/\alpha)$). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be an extended real-valued function and let $\alpha \in \mathbb{R}_{++}$.*

(a) *The conjugate of the function $g(\mathbf{x}) = \alpha f(\mathbf{x})$ is given by*

$$g^*(\mathbf{y}) = \alpha f^*\left(\frac{\mathbf{y}}{\alpha}\right), \quad \mathbf{y} \in \mathbb{E}^*.$$

(b) *The conjugate of the function $h(\mathbf{x}) = \alpha f\left(\frac{\mathbf{x}}{\alpha}\right)$ is given by*

$$h^*(\mathbf{y}) = \alpha f^*(\mathbf{y}), \quad \mathbf{y} \in \mathbb{E}^*.$$

Proof. For any $\mathbf{y} \in \mathbb{E}^*$,

$$
\begin{aligned}
g^*(\mathbf{y}) &= \max_{\mathbf{x}}\{\langle \mathbf{y}, \mathbf{x} \rangle - g(\mathbf{x})\} \\
&= \max_{\mathbf{x}}\{\langle \mathbf{y}, \mathbf{x} \rangle - \alpha f(\mathbf{x})\} \\
&= \alpha \max_{\mathbf{x}}\left\{\left\langle \frac{\mathbf{y}}{\alpha}, \mathbf{x} \right\rangle - f(\mathbf{x})\right\} \\
&= \alpha f^*\left(\frac{\mathbf{y}}{\alpha}\right),
\end{aligned}
$$

proving (a). The proof of (b) follows by the following chain of equalities:

$$
\begin{aligned}
h^*(\mathbf{y}) &= \max_{\mathbf{x}}\{\langle \mathbf{y}, \mathbf{x} \rangle - h(\mathbf{x})\} \\
&= \max_{\mathbf{x}}\left\{\langle \mathbf{y}, \mathbf{x} \rangle - \alpha f\left(\frac{\mathbf{x}}{\alpha}\right)\right\} \\
&= \alpha \max_{\mathbf{x}}\left\{\left\langle \mathbf{y}, \frac{\mathbf{x}}{\alpha}\right\rangle - f\left(\frac{\mathbf{x}}{\alpha}\right)\right\} \\
&\overset{\mathbf{z}\leftarrow\frac{\mathbf{x}}{\alpha}}{=} \alpha \max_{\mathbf{z}}\left\{\langle \mathbf{y}, \mathbf{z} \rangle - f(\mathbf{z})\right\} \\
&= \alpha f^*(\mathbf{y}). \qquad \square
\end{aligned}
$$

The table below summarizes the four calculus rules discussed in this section.

$g(\mathbf{x})$	$g^*(\mathbf{y})$	Reference
$\sum_{i=1}^m f_i(\mathbf{x}_i)$	$\sum_{i=1}^m f_i^*(\mathbf{y}_i)$	Theorem 4.12
$\alpha f(\mathbf{x})\ \ (\alpha > 0)$	$\alpha f^*(\mathbf{y}/\alpha)$	Theorem 4.14
$\alpha f(\mathbf{x}/\alpha)\ \ (\alpha > 0)$	$\alpha f^*(\mathbf{y})$	Theorem 4.14
$f(\mathcal{A}(\mathbf{x} - \mathbf{a})) + \langle \mathbf{b}, \mathbf{x} \rangle + c$	$f^*\left((\mathcal{A}^T)^{-1}(\mathbf{y} - \mathbf{b})\right) + \langle \mathbf{a}, \mathbf{y} \rangle - c - \langle \mathbf{a}, \mathbf{b} \rangle$	Theorem 4.13

4.4 Examples

In this section we compute the conjugate functions of several fundamental convex functions. The first examples are one-dimensional, while the rest are multidimensional.

4.4.1 Exponent

Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = e^x$. Then for any $y \in \mathbb{R}$,

$$
f^*(y) = \max_x \left\{xy - e^x\right\}. \tag{4.5}
$$

If $y < 0$, then the maximum value of the above problem is ∞ (easily seen by taking $x \to -\infty$). If $y = 0$, then obviously the maximal value (which is not attained) is 0. If $y > 0$, the unique maximizer of (4.5) is $x = \tilde{x} \equiv \log y$. Consequently, $f^*(y) = \tilde{x}y - e^{\tilde{x}} = y \log y - y$ for any $y > 0$. Using the convention $0 \log 0 \equiv 0$, we can finally deduce that

$$
f^*(y) = \begin{cases} y \log y - y, & y \geq 0, \\ \infty & \text{else.} \end{cases}
$$

4.4.2 Negative Log

Let $f : \mathbb{R} \to (-\infty, \infty]$ be given by

$$
f(x) = \begin{cases} -\log(x), & x > 0, \\ \infty, & x \leq 0. \end{cases}
$$

For any $y \in \mathbb{R}$,

$$f^*(y) = \max_{x>0}\{xy - f(x)\} = \max_{x>0}\{xy + \log(x)\}. \tag{4.6}$$

If $y \geq 0$, then the maximum value of the above problem is ∞ (since the objective function in (4.6) goes to ∞ as $x \to \infty$). If $y < 0$, the unique optimal solution of (4.6) is attained at $\tilde{x} = -\frac{1}{y}$, and hence for $y < 0$ we have $f^*(y) = \tilde{x}y + \log(\tilde{x}) = -1 - \log(-y)$. To conclude,

$$f^*(y) = \begin{cases} -1 - \log(-y), & y < 0, \\ \\ \infty, & y \geq 0. \end{cases}$$

4.4.3   Hinge Loss

Consider the one-dimensional function $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) = \max\{1 - x, 0\}.$$

Then for any $y \in \mathbb{R}$,

$$f^*(y) = \max_x \left[yx - \max\{1 - x, 0\}\right] = \max_x \left[\min\left\{(1+y)x - 1, yx\right\}\right]. \tag{4.7}$$

The objective function in the above maximization problem can be rewritten as

$$\min\left\{(1+y)x - 1, yx\right\} = \begin{cases} (1+y)x - 1, & x < 1, \\ \\ yx, & x \geq 1. \end{cases}$$

Thus, the objective function is a continuous piecewise linear function comprising two pieces: a line with slope $1 + y$ over $(-\infty, 1]$ and a line with slope y over $[1, \infty)$. Therefore, a maximizer exists if the slope of the left line is nonnegative $(1 + y \geq 0)$ and the slope of the right line is nonpositive $(y \leq 0)$. Consequently, a maximizer exists for the problem in (4.7) if and only if $y \in [-1, 0]$, and in that case it is attained at $x = 1$, with y being the corresponding optimal value. To summarize,

$$f^*(y) = y + \delta_{[-1,0]}(y), \quad y \in \mathbb{R}.$$

4.4.4 $\frac{1}{p}|\cdot|^p$ $(p > 1)$

Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = \frac{1}{p}|x|^p$, where $p > 1$. For any $y \in \mathbb{R}$,

$$f^*(y) = \max_x \left\{xy - \frac{1}{p}|x|^p\right\}. \tag{4.8}$$

Since the problem in (4.8) consists of maximizing a differentiable concave function over $\mathbb{R}$, its optimal solutions are the points $\tilde{x}$ in which the derivative vanishes:

$$y - \text{sgn}(\tilde{x})|\tilde{x}|^{p-1} = 0.$$

Therefore, $\operatorname{sgn}(\tilde{x}) = \operatorname{sgn}(y)$ and $|\tilde{x}|^{p-1} = |y|$, implying that $\tilde{x} = \operatorname{sgn}(y)|y|^{\frac{1}{p-1}}$. Thus,

$$f^*(y) = \tilde{x}y - \frac{1}{p}|\tilde{x}|^p = |y|^{1+\frac{1}{p-1}} - \frac{1}{p}|y|^{\frac{p}{p-1}} = \left(1 - \frac{1}{p}\right)|y|^{\frac{p}{p-1}} = \frac{1}{q}|y|^q,$$

where q is the positive number satisfying $\frac{1}{p} + \frac{1}{q} = 1$. To summarize,

$$f^*(y) = \frac{1}{q}|y|^q, \quad y \in \mathbb{R}.$$

4.4.5 $\quad -\frac{(\cdot)^p}{p} \quad (0 < p < 1)$

Let $f : \mathbb{R} \to (-\infty, \infty]$ be given by

$$f(x) = \begin{cases} -\frac{x^p}{p}, & x \geq 0, \\ \infty, & x < 0. \end{cases}$$

For any $y \in \mathbb{R}$,

$$f^*(y) = \max_x \{xy - f(x)\} = \max_{x \geq 0} \left\{ g(x) \equiv xy + \frac{x^p}{p} \right\}.$$

When $y \geq 0$, the value of the above problem is ∞ since $g(x) \to \infty$ as $x \to \infty$. If $y < 0$, then the derivative of $g(x)$ vanishes at $x = \tilde{x} \equiv (-y)^{\frac{1}{p-1}} > 0$, and since g is concave, it follows that $\tilde{x}$ is a global maximizer of g. Therefore,

$$f^*(y) = \tilde{x}y + \frac{\tilde{x}^p}{p} = -(-y)^{\frac{p}{p-1}} + \frac{1}{p}(-y)^{\frac{p}{p-1}} = -\frac{(-y)^q}{q},$$

where q is the negative number for which $\frac{1}{p} + \frac{1}{q} = 1$. To summarize,

$$f^*(y) = \begin{cases} -\frac{(-y)^q}{q}, & y < 0, \\ \infty, & \text{else.} \end{cases}$$

4.4.6 Strictly Convex Quadratic Functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$, where $\mathbf{A} \in \mathbb{S}_{++}^n$, $\mathbf{b} \in \mathbb{R}^n$, and $c \in \mathbb{R}$. We use our convention that (unless otherwise stated) $\mathbb{R}^n$ is endowed with the dot product, meaning that $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{y}$. For any $\mathbf{y} \in \mathbb{R}^n$,

$$\begin{aligned} f^*(\mathbf{y}) &= \max_{\mathbf{x}} \{\mathbf{y}^T\mathbf{x} - f(\mathbf{x})\} \\ &= \max_{\mathbf{x}} \left\{ \mathbf{y}^T\mathbf{x} - \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - \mathbf{b}^T\mathbf{x} - c \right\} \\ &= \max_{\mathbf{x}} \left\{ -\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} - (\mathbf{b} - \mathbf{y})^T\mathbf{x} - c \right\}. \end{aligned}$$

The maximum in the above problem is attained at $\mathbf{x} = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b})$, leading to the following expression for the conjugate function:

$$f^*(\mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{b})^T \mathbf{A}^{-1}(\mathbf{y} - \mathbf{b}) - c.$$

4.4.7   Convex Quadratic Functions

Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$, where $\mathbf{A} \in \mathbb{S}^n_+$, $\mathbf{b} \in \mathbb{R}^n$ and $c \in \mathbb{R}$. The only difference between this example and the previous one is the fact that here $\mathbf{A}$ is not necessarily positive definite but is assumed to be only positive *semi*definite. For any $\mathbf{y} \in \mathbb{R}^n$,

$$f^*(\mathbf{y}) = \max_{\mathbf{x}} \{\mathbf{y}^T \mathbf{x} - f(\mathbf{x})\} = \max_{\mathbf{x}} \left\{ g(\mathbf{x}) \equiv -\frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + (\mathbf{y} - \mathbf{b})^T \mathbf{x} - c \right\}.$$

Since g is concave and differentiable over $\mathbb{R}^n$, it follows that the maximizers of the above problem are the points for which the gradient vanishes, namely, points $\mathbf{x}$ satisfying

$$\mathbf{A}\mathbf{x} = \mathbf{y} - \mathbf{b}. \tag{4.9}$$

This system has a solution if and only if $\mathbf{y} \in \mathbf{b} + \text{Range}(\mathbf{A})$, and in that case we can choose one of the solutions to the system (4.9), for example, $\tilde{\mathbf{x}} = \mathbf{A}^\dagger(\mathbf{y} - \mathbf{b})$, where $\mathbf{A}^\dagger$ is the Moore–Penrose pseudoinverse of $\mathbf{A}$. We can now compute $f^*(\mathbf{y})$ as follows:

$$\begin{aligned}
f^*(\mathbf{y}) &= -\frac{1}{2}\tilde{\mathbf{x}}^T \mathbf{A}\tilde{\mathbf{x}} - (\mathbf{b} - \mathbf{y})^T \tilde{\mathbf{x}} - c \\
&= -\frac{1}{2}(\mathbf{y} - \mathbf{b})\mathbf{A}^\dagger \mathbf{A}\mathbf{A}^\dagger(\mathbf{y} - \mathbf{b}) - (\mathbf{b} - \mathbf{y})^T \mathbf{A}^\dagger(\mathbf{y} - \mathbf{b}) - c \\
&= \frac{1}{2}(\mathbf{y} - \mathbf{b})^T \mathbf{A}^\dagger(\mathbf{y} - \mathbf{b}) - c,
\end{aligned}$$

where we used the fact that the Moore–Penrose pseudoinverse of a symmetric matrix is symmetric, as well as the known identity $\mathbf{A}^\dagger \mathbf{A}\mathbf{A}^\dagger = \mathbf{A}^\dagger$. We are left with the case where $\mathbf{y} - \mathbf{b} \notin \text{Range}(\mathbf{A})$. We will show that in this case $f^*(\mathbf{y}) = \infty$. Indeed, since $\text{Range}(\mathbf{A}) = \text{Null}(\mathbf{A})^\perp$, it follows that $\mathbf{y} - \mathbf{b} \notin \text{Null}(\mathbf{A})^\perp$, meaning that there exists a vector $\mathbf{v} \in \text{Null}(\mathbf{A})$ such that $(\mathbf{y} - \mathbf{b})^T \mathbf{v} > 0$. Note that for any $\alpha \in \mathbb{R}$,

$$g(\alpha \mathbf{v}) = \alpha(\mathbf{y} - \mathbf{b})^T \mathbf{v} - c,$$

and hence $g(\alpha \mathbf{v}) \to \infty$ as α tends to ∞, establishing the fact that $f^*(\mathbf{y}) = \infty$ whenever $\mathbf{y} \notin \mathbf{b} + \text{Range}(\mathbf{A})$. To conclude,

$$f^*(\mathbf{y}) = \begin{cases} \frac{1}{2}(\mathbf{y} - \mathbf{b})^T \mathbf{A}^\dagger(\mathbf{y} - \mathbf{b}) - c, & \mathbf{y} \in \mathbf{b} + \text{Range}(\mathbf{A}), \\ \infty & \text{else.} \end{cases}$$

4.4.8   Negative Entropy

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be given by

$$f(\mathbf{x}) = \begin{cases} \sum_{i=1}^{n} x_i \log x_i, & \mathbf{x} \geq \mathbf{0}, \\ \infty & \text{else.} \end{cases}$$

Since the function is separable, it is enough to compute the conjugate of the scalar function g defined by $g(t) = t \log t$ for $t \geq 0$ and ∞ for $t < 0$. For any $s \in \mathbb{R}$,

$$g^*(s) = \max_t \{ts - g(t)\} = \max_{t \geq 0} \{ts - t \log t\}.$$

The maximum of the above problem is attained at $t = e^{s-1}$, and hence the conjugate is given by

$$g^*(s) = se^{s-1} - (s-1)e^{s-1} = e^{s-1}.$$

Since $f(\mathbf{x}) = \sum_{i=1}^n g(x_i)$, it follows by Theorem 4.12 that for any $\mathbf{y} \in \mathbb{R}^n$,

$$f^*(\mathbf{y}) = \sum_{i=1}^n g^*(y_i) = \sum_{i=1}^n e^{y_i - 1}.$$

4.4.9 Negative Sum of Logs

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be given by

$$f(\mathbf{x}) = \begin{cases} -\sum_{i=1}^n \log x_i, & \mathbf{x} > \mathbf{0}, \\ \infty & \text{else.} \end{cases}$$

Note that $f(\mathbf{x}) = \sum_{i=1}^n g(x_i)$, where $g(t) = -\log t$ for $t > 0$ and ∞ for $t \leq 0$. Therefore, invoking Theorem 4.12,

$$f^*(\mathbf{x}) = \sum_{i=1}^n g^*(x_i).$$

By Section 4.4.2, $g^*(y) = -1 - \log(-y)$ for $y < 0$ and ∞ otherwise. Therefore,

$$f^*(\mathbf{y}) = \begin{cases} -n - \sum_{i=1}^n \log(-y_i), & \mathbf{y} < \mathbf{0}, \\ \infty & \text{else.} \end{cases}$$

4.4.10 Negative Entropy over the Unit Simplex

Let $f : \mathbb{R}^n \to (-\infty, \infty]$ be given by

$$f(\mathbf{x}) = \begin{cases} \sum_{i=1}^n x_i \log x_i, & \mathbf{x} \in \Delta_n, \\ \infty & \text{else.} \end{cases} \tag{4.10}$$

For any $\mathbf{y} \in \mathbb{R}^n$,

$$f^*(\mathbf{y}) = \max \left\{ \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \log x_i : \sum_{i=1}^n x_i = 1, x_1, x_2, \ldots, x_n \geq 0 \right\}.$$

By Example 3.71, the optimal solution of the above maximization problem is

$$x_i^* = \frac{e^{y_i}}{\sum_{j=1}^n e^{y_j}}, \quad i = 1, 2, \ldots, n,$$

with a corresponding optimal value of

$$f^*(\mathbf{y}) = \sum_{i=1}^n y_i x_i^* - \sum_{i=1}^n x_i^* \log x_i^* = \log\left(\sum_{j=1}^n e^{y_j}\right).$$

That is, the conjugate of the negative entropy is the log-sum-exp function.

4.4.11   log-sum-exp

Let $g : \mathbb{R}^n \to \mathbb{R}$ be given by

$$g(\mathbf{x}) = \log\left(\sum_{j=1}^n e^{x_j}\right).$$

By Section 4.4.10, $g = f^*$, where f is the negative entropy over the unit simplex given by (4.10). Since f is proper closed and convex, it follows by Theorem 4.8 that $f^{**} = f$, and hence

$$g^* = f^{**} = f,$$

meaning that

$$g^*(\mathbf{y}) = \begin{cases} \sum_{i=1}^n y_i \log y_i, & \mathbf{y} \in \Delta_n, \\ \infty & \text{else.} \end{cases}$$

4.4.12   Norms

Let $f : \mathbb{E} \to \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|$. Then, by Example 2.31,

$$f = \sigma_{B_{\|\cdot\|_*}[\mathbf{0},1]},$$

where we used the fact that the bidual norm $\|\cdot\|_{**}$ is identical to the norm $\|\cdot\|$. Hence, by Example 4.9,

$$f^* = \delta_{\mathrm{cl}(\mathrm{conv}(B_{\|\cdot\|_*}[\mathbf{0},1]))},$$

but since $B_{\|\cdot\|_*}[\mathbf{0}, 1]$ is closed and convex, $\mathrm{cl}(\mathrm{conv}(B_{\|\cdot\|_*}[\mathbf{0}, 1])) = B_{\|\cdot\|_*}[\mathbf{0}, 1]$, and therefore for any $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) = \delta_{B_{\|\cdot\|_*}[\mathbf{0},1]}(\mathbf{y}) = \begin{cases} 0, & \|\mathbf{y}\|_* \leq 1, \\ \infty & \text{else.} \end{cases}$$

4.4.13 Ball-Pen

Let $f : \mathbb{E} \to (-\infty, \infty]$ be given by

$$
f(\mathbf{x}) = \begin{cases} -\sqrt{1 - \|\mathbf{x}\|^2}, & \|\mathbf{x}\| \leq 1, \\ \infty & \text{else.} \end{cases}
$$

To compute the conjugate function, we begin by rewriting it in a double maximization form:

$$
f^*(\mathbf{y}) = \max_{\mathbf{x}} \left\{ \langle \mathbf{y}, \mathbf{x} \rangle + \sqrt{1 - \|\mathbf{x}\|^2} : \|\mathbf{x}\| \leq 1 \right\}
$$
$$
= \max_{\alpha \in [0,1]} \max_{\mathbf{x} : \|\mathbf{x}\| = \alpha} \left\{ \langle \mathbf{y}, \mathbf{x} \rangle + \sqrt{1 - \alpha^2} \right\}.
$$

By the definition of the dual norm, the optimal value of the inner maximization problem is $\alpha \|\mathbf{y}\|_* + \sqrt{1 - \alpha^2}$, and we can therefore write, for any $\mathbf{y} \in \mathbb{E}^*$,

$$
f^*(\mathbf{y}) = \max_{\alpha \in [0,1]} \left\{ g(\alpha) \equiv \alpha \|\mathbf{y}\|_* + \sqrt{1 - \alpha^2} \right\}. \tag{4.11}
$$

It is easy to see that the maximizer of g over $[0,1]$ is

$$
\tilde{\alpha} = \frac{\|\mathbf{y}\|_*}{\sqrt{\|\mathbf{y}\|_*^2 + 1}}.
$$

Plugging $\alpha = \tilde{\alpha}$ into (4.11), we finally obtain that for any $\mathbf{y} \in \mathbb{E}^*$,

$$
\boxed{f^*(\mathbf{y}) = \sqrt{\|\mathbf{y}\|_*^2 + 1}.}
$$

It is also possible to generalize the result to functions of the form

$$
f_\alpha(\mathbf{x}) = \begin{cases} -\sqrt{\alpha^2 - \|\mathbf{x}\|^2}, & \|\mathbf{x}\| \leq \alpha, \\ \infty & \text{else,} \end{cases}
$$

where $\alpha \in \mathbb{R}_{++}$. In this notation, $f = f_1$. To compute f_α^*, note that $f_\alpha(\mathbf{x}) = \alpha f\left(\frac{\mathbf{x}}{\alpha}\right)$, and hence by Theorem 4.14(b) it follows that for any $\mathbf{y} \in \mathbb{E}^*$,

$$
\boxed{f_\alpha^*(\mathbf{y}) = \alpha f^*(\mathbf{y}) = \alpha \sqrt{1 + \|\mathbf{y}\|_*^2}.}
$$

4.4.14 $\sqrt{\alpha^2 + \|\cdot\|^2}$

Consider the function $g_\alpha : \mathbb{E} \to \mathbb{R}$ given by $g_\alpha(\mathbf{x}) = \sqrt{\alpha^2 + \|\mathbf{x}\|^2}$, where $\alpha > 0$. Then $g_\alpha(\mathbf{x}) = \alpha g\left(\frac{\mathbf{x}}{\alpha}\right)$, where $g(\mathbf{x}) = \sqrt{1 + \|\mathbf{x}\|^2}$. By Section 4.4.13, it follows that $g = f^*$, where f is given by

$$
f(\mathbf{y}) = \begin{cases} -\sqrt{1 - \|\mathbf{y}\|_*^2}, & \|\mathbf{y}\|_* \leq 1, \\ \infty & \text{else.} \end{cases}
$$

Since f is proper closed and convex, it follows by Theorem 4.8 that

$$g^* = f^{**} = f.$$

Finally, invoking Theorem 4.14(b), we conclude that for any $\mathbf{y} \in \mathbb{E}^*$,

$$g_\alpha^*(\mathbf{y}) = \alpha g^*(\mathbf{y}) = \alpha f(\mathbf{y}) = \begin{cases} -\alpha\sqrt{1 - \|\mathbf{y}\|_*^2}, & \|\mathbf{y}\|_* \leq 1, \\ \infty & \text{else.} \end{cases}$$

4.4.15   Squared Norm

Let $f : \mathbb{E} \to \mathbb{R}$ be given by $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x}\|^2$, where $\|\cdot\|$ is the norm associated with the space $\mathbb{E}$. For any $\mathbf{y} \in \mathbb{E}^*$, we can write $f^*(\mathbf{y})$ as the optimal value of the following double maximization problem:

$$f^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbb{E}} \left\{ \langle \mathbf{y}, \mathbf{x} \rangle - \frac{1}{2}\|\mathbf{x}\|^2 \right\} = \max_{\alpha \geq 0} \max_{\mathbf{x}:\|\mathbf{x}\|=\alpha} \left\{ \langle \mathbf{y}, \mathbf{x} \rangle - \frac{1}{2}\alpha^2 \right\}.$$

Using the definition of the dual norm, it follows that

$$\max_{\mathbf{x} \in \mathbb{E}:\|\mathbf{x}\|=\alpha} \langle \mathbf{y}, \mathbf{x} \rangle = \alpha\|\mathbf{y}\|_*,$$

Hence,

$$f^*(\mathbf{y}) = \max_{\alpha \geq 0} \left\{ \alpha\|\mathbf{y}\|_* - \frac{1}{2}\alpha^2 \right\} = \frac{1}{2}\|\mathbf{y}\|_*^2.$$

4.4.16 Summary of Conjugate Computations

The table below summarizes all the computations of conjugate functions described in this chapter.

$f(\mathbf{x})$	$\mathrm{dom}(f)$	f^*	Assumptions	Reference
e^x	$\mathbb{R}$	$y \log y - y$ ($\mathrm{dom}(f^*) = \mathbb{R}_+$)	–	Section 4.4.1
$-\log x$	$\mathbb{R}_{++}$	$-1 - \log(-y)$ ($\mathrm{dom}(f^*) = \mathbb{R}_{--}$)	–	Section 4.4.2
$\max\{1-x, 0\}$	$\mathbb{R}$	$y + \delta_{[-1,0]}(y)$	–	Section 4.4.3
$\frac{1}{p}|x|^p$	$\mathbb{R}$	$\frac{1}{q}|y|^q$	$p > 1, \frac{1}{p} + \frac{1}{q} = 1$	Section 4.4.4
$-\frac{x^p}{p}$	$\mathbb{R}_+$	$-\frac{(-y)^q}{q}$ ($\mathrm{dom}(f^*) = \mathbb{R}_{--}$)	$0 < p < 1, \frac{1}{p} + \frac{1}{q} = 1$	Section 4.4.5
$\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$	$\mathbb{R}^n$	$\frac{1}{2}(\mathbf{y}-\mathbf{b})^T\mathbf{A}^{-1}(\mathbf{y}-\mathbf{b}) - c$	$\mathbf{A} \in \mathbb{S}^n_{++}, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$	Section 4.4.6
$\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$	$\mathbb{R}^n$	$\frac{1}{2}(\mathbf{y}-\mathbf{b})^T\mathbf{A}^\dagger(\mathbf{y}-\mathbf{b}) - c$ ($\mathrm{dom}(f^*) = \mathbf{b} + \mathrm{Range}(\mathbf{A})$)	$\mathbf{A} \in \mathbb{S}^n_+, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$	Section 4.4.7
$\sum_{i=1}^n x_i \log x_i$	$\mathbb{R}^n_+$	$\sum_{i=1}^n e^{y_i - 1}$	–	Section 4.4.8
$\sum_{i=1}^n x_i \log x_i$	Δ_n	$\log\left(\sum_{i=1}^n e^{y_i}\right)$	–	Section 4.4.10
$-\sum_{i=1}^n \log x_i$	$\mathbb{R}^n_{++}$	$-n - \sum_{i=1}^n \log(-y_i)$ ($\mathrm{dom}(f^*) = \mathbb{R}^n_{--}$)	–	Section 4.4.9
$\log\left(\sum_{i=1}^n e^{x_i}\right)$	$\mathbb{R}^n$	$\sum_{i=1}^n y_i \log y_i$ ($\mathrm{dom}(f^*) = \Delta_n$)	–	Section 4.4.11
$\max_i\{x_i\}$	$\mathbb{R}^n$	$\delta_{\Delta_n}(\mathbf{y})$	–	Example 4.10
$\delta_C(\mathbf{x})$	C	$\sigma_C(\mathbf{y})$	$\emptyset \neq C \subseteq \mathbb{E}$	Example 4.2
$\sigma_C(\mathbf{x})$	$\mathrm{dom}(\sigma_C)$	$\delta_{\mathrm{cl}(\mathrm{conv}(C))}(\mathbf{y})$	$\emptyset \neq C \subseteq \mathbb{E}$	Example 4.9
$\|\mathbf{x}\|$	$\mathbb{E}$	$\delta_{B_{\|\cdot\|_*}[\mathbf{0},1]}(\mathbf{y})$	–	Section 4.4.12
$-\sqrt{\alpha^2 - \|\mathbf{x}\|^2}$	$B[\mathbf{0},\alpha]$	$\alpha\sqrt{\|\mathbf{y}\|_*^2 + 1}$	$\alpha > 0$	Section 4.4.13
$\sqrt{\alpha^2 + \|\mathbf{x}\|^2}$	$\mathbb{E}$	$-\alpha\sqrt{1 - \|\mathbf{y}\|_*^2}$ ($\mathrm{dom} f^* = B_{\|\cdot\|_*}[\mathbf{0},1]$)	$\alpha > 0$	Section 4.4.14
$\frac{1}{2}\|\mathbf{x}\|^2$	$\mathbb{E}$	$\frac{1}{2}\|\mathbf{y}\|_*^2$	–	Section 4.4.15
$\frac{1}{2}\|\mathbf{x}\|^2 + \delta_C(\mathbf{x})$	C	$\frac{1}{2}\|\mathbf{y}\|^2 - \frac{1}{2}d_C^2(\mathbf{y})$	$\emptyset \neq C \subseteq \mathbb{E}$, $\mathbb{E}$ Euclidean	Example 4.4
$\frac{1}{2}\|\mathbf{x}\|^2 - \frac{1}{2}d_C^2(\mathbf{x})$	$\mathbb{E}$	$\frac{1}{2}\|\mathbf{y}\|^2 + \delta_C(\mathbf{y})$	$\emptyset \neq C \subseteq \mathbb{E}$ closed convex. $\mathbb{E}$ Euclidean	Example 4.11

4.4.17   Fenchel's Duality Theorem

Conjugate functions naturally appear in dual problems most prominently in the celebrated *Fenchel's duality theorem*, which we now recall. Consider the problem

$$\text{(P)} \ \min_{\mathbf{x} \in \mathbb{E}} f(\mathbf{x}) + g(\mathbf{x}).$$

We begin by rewriting the problem as

$$\min_{\mathbf{x}, \mathbf{z} \in \mathbb{E}} \{ f(\mathbf{x}) + g(\mathbf{z}) : \mathbf{x} = \mathbf{z} \}$$

and then constructing the Lagrangian

$$L(\mathbf{x}, \mathbf{z}; \mathbf{y}) = f(\mathbf{x}) + g(\mathbf{z}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle = - \left[\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}) \right] - \left[\langle -\mathbf{y}, \mathbf{z} \rangle - g(\mathbf{z}) \right].$$

The dual objective function is computed by minimizing the Lagrangian w.r.t. the primal variables $\mathbf{x}, \mathbf{z}$:

$$q(\mathbf{y}) = \min_{\mathbf{x}, \mathbf{z}} L(\mathbf{x}, \mathbf{z}; \mathbf{y}) = -f^*(\mathbf{y}) - g^*(-\mathbf{y}).$$

We thus obtain the following dual problem, which is also called *Fenchel's dual*:

$$\text{(D)} \ \max_{\mathbf{y} \in \mathbb{E}^*} \{ -f^*(\mathbf{y}) - g^*(-\mathbf{y}) \}.$$

Fenchel's duality theorem, which we recall below, provides conditions under which strong duality holds for the pair of problems (P) and (D).

Theorem 4.15 (Fenchel's duality theorem [108, Theorem 31.1]). *Let* $f, g :$ $\mathbb{E} \to (-\infty, \infty]$ *be proper convex functions. If* $\mathrm{ri}(\mathrm{dom}(f)) \cap \mathrm{ri}(\mathrm{dom}(g)) \neq \emptyset$, *then*

$$\min_{\mathbf{x} \in \mathbb{E}} \{ f(\mathbf{x}) + g(\mathbf{x}) \} = \max_{\mathbf{y} \in \mathbb{E}^*} \{ -f^*(\mathbf{y}) - g^*(-\mathbf{y}) \},$$

and the maximum in the right-hand problem is attained whenever it is finite.

4.5   Infimal Convolution and Conjugacy

We will now show that in some sense the operations of addition and infimal convolution are dual to each other under the conjugacy operation. The first result holds under the very mild condition of properness of the functions.

Theorem 4.16 (conjugate of infimal convolution). *For two proper functions* $h_1, h_2 : \mathbb{E} \to (-\infty, \infty]$ *it holds that*

$$(h_1 \square h_2)^* = h_1^* + h_2^*.$$

Proof. For every $\mathbf{y} \in \mathbb{E}^*$ one has

$$
\begin{aligned}
(h_1 \square h_2)^*(\mathbf{y}) &= \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} \rangle - (h_1 \square h_2)(\mathbf{x})\} \\
&= \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} \rangle - \min_{\mathbf{u} \in \mathbb{E}} \{h_1(\mathbf{u}) + h_2(\mathbf{x} - \mathbf{u})\}\} \\
&= \max_{\mathbf{x} \in \mathbb{E}} \max_{\mathbf{u} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} \rangle - h_1(\mathbf{u}) - h_2(\mathbf{x} - \mathbf{u})\} \\
&= \max_{\mathbf{x} \in \mathbb{E}} \max_{\mathbf{u} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} - \mathbf{u} \rangle + \langle \mathbf{y}, \mathbf{u} \rangle - h_1(\mathbf{u}) - h_2(\mathbf{x} - \mathbf{u})\} \\
&= \max_{\mathbf{u} \in \mathbb{E}} \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} - \mathbf{u} \rangle + \langle \mathbf{y}, \mathbf{u} \rangle - h_1(\mathbf{u}) - h_2(\mathbf{x} - \mathbf{u})\} \\
&= \max_{\mathbf{u} \in \mathbb{E}} \{h_2^*(\mathbf{y}) + \langle \mathbf{y}, \mathbf{u} \rangle - h_1(\mathbf{u})\} \\
&= h_1^*(\mathbf{y}) + h_2^*(\mathbf{y}). \quad \square
\end{aligned}
$$

The second "direction" is a much deeper result requiring additional assumptions like convexity of the functions under consideration.

Theorem 4.17 (conjugate of sum). *Let $h_1 : \mathbb{E} \to (-\infty, \infty]$ be a proper convex function and $h_2 : \mathbb{E} \to \mathbb{R}$ be a real-valued convex function. Then*

$$
(h_1 + h_2)^* = h_1^* \square h_2^*.
$$

Proof. For any $\mathbf{y} \in \mathbb{E}^*$,

$$
\begin{aligned}
(h_1 + h_2)^*(\mathbf{y}) &= \max_{\mathbf{x} \in \mathbb{E}} \{\langle \mathbf{y}, \mathbf{x} \rangle - h_1(\mathbf{x}) - h_2(\mathbf{x})\} \\
&= -\min_{\mathbf{x} \in \mathbb{E}} \{h_1(\mathbf{x}) + h_2(\mathbf{x}) - \langle \mathbf{y}, \mathbf{x} \rangle\} \\
&= -\min_{\mathbf{x} \in \mathbb{E}} \{h_1(\mathbf{x}) + g(\mathbf{x})\}, \quad\quad (4.12)
\end{aligned}
$$

where $g(\mathbf{x}) \equiv h_2(\mathbf{x}) - \langle \mathbf{y}, \mathbf{x} \rangle$. Note that

$$
\mathrm{ri}(\mathrm{dom}(h_1)) \cap \mathrm{ri}(\mathrm{dom}(g)) = \mathrm{ri}(\mathrm{dom}(h_1)) \cap \mathbb{E} = \mathrm{ri}(\mathrm{dom}(h_1)) \neq \emptyset,
$$

and we can thus employ Fenchel's duality theorem (Theorem 4.15) and obtain the following equality:

$$
\min_{\mathbf{x} \in \mathbb{E}} \{h_1(\mathbf{x}) + g(\mathbf{x})\} = \max_{\mathbf{z} \in \mathbb{E}^*} \{-h_1^*(\mathbf{z}) - g^*(-\mathbf{z})\} = \max_{\mathbf{z} \in \mathbb{E}^*} \{-h_1^*(\mathbf{z}) - h_2^*(\mathbf{y} - \mathbf{z})\}.
$$
$$(4.13)$$

Combining (4.12) and (4.13), we finally obtain that for any $\mathbf{y} \in \mathbb{E}^*$,

$$
(h_1 + h_2)^*(\mathbf{y}) = \min_{\mathbf{z} \in \mathbb{E}^*} \{h_1^*(\mathbf{z}) + h_2^*(\mathbf{y} - \mathbf{z})\} = (h_1^* \square h_2^*)(\mathbf{y}),
$$

establishing the desired result. $\quad \square$

Corollary 4.18. *Let $h_1 : \mathbb{E} \to (-\infty, \infty]$ be a proper closed convex function and $h_2 : \mathbb{E} \to \mathbb{R}$ be a real-valued convex function. Then*

$$
h_1 + h_2 = (h_1^* \square h_2^*)^*.
$$

Proof. The function $h_1 + h_2$ is obviously proper and is closed by the closedness of h_1, h_2 (Theorem 2.7(b)). Therefore, by Theorem 4.8, $(h_1 + h_2)^{**} = h_1 + h_2$, which, combined with Theorem 4.17, yields

$$h_1 + h_2 = (h_1 + h_2)^{**} = [(h_1 + h_2)^*]^* = (h_1^* \square h_2^*)^*. \qquad \square$$

The next result shows a representation of the infimal convolution in terms of the corresponding conjugate functions.

Theorem 4.19 (representation of the infimal convolution by conjugates). *Let $h_1 : \mathbb{E} \to (-\infty, \infty]$ be a proper convex function, and let $h_2 : \mathbb{E} \to \mathbb{R}$ be a real-valued convex function. Suppose that $h_1 \square h_2$ is a real-valued function. Then*

$$h_1 \square h_2 = (h_1^* + h_2^*)^*. \tag{4.14}$$

Proof. By Theorem 4.16,

$$(h_1 \square h_2)^* = h_1^* + h_2^*. \tag{4.15}$$

Since h_1 is proper and convex and h_2 is real-valued and convex, it follows by Theorem 2.19 that $h_1 \square h_2$ is convex. Since $h_1 \square h_2$ is real-valued, it is in particular proper and closed. Therefore, by Theorem 4.8, $(h_1 \square h_2)^{**} = h_1 \square h_2$. Hence, taking the conjugate of both sides of (4.15), the identity (4.14) follows. $\square$

4.6   Subdifferentials of Conjugate Functions

The main result concerning the subdifferential of a conjugate function is the so-called *conjugate subgradient theorem*.

Theorem 4.20 (conjugate subgradient theorem). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be proper and convex. The following two claims are equivalent for any $\mathbf{x} \in \mathbb{E}, \mathbf{y} \in \mathbb{E}^*$:*

(i) $\langle \mathbf{x}, \mathbf{y} \rangle = f(\mathbf{x}) + f^*(\mathbf{y})$.

(ii) $\mathbf{y} \in \partial f(\mathbf{x})$.

If in addition f is closed, then (i) *and* (ii) *are equivalent to*

(iii) $\mathbf{x} \in \partial f^*(\mathbf{y})$.

Proof. The relation $\mathbf{y} \in \partial f(\mathbf{x})$ holds if and only if

$$f(\mathbf{z}) \geq f(\mathbf{x}) + \langle \mathbf{y}, \mathbf{z} - \mathbf{x} \rangle \text{ for all } \mathbf{z} \in \mathbb{E},$$

which is the same as

$$\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}) \geq \langle \mathbf{y}, \mathbf{z} \rangle - f(\mathbf{z}) \text{ for all } \mathbf{z} \in \mathbb{E}. \tag{4.16}$$

Taking the maximum over $\mathbf{z}$, we obtain that (4.16) is the same as

$$\langle \mathbf{y}, \mathbf{x} \rangle - f(\mathbf{x}) \geq f^*(\mathbf{y}),$$

which by Fenchel's inequality (Theorem 4.6) is equivalent to the equality $\langle \mathbf{x}, \mathbf{y} \rangle = f(\mathbf{x}) + f^*(\mathbf{y})$. We have thus established the equivalence between (i) and (ii). Assume now that in addition f is closed. Then by Theorem 4.8, $f^{**} = f$, which in particular implies that (i) is equivalent to

$$\langle \mathbf{x}, \mathbf{y} \rangle = g(\mathbf{y}) + g^*(\mathbf{x}),$$

where $g = f^*$. By the same equivalence that was already established between (i) and (ii) (but here employed on g), we conclude that (i) is equivalent to $\mathbf{x} \in \partial g(\mathbf{y}) = \partial f^*(\mathbf{y})$. $\square$

By the definition of the conjugate function, claim (i) in Theorem 4.20 can be rewritten as

$$\mathbf{x} \in \operatorname{argmax}_{\tilde{\mathbf{x}} \in \mathbb{E}} \left\{ \langle \mathbf{y}, \tilde{\mathbf{x}} \rangle - f(\tilde{\mathbf{x}}) \right\},$$

and, when f is closed, also as

$$\mathbf{y} \in \operatorname{argmax}_{\tilde{\mathbf{y}} \in \mathbb{E}^*} \left\{ \langle \mathbf{x}, \tilde{\mathbf{y}} \rangle - f^*(\tilde{\mathbf{y}}) \right\}.$$

Equipped with the above observation, we can conclude that the conjugate subgradient theorem, in the case where f is closed, can also be equivalently formulated as follows.

Corollary 4.21 (conjugate subgradient theorem—second formulation). *Let* $f : \mathbb{E} \to (-\infty, \infty]$ *be a proper closed convex function. Then for any* $\mathbf{x} \in \mathbb{E}, \mathbf{y} \in \mathbb{E}^*$,

$$\partial f(\mathbf{x}) = \operatorname{argmax}_{\tilde{\mathbf{y}} \in \mathbb{E}^*} \left\{ \langle \mathbf{x}, \tilde{\mathbf{y}} \rangle - f^*(\tilde{\mathbf{y}}) \right\}$$

and

$$\partial f^*(\mathbf{y}) = \operatorname{argmax}_{\tilde{\mathbf{x}} \in \mathbb{E}} \left\{ \langle \mathbf{y}, \tilde{\mathbf{x}} \rangle - f(\tilde{\mathbf{x}}) \right\}.$$

In particular, we can also conclude that for any proper closed convex function f,

$$\partial f(0) = \operatorname{argmin}_{\mathbf{y} \in \mathbb{E}^*} f^*(\mathbf{y})$$

and

$$\partial f^*(0) = \operatorname{argmin}_{\mathbf{x} \in \mathbb{E}} f(\mathbf{x}).$$

Example 4.22. Let $f : \mathbb{E} \to \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|$. Obviously, f is proper, closed, and convex. By Example 2.31, $f = \sigma_{B_{\|\cdot\|_*}[\mathbf{0},1]}$. Therefore, by Example 4.9, $f^* = \delta_{B_{\|\cdot\|_*}[\mathbf{0},1]}$. We can now use the conjugate subgradient theorem (Corollary 4.21) and compute the subdifferential set of f at $\mathbf{0}$ as follows:

$$\partial f(\mathbf{0}) = \operatorname{argmin}_{\mathbf{y} \in \mathbb{E}^*} f^*(\mathbf{y}) = \operatorname{argmin}_{\mathbf{y} \in \mathbb{E}^*} \delta_{B_{\|\cdot\|_*}[\mathbf{0},1]} = B_{\|\cdot\|_*}[\mathbf{0}, 1].$$

This result was already established in Example 3.3. ∎

A relation between Lipschitz continuity of a function and the boundedness of its subgradients over a given set was established in Theorem 3.61. We end this chapter with a related result showing that Lipschitz continuity over the entire space is also equivalent to boundedness of the domain of the conjugate.

Theorem 4.23 (Lipschitz continuity and boundedness of the domain of the conjugate). *Let $f : \mathbb{E} \to \mathbb{R}$ be convex. Then the following three claims are equivalent for a given constant $L > 0$:*

(i) $|f(\mathbf{x}) - f(\mathbf{y})| \leq L\|\mathbf{x} - \mathbf{y}\|$ *for any* $\mathbf{x}, \mathbf{y} \in \mathbb{E}$.

(ii) $\|\mathbf{g}\|_* \leq L$ *for any* $\mathbf{g} \in \partial f(\mathbf{x}), \mathbf{x} \in \mathbb{E}$.

(iii) $\mathrm{dom}(f^*) \subseteq B_{\|\cdot\|_*}[\mathbf{0}, L]$.

Proof. The equivalence between (i) and (ii) follows from Theorem 3.61. We will show that (iii) implies (ii). Indeed, assume that (iii) holds, that is, $\mathrm{dom}(f^*) \subseteq B_{\|\cdot\|_*}[\mathbf{0}, L]$. Since by the conjugate subgradient theorem (Corollary 4.21) for any $\mathbf{x} \in \mathbb{E}$,

$$\partial f(\mathbf{x}) = \mathrm{argmax}_{\mathbf{y} \in \mathbb{E}^*} \left\{ \langle \mathbf{x}, \mathbf{y} \rangle - f^*(\mathbf{y}) \right\},$$

it follows that $\partial f(\mathbf{x}) \subseteq \mathrm{dom}(f^*)$, and hence in particular $\partial f(\mathbf{x}) \subseteq B_{\|\cdot\|_*}[\mathbf{0}, L]$ for any $\mathbf{x} \in \mathbb{E}$, establishing (ii). In the reverse direction, we will show that the implication (i) $\Rightarrow$ (iii) holds. Suppose that (i) holds. Then in particular

$$f(\mathbf{x}) - f(\mathbf{0}) \leq |f(\mathbf{x}) - f(\mathbf{0})| \leq L\|\mathbf{x}\|,$$

and hence

$$-f(\mathbf{x}) \geq -f(\mathbf{0}) - L\|\mathbf{x}\|.$$

Therefore, for any $\mathbf{y} \in \mathbb{E}^*$,

$$f^*(\mathbf{y}) = \max_{\mathbf{x} \in \mathbb{E}} \left\{ \langle \mathbf{x}, \mathbf{y} \rangle - f(\mathbf{x}) \right\} \geq \max_{\mathbf{x} \in \mathbb{E}} \left\{ \langle \mathbf{x}, \mathbf{y} \rangle - f(\mathbf{0}) - L\|\mathbf{x}\| \right\}. \tag{4.17}$$

To show (iii), we take $\tilde{\mathbf{y}} \in \mathbb{E}^*$ that satisfies $\|\tilde{\mathbf{y}}\|_* > L$ and show that $\tilde{\mathbf{y}} \notin \mathrm{dom}(f^*)$. Take a vector $\mathbf{y}^\dagger \in \mathbb{E}$ satisfying $\|\mathbf{y}^\dagger\| = 1$ for which $\langle \tilde{\mathbf{y}}, \mathbf{y}^\dagger \rangle = \|\tilde{\mathbf{y}}\|_*$ (such a vector exists by the definition of the dual norm). Define $C = \{\alpha \mathbf{y}^\dagger : \alpha \geq 0\} \subseteq \mathbb{E}$. We can now continue (4.17) (with $\mathbf{y} = \tilde{\mathbf{y}}$) and write

$$
\begin{aligned}
f^*(\tilde{\mathbf{y}}) &\geq \max_{\mathbf{x} \in \mathbb{E}} \left\{ \langle \mathbf{x}, \tilde{\mathbf{y}} \rangle - f(\mathbf{0}) - L\|\mathbf{x}\| \right\} \\
&\geq \max_{\mathbf{x} \in C} \left\{ \langle \mathbf{x}, \tilde{\mathbf{y}} \rangle - f(\mathbf{0}) - L\|\mathbf{x}\| \right\} \\
&= \max_{\alpha \geq 0} \left\{ \langle \alpha \tilde{\mathbf{y}}, \mathbf{y}^\dagger \rangle - f(\mathbf{0}) - L\alpha\|\mathbf{y}^\dagger\| \right\} \\
&= \max_{\alpha \geq 0} \left\{ \alpha\|\tilde{\mathbf{y}}\|_* - f(\mathbf{0}) - L\alpha \right\} \\
&= \max_{\alpha \geq 0} \left\{ \alpha(\|\tilde{\mathbf{y}}\|_* - L) - f(\mathbf{0}) \right\} \\
&\overset{\|\tilde{\mathbf{y}}\|_* > L}{=} \infty,
\end{aligned}
$$

thus showing that $\tilde{\mathbf{y}} \notin \mathrm{dom}(f^*)$, establishing claim (iii). $\qquad \square$