Chapter 6

The Proximal Operator

> **Underlying Space:** In this chapter $\mathbb{E}$ is a Euclidean space, meaning a finite dimensional space endowed with an inner product $\langle \cdot, \cdot \rangle$ and the Euclidean norm $\|\cdot\| = \sqrt{\langle \cdot, \cdot \rangle}$.

This chapter is devoted to the study of the proximal mapping, which will be fundamental in many of the algorithms that will be explored later in the book. The operator and its properties were first studied by Moreau, and hence it is also referred to as "Moreau's proximal mapping."

6.1 Definition, Existence, and Uniqueness

Definition 6.1 (proximal mapping). *Given a function $f : \mathbb{E} \to (-\infty, \infty]$, the* **proximal mapping** *of f is the operator given by*

$$\operatorname{prox}_f(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\} \text{ for any } \mathbf{x} \in \mathbb{E}.$$

We will often use the term "prox" instead of "proximal." The mapping prox_f takes a vector $\mathbf{x} \in \mathbb{E}$ and maps it into a subset of $\mathbb{E}$, which might be empty, a singleton, or a set with multiple vectors as the following example illustrates.

Example 6.2. Consider the following three functions from $\mathbb{R}$ to $\mathbb{R}$:

$$g_1(x) \equiv 0,$$

$$g_2(x) = \begin{cases} 0, & x \neq 0, \\ -\lambda, & x = 0, \end{cases}$$

$$g_3(x) = \begin{cases} 0, & x \neq 0, \\ \lambda, & x = 0, \end{cases}$$
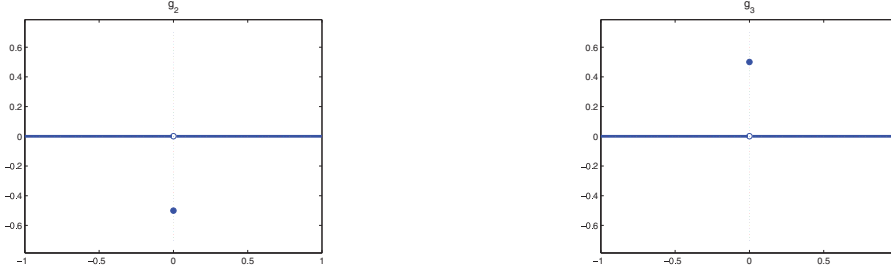
Figure 6.1. *The left and right images are the plots of the functions g_2 and g_3, respectively, with $\lambda = 0.5$ from Example* 6.2.

where $\lambda > 0$ is a given constant. The plots of the noncontinuous functions g_2 and g_3 are given in Figure 6.1. The prox of g_1 can computed as follows:

$$\mathrm{prox}_{g_1}(x) = \mathrm{argmin}_{u \in \mathbb{R}} \left\{ g_1(u) + \frac{1}{2}(u-x)^2 \right\} = \mathrm{argmin}_{u \in \mathbb{R}} \left\{ \frac{1}{2}(u-x)^2 \right\} = \{x\}.$$

To compute the prox of g_2, note that $\mathrm{prox}_{g_2}(x) = \mathrm{argmin}_{u \in \mathbb{R}} \tilde{g}_2(u, x)$, where

$$\tilde{g}_2(u,x) \equiv g_2(u) + \frac{1}{2}(u-x)^2 = \begin{cases} -\lambda + \frac{x^2}{2}, & u = 0, \\ \frac{1}{2}(u-x)^2, & u \neq 0. \end{cases}$$

For $x \neq 0$, the minimum of $\frac{1}{2}(u-x)^2$ over $\mathbb{R} \setminus \{0\}$ is attained at $u = x(\neq 0)$ with a minimal value of 0. Therefore, in this case, if $0 > -\lambda + \frac{x^2}{2}$, then the unique minimizer of $\tilde{g}_2(\cdot, x)$ is $u = 0$, and if $0 < -\lambda + \frac{x^2}{2}$, then $u = x$ is the unique minimizer of $\tilde{g}_2(\cdot, x)$; finally, if $0 = -\lambda + \frac{x^2}{2}$, then 0 and x are the two minimizers $\tilde{g}_2(\cdot, x)$. When $x = 0$, the minimizer of $\tilde{g}_2(\cdot, 0)$ is $u = 0$. To conclude,

$$\mathrm{prox}_{g_2}(x) = \begin{cases} \{0\}, & |x| < \sqrt{2\lambda}, \\ \{x\}, & |x| > \sqrt{2\lambda}, \\ \{0, x\}, & |x| = \sqrt{2\lambda}. \end{cases}$$

Similar arguments show that

$$\mathrm{prox}_{g_3}(x) = \begin{cases} \{x\}, & x \neq 0, \\ \emptyset, & x = 0. \end{cases} \quad \blacksquare$$

The next theorem, called the *first prox theorem*, states that if f is proper closed and convex, then $\mathrm{prox}_f(\mathbf{x})$ is always a singleton, meaning that the prox exists and is unique. This is the reason why in the last example only g_1, which was proper closed and convex, had a unique prox at any point.

Theorem 6.3 (first prox theorem). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed and convex function. Then $\operatorname{prox}_f(\mathbf{x})$ is a singleton for any $\mathbf{x} \in \mathbb{E}$.*

Proof. For any $\mathbf{x} \in \mathbb{E}$,

$$\operatorname{prox}_f(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u} \in \mathbb{E}} \tilde{f}(\mathbf{u}, \mathbf{x}), \tag{6.1}$$

where $\tilde{f}(\mathbf{u}, \mathbf{x}) \equiv f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2$. The function $\tilde{f}(\cdot, \mathbf{x})$ is a closed and strongly convex function as a sum of the closed and strongly convex function $\frac{1}{2}\| \cdot - \mathbf{x}\|^2$ and the closed and convex function f (see Lemma 5.20 and Theorem 2.7(b)). The properness of $\tilde{f}(\cdot, \mathbf{x})$ immediately follows from the properness of f. Therefore, by Theorem 5.25(a), there exists a unique minimizer to the problem in (6.1). $\quad\square$

When f is proper closed and convex, the last result shows that $\operatorname{prox}_f(\mathbf{x})$ is a singleton for any $\mathbf{x} \in \mathbb{E}$. In these cases, which will constitute the vast majority of cases that will be discussed in this chapter, we will treat prox_f as a single-valued mapping from $\mathbb{E}$ to $\mathbb{E}$, meaning that we will write $\operatorname{prox}_f(\mathbf{x}) = \mathbf{y}$ and not $\operatorname{prox}_f(\mathbf{x}) = \{\mathbf{y}\}$.

If we relax the assumptions in the first prox theorem and only require closedness of the function, then it is possible to show under some coerciveness assumptions that $\operatorname{prox}_f(\mathbf{x})$ is never an empty set.

Theorem 6.4 (nonemptiness of the prox under closedness and coerciveness). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed function, and assume that the following condition is satisfied:*

$$\text{the function } \mathbf{u} \mapsto f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \text{ is coercive for any } \mathbf{x} \in \mathbb{E}. \tag{6.2}$$

Then $\operatorname{prox}_f(\mathbf{x})$ is nonempty for any $\mathbf{x} \in \mathbb{E}$.

Proof. For any $\mathbf{x} \in \mathbb{E}$, the proper function $h(\mathbf{u}) \equiv f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2$ is closed as a sum of two closed functions. Since by the premise of the theorem it is also coercive, it follows by Theorem 2.14 (with $S = \mathbb{E}$) that $\operatorname{prox}_f(\mathbf{x})$, which consists of the minimizers of h, is nonempty. $\quad\square$

Example 6.2 actually gave an illustration of Theorem 6.4 since although both g_2 and g_3 satisfy the coercivity assumption (6.2), only g_2 was closed, and thus the fact that $\operatorname{prox}_{g_3}(x)$ was empty for a certain value of x, as opposed to $\operatorname{prox}_{g_2}(x)$, which was never empty, is not surprising.

6.2 First Set of Examples of Proximal Mappings

Equipped just with the definition of the proximal mapping, we will now compute the proximal mapping of several proper closed and convex functions.

6.2.1   Constant

If $f \equiv c$ for some $c \in \mathbb{R}$, then

$$\text{prox}_f(\mathbf{x}) = \text{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ c + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} = \mathbf{x}.$$

Therefore,

$$\text{prox}_f(\mathbf{x}) = \mathbf{x}$$

is the identity mapping.

6.2.2   Affine

Let $f(\mathbf{x}) = \langle \mathbf{a}, \mathbf{x} \rangle + b$, where $\mathbf{a} \in \mathbb{E}$ and $b \in \mathbb{R}$. Then

$$\begin{aligned}
\text{prox}_f(\mathbf{x}) &= \text{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ \langle \mathbf{a}, \mathbf{u} \rangle + b + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\} \\
&= \text{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ \langle \mathbf{a}, \mathbf{x} \rangle + b - \frac{1}{2} \|\mathbf{a}\|^2 + \frac{1}{2} \|\mathbf{u} - (\mathbf{x} - \mathbf{a})\|^2 \right\} \\
&= \mathbf{x} - \mathbf{a}.
\end{aligned}$$

Therefore,

$$\text{prox}_f(\mathbf{x}) = \mathbf{x} - \mathbf{a}$$

is a translation mapping.

6.2.3   Convex Quadratic

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 vector $\text{prox}_f(\mathbf{x})$ is the minimizer of the problem

$$\min_{\mathbf{u} \in \mathbb{E}} \left\{ \frac{1}{2} \mathbf{u}^T \mathbf{A} \mathbf{u} + \mathbf{b}^T \mathbf{u} + c + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\}.$$

The optimal solution of the last problem is attained when the gradient of the objective function vanishes:

$$\mathbf{A} \mathbf{u} + \mathbf{b} + \mathbf{u} - \mathbf{x} = \mathbf{0},$$

that is, when

$$(\mathbf{A} + \mathbf{I}) \mathbf{u} = \mathbf{x} - \mathbf{b},$$

and hence

$$\text{prox}_f(\mathbf{x}) = (\mathbf{A} + \mathbf{I})^{-1} (\mathbf{x} - \mathbf{b}).$$

6.2.4 One-Dimensional Examples

The following lemma contains several prox computations of one-dimensional functions.

Lemma 6.5. *The following are pairs of proper closed and convex functions and their prox mappings:*

$$g_1(x) = \begin{cases} \mu x, & x \geq 0, \\ \infty, & x < 0, \end{cases} \qquad \text{prox}_{g_1}(x) = [x - \mu]_+,$$

$$g_2(x) = \lambda|x|, \qquad \text{prox}_{g_2}(x) = [|x| - \lambda]_+ \text{sgn}(x),$$

$$g_3(x) = \begin{cases} \lambda x^3, & x \geq 0, \\ \infty, & x < 0, \end{cases} \qquad \text{prox}_{g_3}(x) = \frac{-1+\sqrt{1+12\lambda[x]_+}}{6\lambda},$$

$$g_4(x) = \begin{cases} -\lambda \log x, & x > 0, \\ \infty, & x \leq 0, \end{cases} \qquad \text{prox}_{g_4}(x) = \frac{x+\sqrt{x^2+4\lambda}}{2},$$

$$g_5(x) = \delta_{[0,\eta]\cap\mathbb{R}}(x), \qquad \text{prox}_{g_5}(x) = \min\{\max\{x,0\},\eta\},$$

where $\lambda \in \mathbb{R}_+, \eta \in [0,\infty]$, *and* $\mu \in \mathbb{R}$.

Proof. The proofs repeatedly use the following trivial arguments: (i) if $f'(u) = 0$ for a convex function f, then u must be one of its minimizers; (ii) if a minimizer of a convex function exists and is *not* attained at any point of differentiability, then it must be attained at a point of nondifferentiability.

[prox of g_1] By definition, $\text{prox}_{g_1}(x)$ is the minimizer of the function

$$f(u) = \begin{cases} \infty, & u < 0, \\ f_1(u), & u \geq 0, \end{cases}$$

where $f_1(u) = \mu u + \frac{1}{2}(u-x)^2$. First note that $f_1'(u) = 0$ if and only if $u = x - \mu$. If $x > \mu$, then $f'(x-\mu) = f_1'(x-\mu) = 0$, implying that in this case $\text{prox}_{g_1}(x) = x - \mu$. Otherwise, if $x \leq \mu$, the minimizer of f is not attained at a point of differentiability, meaning that it has to be attained at 0, which is the only point of nondifferentiability in the domain of f, so that $\text{prox}_{g_1}(x) = 0$.

[prox of g_2] $\text{prox}_{g_2}(x)$ is the minimizer of the function

$$h(u) = \begin{cases} h_1(u) \equiv \lambda u + \frac{1}{2}(u-x)^2, & u > 0, \\ h_2(u) \equiv -\lambda u + \frac{1}{2}(u-x)^2, & u \leq 0. \end{cases}$$

If the minimizer is attained at $u > 0$, then $0 = h_1'(u) = \lambda + u - x$, meaning that $u = x - \lambda$. Therefore, if $x > \lambda$, then $\text{prox}_{g_2}(x) = x - \lambda$. The same argument shows that if $x < -\lambda$, then $\text{prox}_{g_2}(x) = x + \lambda$. If $|x| \leq \lambda$, then $\text{prox}_{g_2}(x)$ must be the only point of nondifferentiability of h, namely, 0.

[prox of g_3] $\text{prox}_{g_3}(x)$ is the minimizer of the function

$$s(u) = \begin{cases} \lambda u^3 + \frac{1}{2}(u-x)^2, & u \geq 0, \\ \infty, & u < 0. \end{cases}$$

If the minimizer is positive, then $\tilde{u} = \text{prox}_{g_3}(x)$ satisfies $s'(\tilde{u}) = 0$, that is,

$$3\lambda\tilde{u}^2 + \tilde{u} - x = 0.$$

The above equation has a positive root if and only if $x > 0$, and in this case the (unique) positive root is $\text{prox}_{g_3}(x) = \tilde{u} = \frac{-1+\sqrt{1+12\lambda x}}{6\lambda}$. If $x \leq 0$, the minimizer of s is attained at the only point of nondifferentiability of s in its domain, that is, at 0.

[prox of g_4] $\tilde{u} = \text{prox}_{g_4}(x)$ is a minimizer over $\mathbb{R}_{++}$ of

$$t(u) = -\lambda \log u + \frac{1}{2}(u-x)^2,$$

which is determined by the condition that the derivative vanishes:

$$-\frac{\lambda}{\tilde{u}} + (\tilde{u} - x) = 0,$$

that is,

$$\tilde{u}^2 - \tilde{u}x - \lambda = 0.$$

Therefore (taking the positive root),

$$\text{prox}_{g_4}(x) = \tilde{u} = \frac{x + \sqrt{x^2 + 4\lambda}}{2}.$$

[prox of g_5] We will first assume that $\eta < \infty$. Note that $\tilde{u} = \text{prox}_{g_5}(x)$ is the minimizer of

$$w(u) = \frac{1}{2}(u-x)^2$$

over $[0, \eta]$. The minimizer of w over $\mathbb{R}$ is $u = x$. Therefore, if $0 \leq x \leq \eta$, then $\tilde{u} = x$. If $x < 0$, then w is increasing over $[0, \eta]$, and hence $\tilde{u} = 0$. Finally, if $x > \eta$, then w is decreasing over $[0, \eta]$, and thus $\tilde{u} = \eta$. To conclude,

$$\text{prox}_{g_5}(x) = \tilde{u} = \begin{cases} x, & 0 \leq x \leq \eta, \\ 0, & x < 0, \\ \eta, & x > \eta, \end{cases} = \min\{\max\{x,0\}, \eta\}.$$

For $\eta = \infty$, $g_5(x) = \delta_{[0,\infty)}(x)$, and in this case, g_5 is identical to g_1 with $\mu = 0$, implying that $\text{prox}_{g_5}(x) = [x]_+$, which can also be written as

$$\text{prox}_{g_5}(x) = \min\{\max\{x,0\}, \infty\}. \quad \square$$

6.3   Prox Calculus Rules

In this section we gather several important results on the calculus of proximal mappings. Note that some of the results do not require any convexity/closedness assumptions.

Theorem 6.6 (prox of separable functions). *Suppose that* $f : \mathbb{E}_1 \times \mathbb{E}_2 \times \cdots \times \mathbb{E}_m \to (-\infty, \infty]$ *is given by*

$$f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) = \sum_{i=1}^{m} f_i(\mathbf{x}_i) \text{ for any } \mathbf{x}_i \in \mathbb{E}_i, \quad i = 1, 2, \ldots, m.$$

Then for any $\mathbf{x}_1 \in \mathbb{E}_1, \mathbf{x}_2 \in \mathbb{E}_2, \ldots, \mathbf{x}_m \in \mathbb{E}_m,$

$$\text{prox}_f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) = \text{prox}_{f_1}(\mathbf{x}_1) \times \text{prox}_{f_2}(\mathbf{x}_2) \times \cdots \times \text{prox}_{f_m}(\mathbf{x}_m). \qquad (6.3)$$

Proof. Formula (6.3) is a result of the following chain of equalities:

$$\begin{aligned}
\text{prox}_f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) &= \text{argmin}_{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m} \sum_{i=1}^{m} \left[\frac{1}{2} \|\mathbf{y}_i - \mathbf{x}_i\|^2 + f_i(\mathbf{y}_i) \right] \\
&= \prod_{i=1}^{m} \text{argmin}_{\mathbf{y}_i} \left[\frac{1}{2} \|\mathbf{y}_i - \mathbf{x}_i\|^2 + f_i(\mathbf{y}_i) \right] \\
&= \prod_{i=1}^{m} \text{prox}_{f_i}(\mathbf{x}_i). \qquad \square
\end{aligned}$$

Remark 6.7. *If* $f : \mathbb{R}^n \to \mathbb{R}$ *is proper closed convex and separable,*

$$f(\mathbf{x}) = \sum_{i=1}^{n} f_i(x_i),$$

with f_i *being proper closed and convex univariate functions, then the result of Theorem* 6.6 *can be rewritten as*

$$\text{prox}_f(\mathbf{x}) = (\text{prox}_{f_i}(x_i))_{i=1}^{n}.$$

Example 6.8 (l_1-norm). Suppose that $g : \mathbb{R}^n \to \mathbb{R}$ is given by $g(\mathbf{x}) = \lambda \|\mathbf{x}\|_1$, where $\lambda > 0$. Then

$$g(\mathbf{x}) = \sum_{i=1}^{n} \varphi(x_i), \qquad (6.4)$$

where $\varphi(t) = \lambda|t|$. By Lemma 6.5 (computation of prox_{g_2}), $\text{prox}_\varphi(s) = \mathcal{T}_\lambda(s)$, where $\mathcal{T}_\lambda$ is defined as

$$\mathcal{T}_\lambda(y) = [|y| - \lambda]_+ \text{sgn}(y) = \begin{cases} y - \lambda, & y \geq \lambda, \\ 0, & |y| < \lambda, \\ y + \lambda, & y \leq -\lambda. \end{cases}$$
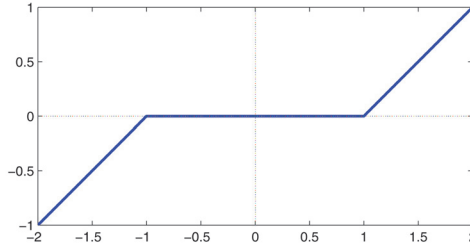
Figure 6.2. *The soft thresholding function $\mathcal{T}_1$.*

The function $\mathcal{T}_\lambda$ is called the *soft thresholding* function, and its description is given in Figure 6.2.

By Theorem 6.6,
$$\mathrm{prox}_g(\mathbf{x}) = (\mathcal{T}_\lambda(x_j))_{j=1}^n.$$

We will expand the definition of the soft thresholding function for vectors by applying it componentwise, that is, for any $\mathbf{x} \in \mathbb{R}^n$,
$$\mathcal{T}_\lambda(\mathbf{x}) \equiv (\mathcal{T}_\lambda(x_j))_{j=1}^n = [|\mathbf{x}| - \lambda\mathbf{e}]_+ \odot \mathrm{sgn}(\mathbf{x}).$$

In this notation,

$$\mathrm{prox}_g(\mathbf{x}) = \mathcal{T}_\lambda(\mathbf{x}).$$

∎

Example 6.9 (negative sum of logs). Let $g : \mathbb{R}^n \to (-\infty, \infty]$ be given by
$$g(\mathbf{x}) = \begin{cases} -\lambda \sum_{j=1}^n \log x_j, & \mathbf{x} > \mathbf{0}, \\ \infty & \text{else,} \end{cases}$$

where $\lambda > 0$. Then $g(\mathbf{x}) = \sum_{i=1}^n \varphi(x_i)$, where
$$\varphi(t) = \begin{cases} -\lambda \log t, & t > 0, \\ \infty, & t < 0. \end{cases}$$

By Lemma 6.5 (computation of prox_{g_4}),
$$\mathrm{prox}_\varphi(s) = \frac{s + \sqrt{s^2 + 4\lambda}}{2}.$$

Thus, by Theorem 6.6,

$$\text{prox}_g(\mathbf{x}) = (\text{prox}_\varphi(x_j))_{j=1}^n = \left(\frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2} \right)_{j=1}^n.$$

∎

Example 6.10 (l_0-norm). Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(\mathbf{x}) = \lambda \|\mathbf{x}\|_0$, where $\lambda > 0$ and $\|\mathbf{x}\|_0 = \#\{i : x_i \neq 0\}$ is the l_0-norm discussed in Example 2.11. For any $\mathbf{x} \in \mathbb{R}^n$,

$$f(\mathbf{x}) = \sum_{i=1}^n I(x_i),$$

where

$$I(t) = \begin{cases} \lambda, & t \neq 0, \\ 0, & t = 0. \end{cases}$$

Note that $I(\cdot) = J(\cdot) + \lambda$, where

$$J(t) = \begin{cases} 0, & t \neq 0, \\ -\lambda, & t = 0, \end{cases}$$

and that by Example 6.2,

$$\text{prox}_J(s) = \begin{cases} \{0\}, & |s| < \sqrt{2\lambda}, \\ \{s\}, & |s| > \sqrt{2\lambda}, \\ \{0, s\}, & |s| = \sqrt{2\lambda}. \end{cases} \tag{6.5}$$

We can write the above as $\text{prox}_J(s) = \mathcal{H}_{\sqrt{2\lambda}}(s)$, where $\mathcal{H}_\alpha$ is the so-called *hard thresholding* operator defined by

$$\mathcal{H}_\alpha(s) \equiv \begin{cases} \{0\}, & |s| < \alpha, \\ \{s\}, & |s| > \alpha, \\ \{0, s\}, & |s| = \alpha. \end{cases}$$

The operators prox_J and prox_I are the same since for any $s \in \mathbb{R}$,

$$\begin{aligned} \text{prox}_I(s) &= \text{argmin}_t \left\{ I(t) + \frac{1}{2}(t-s)^2 \right\} \\ &= \text{argmin}_t \left\{ J(t) + \lambda + \frac{1}{2}(t-s)^2 \right\} \\ &= \text{argmin}_t \left\{ J(t) + \frac{1}{2}(t-s)^2 \right\} \\ &= \text{prox}_J(s). \end{aligned}$$

Thus, invoking Theorem 6.6, it follows that[27]

$$\text{prox}_g(\mathbf{x}) = \mathcal{H}_{\sqrt{2\lambda}}(x_1) \times \mathcal{H}_{\sqrt{2\lambda}}(x_2) \times \cdots \times \mathcal{H}_{\sqrt{2\lambda}}(x_n).$$

∎

Theorem 6.11 (scaling and translation). *Let $g : \mathbb{E} \to (-\infty, \infty]$ be a proper function. Let $\lambda \neq 0$ and $\mathbf{a} \in \mathbb{E}$. Define $f(\mathbf{x}) = g(\lambda \mathbf{x} + \mathbf{a})$. Then*

$$\text{prox}_f(\mathbf{x}) = \frac{1}{\lambda}\left[\text{prox}_{\lambda^2 g}(\lambda \mathbf{x} + \mathbf{a}) - \mathbf{a}\right]. \tag{6.6}$$

Proof. By definition of the prox,

$$\text{prox}_f(\mathbf{x}) = \text{argmin}_{\mathbf{u}}\left\{f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2\right\}$$
$$= \text{argmin}_{\mathbf{u}}\left\{g(\lambda\mathbf{u} + \mathbf{a}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2\right\}. \tag{6.7}$$

Making the change of variables

$$\mathbf{z} = \lambda\mathbf{u} + \mathbf{a}, \tag{6.8}$$

the objective function in the minimization problem (6.7) becomes

$$g(\mathbf{z}) + \frac{1}{2}\left\|\frac{1}{\lambda}(\mathbf{z} - \mathbf{a}) - \mathbf{x}\right\|^2 = \frac{1}{\lambda^2}\left[\lambda^2 g(\mathbf{z}) + \frac{1}{2}\|\mathbf{z} - (\lambda\mathbf{x} + \mathbf{a})\|^2\right]. \tag{6.9}$$

The minimizer of (6.9) is $\mathbf{z} = \text{prox}_{\lambda^2 g}(\lambda\mathbf{x} + \mathbf{a})$, and hence by (6.8), it follows that (6.6) holds. ▢

Theorem 6.12 (prox of $\lambda g(\cdot/\lambda)$). *Let $g : \mathbb{E} \to (-\infty, \infty]$ be proper, and let $\lambda \neq 0$. Define $f(\mathbf{x}) = \lambda g(\mathbf{x}/\lambda)$. Then*

$$\text{prox}_f(\mathbf{x}) = \lambda\text{prox}_{g/\lambda}(\mathbf{x}/\lambda).$$

Proof. Note that

$$\text{prox}_f(\mathbf{x}) = \text{argmin}_{\mathbf{u}}\left\{f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2\right\} = \text{argmin}_{\mathbf{u}}\left\{\lambda g\left(\frac{\mathbf{u}}{\lambda}\right) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2\right\}.$$

---

[27]Actually, $\text{prox}_g(\mathbf{x})$ should be a subset of $\mathbb{R}^n$, meaning the space of n-length *column* vectors, but here we practice some abuse of notation and represent $\text{prox}_g(\mathbf{x})$ as a set of n-length *row* vectors.

Making the change of variables $\mathbf{z} = \frac{\mathbf{u}}{\lambda}$, we can continue to write

$$
\begin{aligned}
\mathrm{prox}_f(\mathbf{x}) &= \lambda \mathrm{argmin}_{\mathbf{z}} \left\{ \lambda g(\mathbf{z}) + \frac{1}{2}\|\lambda \mathbf{z} - \mathbf{x}\|^2 \right\} \\
&= \lambda \mathrm{argmin}_{\mathbf{z}} \left\{ \lambda^2 \left[\frac{g(\mathbf{z})}{\lambda} + \frac{1}{2} \left\| \mathbf{z} - \frac{\mathbf{x}}{\lambda} \right\|^2 \right] \right\} \\
&= \lambda \mathrm{argmin}_{\mathbf{z}} \left\{ \frac{g(\mathbf{z})}{\lambda} + \frac{1}{2} \left\| \mathbf{z} - \frac{\mathbf{x}}{\lambda} \right\|^2 \right\} \\
&= \lambda \mathrm{prox}_{g/\lambda}(\mathbf{x}/\lambda). \quad \square
\end{aligned}
$$

Theorem 6.13 (quadratic perturbation). *Let $g : \mathbb{E} \to (-\infty, \infty]$ be proper, and let $f(\mathbf{x}) = g(\mathbf{x}) + \frac{c}{2}\|\mathbf{x}\|^2 + \langle \mathbf{a}, \mathbf{x} \rangle + \gamma$, where $c > 0, \mathbf{a} \in \mathbb{E}$, and $\gamma \in \mathbb{R}$. Then*

$$
\mathrm{prox}_f(\mathbf{x}) = \mathrm{prox}_{\frac{1}{c+1}g} \left(\frac{\mathbf{x} - \mathbf{a}}{c+1} \right).
$$

Proof. Follows by the following simple computation:

$$
\begin{aligned}
\mathrm{prox}_f(\mathbf{x}) &= \mathrm{argmin}_{\mathbf{u}} \left\{ f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\} \\
&= \mathrm{argmin}_{\mathbf{u}} \left\{ g(\mathbf{u}) + \frac{c}{2}\|\mathbf{u}\|^2 + \langle \mathbf{a}, \mathbf{u} \rangle + \gamma + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\} \\
&= \mathrm{argmin}_{\mathbf{u}} \left\{ g(\mathbf{u}) + \frac{c+1}{2} \left\| \mathbf{u} - \left(\frac{\mathbf{x} - \mathbf{a}}{c+1} \right) \right\|^2 \right\} \\
&= \mathrm{prox}_{\frac{1}{c+1}g} \left(\frac{\mathbf{x} - \mathbf{a}}{c+1} \right). \quad \square
\end{aligned}
$$

Example 6.14. Consider the function $f : \mathbb{R} \to (-\infty, \infty]$ given for any $x \in \mathbb{R}$ by

$$
f(x) = \begin{cases} \mu x, & 0 \le x \le \alpha, \\ \infty & \text{else,} \end{cases}
$$

where $\mu \in \mathbb{R}$ and $\alpha \in [0, \infty]$. To compute the prox of f, note first that f can be represented as

$$
f(x) = \delta_{[0,\alpha] \cap \mathbb{R}}(x) + \mu x.
$$

By Lemma 6.5 (computation of prox_{g_5}), $\mathrm{prox}_{\delta_{[0,\alpha] \cap \mathbb{R}}}(x) = \min\{\max\{x, 0\}, \alpha\}$. Therefore, using Theorem 6.13 with $c = 0, \mathbf{a} = \mu, \gamma = 0$, we obtain that for any $x \in \mathbb{R}$,

$$
\boxed{\mathrm{prox}_f(x) = \mathrm{prox}_g(x - \mu) = \min\{\max\{x - \mu, 0\}, \alpha\}.}
$$

■

Unfortunately, there is no useful calculus rule for computing the prox mapping of a composition of a function with a general affine mapping. However, if the associated linear transformation satisfies a certain orthogonality condition, such a rule exists.

Theorem 6.15 (composition with an affine mapping). *Let* $g : \mathbb{R}^m \to (-\infty, \infty]$ *be a proper closed convex function, and let* $f(\mathbf{x}) = g(\mathcal{A}(\mathbf{x}) + \mathbf{b})$, *where* $\mathbf{b} \in \mathbb{R}^m$ *and* $\mathcal{A} : \mathbb{V} \to \mathbb{R}^m$ *is a linear transformation satisfying*[28] $\mathcal{A} \circ \mathcal{A}^T = \alpha \mathcal{I}$ *for some constant* $\alpha > 0$. *Then for any* $\mathbf{x} \in \mathbb{V}$,

$$\text{prox}_f(\mathbf{x}) = \mathbf{x} + \frac{1}{\alpha} \mathcal{A}^T (\text{prox}_{\alpha g}(\mathcal{A}(\mathbf{x}) + \mathbf{b}) - \mathcal{A}(\mathbf{x}) - \mathbf{b}).$$

Proof. By definition, $\text{prox}_f(\mathbf{x})$ is the optimal solution of

$$\min_{\mathbf{u} \in \mathbb{V}} \left\{ f(\mathbf{u}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\},$$

which can be rewritten as

$$\min_{\mathbf{u} \in \mathbb{V}} \left\{ g(\mathcal{A}(\mathbf{u}) + \mathbf{b}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \right\}.$$

The above problem can be formulated as the following constrained problem:

$$\begin{aligned} \min_{\mathbf{u} \in \mathbb{V}, \mathbf{z} \in \mathbb{R}^m} \quad & g(\mathbf{z}) + \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \\ \text{s.t.} \quad & \mathbf{z} = \mathcal{A}(\mathbf{u}) + \mathbf{b}. \end{aligned} \tag{6.10}$$

Denote the optimal solution of (6.10) by $(\tilde{\mathbf{z}}, \tilde{\mathbf{u}})$ (the existence and uniqueness of $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{u}}$ follow by the underlying assumption that g is proper closed and convex). Note that $\tilde{\mathbf{u}} = \text{prox}_f(\mathbf{x})$. Fixing $\mathbf{z} = \tilde{\mathbf{z}}$, we obtain that $\tilde{\mathbf{u}}$ is the optimal solution of

$$\begin{aligned} \min_{\mathbf{u} \in \mathbb{V}} \quad & \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 \\ \text{s.t.} \quad & \mathcal{A}(\mathbf{u}) = \tilde{\mathbf{z}} - \mathbf{b}. \end{aligned} \tag{6.11}$$

Since strong duality holds for problem (6.11) (see Theorem A.1), by Theorem A.2, it follows that there exists $\mathbf{y} \in \mathbb{R}^m$ for which

$$\tilde{\mathbf{u}} \in \text{argmin}_{\mathbf{u} \in \mathbb{V}} \left\{ \frac{1}{2} \|\mathbf{u} - \mathbf{x}\|^2 + \langle \mathbf{y}, \mathcal{A}(\mathbf{u}) - \tilde{\mathbf{z}} + \mathbf{b} \rangle \right\} \tag{6.12}$$

$$\mathcal{A}(\tilde{\mathbf{u}}) = \tilde{\mathbf{z}} - \mathbf{b}. \tag{6.13}$$

By (6.12),

$$\tilde{\mathbf{u}} = \mathbf{x} - \mathcal{A}^T(\mathbf{y}). \tag{6.14}$$

---

[28]The identity transformation $\mathcal{I}$ was defined in Section 1.10.

Substituting this expression of $\tilde{\mathbf{u}}$ into (6.13), we obtain

$$\mathcal{A}(\mathbf{x} - \mathcal{A}^T(\mathbf{y})) = \tilde{\mathbf{z}} - \mathbf{b},$$

and hence, using the assumption that $\mathcal{A} \circ \mathcal{A}^T = \alpha\mathcal{I}$,

$$\alpha\mathbf{y} = \mathcal{A}(\mathbf{x}) + \mathbf{b} - \tilde{\mathbf{z}},$$

which, combined with (6.14), yields an explicit expression for $\tilde{\mathbf{u}} = \text{prox}_f(\mathbf{x})$ in terms of $\tilde{\mathbf{z}}$:

$$\text{prox}_f(\mathbf{x}) = \tilde{\mathbf{u}} = \mathbf{x} + \frac{1}{\alpha}\mathcal{A}^T(\tilde{\mathbf{z}} - \mathcal{A}(\mathbf{x}) - \mathbf{b}). \tag{6.15}$$

Substituting $\mathbf{u} = \tilde{\mathbf{u}}$ in the minimization problem (6.10), we obtain that $\tilde{\mathbf{z}}$ is given by

$$\tilde{\mathbf{z}} = \text{argmin}_{\mathbf{z} \in \mathbb{R}^m} \left\{ g(\mathbf{z}) + \frac{1}{2} \left\| \mathbf{x} + \frac{1}{\alpha}\mathcal{A}^T(\mathbf{z} - \mathcal{A}(\mathbf{x}) - \mathbf{b}) - \mathbf{x} \right\|^2 \right\}$$

$$= \text{argmin}_{\mathbf{z} \in \mathbb{R}^m} \left\{ g(\mathbf{z}) + \frac{1}{2\alpha^2} \|\mathcal{A}^T(\mathbf{z} - \mathcal{A}(\mathbf{x}) - \mathbf{b})\|^2 \right\}$$

$$\overset{(*)}{=} \text{argmin}_{\mathbf{z} \in \mathbb{R}^m} \left\{ \alpha g(\mathbf{z}) + \frac{1}{2} \|\mathbf{z} - \mathcal{A}(\mathbf{x}) - \mathbf{b}\|^2 \right\}$$

$$= \text{prox}_{\alpha g}(\mathcal{A}(\mathbf{x}) + \mathbf{b}),$$

where the equality $(*)$ uses the assumption that $\mathcal{A} \circ \mathcal{A}^T = \alpha\mathcal{I}$. Plugging the expression for $\tilde{\mathbf{z}}$ into (6.15) produces the desired result. $\square$

Example 6.16. Let $g : \mathbb{E} \to (-\infty, \infty]$ be proper closed and convex where $\mathbb{E} = \mathbb{R}^d$, and let $f : \mathbb{E}^m \to (-\infty, \infty]$ be defined as

$$f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) = g(\mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_m).$$

The above can be written as $f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) = g(\mathcal{A}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m))$, where $\mathcal{A} : \mathbb{E}^m \to \mathbb{E}$ is the linear transformation

$$\mathcal{A}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) = \mathbf{x}_1 + \mathbf{x}_2 + \cdots + \mathbf{x}_m.$$

Obviously, the adjoint operator $\mathcal{A}^T : \mathbb{E} \to \mathbb{E}^m$ is given by

$$\mathcal{A}^T(\mathbf{x}) = (\mathbf{x}, \mathbf{x}, \ldots, \mathbf{x}),$$

and for any $\mathbf{x} \in \mathbb{E}$,

$$\mathcal{A}(\mathcal{A}^T(\mathbf{x})) = m\mathbf{x}.$$

Thus, the conditions of Theorem 6.15 are satisfied with $\alpha = m$ and $\mathbf{b} = 0$, and consequently, for any $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) \in \mathbb{E}^m$,

$$\text{prox}_f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)_j = \mathbf{x}_j + \frac{1}{m}\left(\text{prox}_{mg}\left(\sum_{i=1}^m \mathbf{x}_i\right) - \sum_{i=1}^m \mathbf{x}_i\right), \quad j = 1, 2, \ldots, m.$$

∎

Example 6.17. Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(\mathbf{x}) = |\mathbf{a}^T\mathbf{x}|$, where $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$. We can write f as $f(\mathbf{x}) = g(\mathbf{a}^T\mathbf{x})$, where $g(t) = |t|$. By Lemma 6.5 (prox_{g_2} computation), $\text{prox}_{\lambda g} = \mathcal{T}_\lambda$, with $\mathcal{T}_\lambda(x) = [|x| - \lambda]_+ \text{sgn}(x)$ being the soft thresholding operator defined in Example 6.8. Invoking Theorem 6.15 with $\alpha = \|\mathbf{a}\|^2$, $\mathbf{b} = 0$, and $\mathcal{A}$ defined as the transformation $\mathbf{x} \mapsto \mathbf{a}^T\mathbf{x}$, we obtain that

$$\text{prox}_f(\mathbf{x}) = \mathbf{x} + \frac{1}{\|\mathbf{a}\|^2}(\mathcal{T}_{\|\mathbf{a}\|^2}(\mathbf{a}^T\mathbf{x}) - \mathbf{a}^T\mathbf{x})\mathbf{a}.$$

∎

Theorem 6.18 (norm composition). *Let $f : \mathbb{E} \to \mathbb{R}$ be given by $f(\mathbf{x}) = g(\|\mathbf{x}\|)$, where $g : \mathbb{R} \to (-\infty, \infty]$ is a proper closed and convex function satisfying $\text{dom}(g) \subseteq [0, \infty)$. Then*

$$\text{prox}_f(\mathbf{x}) = \begin{cases} \text{prox}_g(\|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} \in \mathbb{E} : \|\mathbf{u}\| = \text{prox}_g(0)\}, & \mathbf{x} = \mathbf{0}. \end{cases} \quad (6.16)$$

Proof. By definition, $\text{prox}_f(\mathbf{0})$ is the set of minimizers of the problem

$$\min_{\mathbf{u} \in \mathbb{E}} \left\{ f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u}\|^2 \right\} = \min_{\mathbf{u} \in \mathbb{E}} \left\{ g(\|\mathbf{u}\|) + \frac{1}{2}\|\mathbf{u}\|^2 \right\}.$$

Making the change of variables $w = \|\mathbf{u}\|$, the problem reduces to (recalling that $\text{dom}(g) \subseteq [0, \infty)$)

$$\min_{w \in \mathbb{R}} \left\{ g(w) + \frac{1}{2}w^2 \right\}.$$

The optimal set of the above problem is $\text{prox}_g(0)$, and hence $\text{prox}_f(\mathbf{0})$ is the set of vectors $\mathbf{u}$ satisfying $\|\mathbf{u}\| = \text{prox}_g(0)$. We will now compute $\text{prox}_f(\mathbf{x})$ for $\mathbf{x} \neq \mathbf{0}$. The optimization problem associated with the prox computation can be rewritten as the following double minimization problem:

$$\min_{\mathbf{u} \in \mathbb{E}} \left\{ g(\|\mathbf{u}\|) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\} = \min_{\mathbf{u} \in \mathbb{E}} \left\{ g(\|\mathbf{u}\|) + \frac{1}{2}\|\mathbf{u}\|^2 - \langle \mathbf{u}, \mathbf{x} \rangle + \frac{1}{2}\|\mathbf{x}\|^2 \right\}$$
$$= \min_{\alpha \in \mathbb{R}_+} \min_{\mathbf{u} \in \mathbb{E}: \|\mathbf{u}\| = \alpha} \left\{ g(\alpha) + \frac{1}{2}\alpha^2 - \langle \mathbf{u}, \mathbf{x} \rangle + \frac{1}{2}\|\mathbf{x}\|^2 \right\}.$$

Using the Cauchy–Schwarz inequality, it is easy to see that the minimizer of the inner minimization problem is

$$\mathbf{u} = \alpha \frac{\mathbf{x}}{\|\mathbf{x}\|}, \tag{6.17}$$

and the corresponding optimal value is

$$g(\alpha) + \frac{1}{2}\alpha^2 - \alpha\|\mathbf{x}\| + \frac{1}{2}\|\mathbf{x}\|^2 = g(\alpha) + \frac{1}{2}(\alpha - \|\mathbf{x}\|)^2.$$

Therefore, $\mathrm{prox}_f(\mathbf{x})$ is given by $\mathbf{u}$ in (6.17) with α given by

$$
\begin{aligned}
\alpha &= \mathrm{argmin}_{\alpha \in \mathbb{R}_+} \left\{ g(\alpha) + \frac{1}{2}(\alpha - \|\mathbf{x}\|)^2 \right\} \\
&= \mathrm{argmin}_{\alpha \in \mathbb{R}} \left\{ g(\alpha) + \frac{1}{2}(\alpha - \|\mathbf{x}\|)^2 \right\} \\
&= \mathrm{prox}_g(\|\mathbf{x}\|),
\end{aligned}
$$

where the second equality is due to the assumption that $\mathrm{dom}(g) \subseteq [0, \infty)$. Thus, $\mathrm{prox}_f(\mathbf{x}) = \mathrm{prox}_g(\|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|}$. $\quad\square$

Example 6.19 (prox of Euclidean norm). Let $f : \mathbb{E} \to \mathbb{R}$ be given by $f(\mathbf{x}) = \lambda\|\mathbf{x}\|$, where $\lambda > 0$ and $\|\cdot\|$ is the underlying Euclidean norm (recall that in this section we assume that the underlying space is Euclidean). Then $f(\mathbf{x}) = g(\|\mathbf{x}\|)$, where

$$g(t) = \begin{cases} \lambda t, & t \geq 0, \\ \infty, & t < 0. \end{cases}$$

Then by Theorem 6.18, for any $\mathbf{x} \in \mathbb{E}$,

$$\mathrm{prox}_f(\mathbf{x}) = \begin{cases} \mathrm{prox}_g(\|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} \in \mathbb{E} : \|\mathbf{u}\| = \mathrm{prox}_g(0)\}, & \mathbf{x} = \mathbf{0}. \end{cases}$$

By Lemma 6.5 (computation of prox_{g_1}), $\mathrm{prox}_g(t) = [t - \lambda]_+$. Thus, $\mathrm{prox}_g(0) = 0$ and $\mathrm{prox}_g(\|\mathbf{x}\|) = [\|\mathbf{x}\| - \lambda]_+$, and therefore

$$\mathrm{prox}_f(\mathbf{x}) = \begin{cases} [\|\mathbf{x}\| - \lambda]_+\frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \mathbf{0}, & \mathbf{x} = \mathbf{0}. \end{cases}$$

Finally, we can write the above formula in the following compact form:

$$\mathrm{prox}_{\lambda\|\cdot\|}(\mathbf{x}) = \left(1 - \frac{\lambda}{\max\{\|\mathbf{x}\|, \lambda\}} \right) \mathbf{x}.$$

■

Example 6.20 (prox of cubic Euclidean norm). Let $f(\mathbf{x}) = \lambda\|\mathbf{x}\|^3$, where $\lambda > 0$. Then $f(\mathbf{x}) = \lambda g(\|\mathbf{x}\|)$, where

$$g(t) = \begin{cases} t^3, & t \geq 0, \\ \infty, & t < 0. \end{cases}$$

Then by Theorem 6.18, for any $\mathbf{x} \in \mathbb{R}$,

$$\mathrm{prox}_f(\mathbf{x}) = \begin{cases} \mathrm{prox}_g(\|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} \in \mathbb{E} : \|\mathbf{u}\| = \mathrm{prox}_g(0)\}, & \mathbf{x} = \mathbf{0}. \end{cases}$$

By Lemma 6.5 (computation of prox_{g_3}), $\mathrm{prox}_g(t) = \frac{-1+\sqrt{1+12\lambda[t]_+}}{6\lambda}$.   Therefore, $\mathrm{prox}_g(0) = 0$ and

$$\mathrm{prox}_f(\mathbf{x}) = \begin{cases} \frac{-1+\sqrt{1+12\lambda\|\mathbf{x}\|}}{6\lambda}\frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \mathbf{0}, & \mathbf{x} = \mathbf{0}, \end{cases}$$

and thus

$$\boxed{\mathrm{prox}_{\lambda\|\cdot\|^3}(\mathbf{x}) = \frac{2}{1 + \sqrt{1 + 12\lambda\|\mathbf{x}\|}}\mathbf{x}.}$$

■

Example 6.21 (prox of negative Euclidean norm). Let $f : \mathbb{E} \to \mathbb{R}$ be given by $f(\mathbf{x}) = -\lambda\|\mathbf{x}\|$, where $\lambda > 0$. Since f is not convex, we do not expect the prox to be a single-valued mapping. However, since f is closed, and since the function $\mathbf{u} \mapsto f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2$ is coercive for any $\mathbf{x} \in \mathbb{E}$, it follows by Theorem 6.4 that the set $\mathrm{prox}_f(\mathbf{x})$ is always nonempty. To compute the prox, note that $f(\mathbf{x}) = g(\|\mathbf{x}\|)$, where

$$g(t) = \begin{cases} -\lambda t, & t \geq 0, \\ \infty, & t < 0. \end{cases}$$

By Theorem 6.18, for any $\mathbf{x} \in \mathbb{R}$,

$$\mathrm{prox}_f(\mathbf{x}) = \begin{cases} \mathrm{prox}_g(\|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} \in \mathbb{E} : \|\mathbf{u}\| = \mathrm{prox}_g(0)\}, & \mathbf{x} = \mathbf{0}. \end{cases}$$

By Lemma 6.5 (computation of prox_{g_1}), $\mathrm{prox}_g(t) = [t+\lambda]_+$. Therefore, $\mathrm{prox}_g(0) = \lambda$ and

$$\text{prox}_{-\lambda\|\cdot\|}(\mathbf{x}) = \begin{cases} \left(1 + \frac{\lambda}{\|\mathbf{x}\|}\right)\mathbf{x}, & \mathbf{x} \neq \mathbf{0}, \\ \{\mathbf{u} : \|\mathbf{u}\| = \lambda\}, & \mathbf{x} = \mathbf{0}. \end{cases}$$

∎

Example 6.22 (prox of absolute value over symmetric intervals). Consider the function $f : \mathbb{R} \to (-\infty, \infty]$ given by

$$f(x) = \begin{cases} \lambda|x|, & |x| \leq \alpha, \\ \infty & \text{else}, \end{cases}$$

where $\lambda \in [0, \infty)$ and $\alpha \in [0, \infty]$. Then $f(x) = g(|x|)$, where

$$g(x) = \begin{cases} \lambda x, & 0 \leq x \leq \alpha, \\ \infty & \text{else}. \end{cases}$$

Thus, by Theorem 6.18, for any x,

$$\text{prox}_f(\mathbf{x}) = \begin{cases} \text{prox}_g(|x|)\frac{x}{|x|}, & x \neq 0, \\ \{u \in \mathbb{R} : |u| = \text{prox}_g(0)\}, & x = 0. \end{cases} \tag{6.18}$$

By Example 6.14, $\text{prox}_g(x) = \min\{\max\{x - \lambda, 0\}, \alpha\}$, which, combined with (6.18) and the fact that $\frac{x}{|x|} = \text{sgn}(x)$ for any $x \neq 0$, yields the formula

$$\text{prox}_{\lambda|\cdot| + \delta_{[-\alpha,\alpha]}}(x) = \min\{\max\{|x| - \lambda, 0\}, \alpha\}\text{sgn}(x).$$

∎

Using the previous example, we can compute the prox of weighted l_1-norms over boxes.

Example 6.23 (prox of weighted l_1 over a box). Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by

$$f(\mathbf{x}) = \begin{cases} \sum_{i=1}^n \omega_i|x_i|, & -\boldsymbol{\alpha} \leq \mathbf{x} \leq \boldsymbol{\alpha}, \\ \infty, & \text{else}, \end{cases}$$

for any $\mathbf{x} \in \mathbb{R}^n$, where $\boldsymbol{\omega} \in \mathbb{R}_+^n$ and $\boldsymbol{\alpha} \in [0, \infty]^n$. Then $f = \sum_{i=1}^n f_i$, where

$$f_i(x) = \begin{cases} w_i|x|, & -\alpha_i \leq x \leq \alpha_i, \\ \infty, & \text{else}. \end{cases}$$

Using Example 6.22 and invoking Theorem 6.6, we finally obtain that

$$\text{prox}_f(\mathbf{x}) = (\min\{\max\{|x_i| - \omega_i, 0\}, \alpha_i\}\text{sgn}(x_i))_{i=1}^n.$$

■

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

$f(\mathbf{x})$	$\text{prox}_f(\mathbf{x})$	Assumptions	Reference
$\sum_{i=1}^m f_i(\mathbf{x}_i)$	$\text{prox}_{f_1}(\mathbf{x}_1) \times \cdots \times \text{prox}_{f_m}(\mathbf{x}_m)$		Theorem 6.6
$g(\lambda\mathbf{x} + \mathbf{a})$	$\frac{1}{\lambda}\left[\text{prox}_{\lambda^2 g}(\lambda\mathbf{x} + \mathbf{a}) - \mathbf{a}\right]$	$\lambda \neq 0, \mathbf{a} \in \mathbb{E}, g$ proper	Theorem 6.11
$\lambda g(\mathbf{x}/\lambda)$	$\lambda\text{prox}_{g/\lambda}(\mathbf{x}/\lambda)$	$\lambda \neq 0, g$ proper	Theorem 6.12
$g(\mathbf{x}) + \frac{c}{2}\|\mathbf{x}\|^2 + \langle\mathbf{a}, \mathbf{x}\rangle + \gamma$	$\text{prox}_{\frac{1}{c+1}g}(\frac{\mathbf{x}-\mathbf{a}}{c+1})$	$\mathbf{a} \in \mathbb{E}, c > 0, \gamma \in \mathbb{R}, g$ proper	Theorem 6.13
$g(\mathcal{A}(\mathbf{x}) + \mathbf{b})$	$\mathbf{x} + \frac{1}{\alpha}\mathcal{A}^T(\text{prox}_{\alpha g}(\mathcal{A}(\mathbf{x}) + \mathbf{b}) - \mathcal{A}(\mathbf{x}) - \mathbf{b})$	$\mathbf{b} \in \mathbb{R}^m,$ $\mathcal{A} : \mathbb{V} \to \mathbb{R}^m,$ g proper closed convex, $\mathcal{A} \circ \mathcal{A}^T = \alpha I,$ $\alpha > 0$	Theorem 6.15
$g(\|\mathbf{x}\|)$	$\text{prox}_g(\|\mathbf{x}\|)\frac{\mathbf{x}}{\|\mathbf{x}\|}, \qquad \mathbf{x} \neq \mathbf{0}$ $\{\mathbf{u} : \|\mathbf{u}\| = \text{prox}_g(0)\}, \quad \mathbf{x} = \mathbf{0}$	g proper closed convex, $\text{dom}(g) \subseteq [0, \infty)$	Theorem 6.18

6.4 Prox of Indicators—Orthogonal Projections

6.4.1 The First Projection Theorem

Let $g : \mathbb{E} \to (-\infty, \infty]$ be given by $g(\mathbf{x}) = \delta_C(\mathbf{x})$, where C is a nonempty set. Then

$$\text{prox}_g(\mathbf{x}) = \text{argmin}_{\mathbf{u}\in\mathbb{E}}\left\{\delta_C(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2\right\} = \text{argmin}_{\mathbf{u}\in C}\|\mathbf{u} - \mathbf{x}\|^2 = P_C(\mathbf{x}).$$

Thus, the proximal mapping of the indicator function of a given set is the orthogonal projection[29] operator onto the same set.

Theorem 6.24. *Let $C \subseteq \mathbb{E}$ be nonempty. Then $\text{prox}_{\delta_C}(\mathbf{x}) = P_C(\mathbf{x})$ for any $\mathbf{x} \in \mathbb{E}$.*

If C is closed and convex, in addition to being nonempty, the indicator function δ_C is proper closed and convex, and hence by the first prox theorem (Theorem 6.3), the orthogonal projection mapping (which coincides with the proximal mapping) exists and is unique. This is the first projection theorem.

---

[29]The orthogonal projection operator was introduced in Example 3.31.

Theorem 6.25 (first projection theorem). *Let $C \subseteq \mathbb{E}$ be a nonempty closed convex set. Then $P_C(\mathbf{x})$ is a singleton for any $\mathbf{x} \in \mathbb{E}$.*

6.4.2 First Examples in $\mathbb{R}^n$

We begin by recalling[30] several known expressions for the orthogonal projection onto some basic subsets of $\mathbb{R}^n$. Since the assumption made throughout the book is that (unless otherwise stated) $\mathbb{R}^n$ is endowed with the dot product, and since the standing assumption in this chapter is that the underlying space is Euclidean, it follows that the endowed norm is the l_2-norm.

Lemma 6.26 (projection onto subsets of $\mathbb{R}^n$). *Following are pairs of nonempty closed and convex sets and their corresponding orthogonal projections:*

nonnegative orthant	$C_1 = \mathbb{R}^n_+,$	$[\mathbf{x}]_+,$
box	$C_2 = \mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}],$	$(\min\{\max\{x_i, \ell_i\}, u_i\})_{i=1}^n,$
affine set	$C_3 = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\},$	$\mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{x} - \mathbf{b}),$
l_2 ball	$C_4 = B_{\|\cdot\|_2}[\mathbf{c}, r],$	$\mathbf{c} + \frac{r}{\max\{\|\mathbf{x}-\mathbf{c}\|_2, r\}}(\mathbf{x} - \mathbf{c}),$
half-space	$C_5 = \{\mathbf{x} : \mathbf{a}^T\mathbf{x} \leq \alpha\},$	$\mathbf{x} - \frac{[\mathbf{a}^T\mathbf{x}-\alpha]_+}{\|\mathbf{a}\|^2}\mathbf{a},$

where $\boldsymbol{\ell} \in [-\infty, \infty)^n, \mathbf{u} \in (-\infty, \infty]^n$ are such that $\boldsymbol{\ell} \leq \mathbf{u}$, $\mathbf{A} \in \mathbb{R}^{m \times n}$ has full row rank, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{c} \in \mathbb{R}^n$, $r > 0$, $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, and $\alpha \in \mathbb{R}$.

Note that we extended the definition of box sets given in Section 1.7.1 to include unbounded intervals, meaning that $\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]$ is also defined when the components of $\boldsymbol{\ell}$ might also take the value $-\infty$, and the components of $\mathbf{u}$ might take the value ∞. However, boxes are always subsets of $\mathbb{R}^n$, and the formula

$$\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}] = \{\mathbf{x} \in \mathbb{R}^n : \boldsymbol{\ell} \leq \mathbf{x} \leq \mathbf{u}\}$$

still holds. For example, $\mathrm{Box}[\mathbf{0}, \infty\mathbf{e}] = \mathbb{R}^n_+$.

6.4.3 Projection onto the Intersection of a Hyperplane and a Box

The next result develops an expression for the orthogonal projection onto another subset of $\mathbb{R}^n$—the intersection of an hyperplane and a box.

Theorem 6.27 (projection onto the intersection of a hyperplane and a box). *Let $C \subseteq \mathbb{R}^n$ be given by*

$$C = H_{\mathbf{a},b} \cap \mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}] = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T\mathbf{x} = b, \boldsymbol{\ell} \leq \mathbf{x} \leq \mathbf{u}\},$$

where $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, b \in \mathbb{R}, \boldsymbol{\ell} \in [-\infty, \infty)^n, \mathbf{u} \in (-\infty, \infty]^n$. Assume that $C \neq \emptyset$. Then

$$P_C(\mathbf{x}) = P_{\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x} - \mu^*\mathbf{a}),$$

---

[30]The derivations of the orthogonal projection expressions in Lemma 6.26 can be found, for example, in [10].

where $\text{Box}[\boldsymbol{\ell}, \mathbf{u}] = \{\mathbf{y} \in \mathbb{R}^n : \ell_i \leq y_i \leq u_i, i = 1, 2, \ldots, n\}$ *and* μ^* *is a solution of the equation*

$$\mathbf{a}^T P_{\text{Box}[\boldsymbol{\ell},\mathbf{u}]}(\mathbf{x} - \mu\mathbf{a}) = b. \tag{6.19}$$

Proof. The orthogonal projection of $\mathbf{x}$ onto C is the unique optimal solution of

$$\min_{\mathbf{y}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 : \mathbf{a}^T\mathbf{y} = b, \boldsymbol{\ell} \leq \mathbf{y} \leq \mathbf{u} \right\}. \tag{6.20}$$

A Lagrangian of the problem is

$$L(\mathbf{y}; \mu) = \frac{1}{2}\|\mathbf{y}-\mathbf{x}\|_2^2 + \mu(\mathbf{a}^T\mathbf{y}-b) = \frac{1}{2}\|\mathbf{y}-(\mathbf{x}-\mu\mathbf{a})\|_2^2 - \frac{\mu^2}{2}\|\mathbf{a}\|_2^2 + \mu(\mathbf{a}^T\mathbf{x}-b). \tag{6.21}$$

Since strong duality holds for problem (6.20) (see Theorem A.1), it follows by Theorem A.2 that $\mathbf{y}^*$ is an optimal solution of problem (6.20) if and only if there exists $\mu^* \in \mathbb{R}$ (which will actually be an optimal solution of the dual problem) for which

$$\mathbf{y}^* \in \operatorname{argmin}_{\boldsymbol{\ell}\leq\mathbf{y}\leq\mathbf{u}}L(\mathbf{y}; \mu^*), \tag{6.22}$$
$$\mathbf{a}^T\mathbf{y}^* = b. \tag{6.23}$$

Using the expression of the Lagrangian given in (6.21), the relation (6.22) can be equivalently written as

$$\mathbf{y}^* = P_{\text{Box}[\boldsymbol{\ell},\mathbf{u}]}(\mathbf{x} - \mu^*\mathbf{a}).$$

The feasibility condition (6.23) can then be rewritten as

$$\mathbf{a}^T P_{\text{Box}[\boldsymbol{\ell},\mathbf{u}]}(\mathbf{x} - \mu^*\mathbf{a}) = b. \quad \square$$

Remark 6.28. *The projection onto the box* $\text{Box}[\boldsymbol{\ell}, \mathbf{u}]$ *is extremely simple and is done component-wise as described in Lemma* 6.26. *Note also that* (6.19) *actually consists in finding a root of the nonincreasing function* $\varphi(\mu) = \mathbf{a}^T P_{\text{Box}}(\mathbf{x} - \mu\mathbf{a}) - b$, *which is a task that can be performed efficiently even by simple procedures such as bisection. The fact that* φ *is nonincreasing follows from the observation that* $\varphi(\mu) = \sum_{i=1}^n a_i \min\{\max\{x_i - \mu a_i, \ell_i\}, u_i\} - b$ *and the fact that* $\mu \mapsto a_i \min\{\max\{x_i - \mu a_i, \ell_i\}, u_i\}$ *is a nonincreasing function for any* i.

A direct consequence of Theorem 6.27 is an expression for the orthogonal projection onto the unit simplex.

Corollary 6.29 (orthogonal projection onto the unit simplex). *For any* $\mathbf{x} \in \mathbb{R}^n$,

$$P_{\Delta_n}(\mathbf{x}) = [\mathbf{x} - \mu^*\mathbf{e}]_+,$$

where μ^* *is a root of the equation*

$$\mathbf{e}^T[\mathbf{x} - \mu^*\mathbf{e}]_+ - 1 = 0.$$

Proof. Invoking Theorem 6.27 with $\mathbf{a} = \mathbf{e}$, $b = 1$, $\ell_i = 0, u_i = \infty, i = 1, 2, \ldots, n$, and noting that in this case $P_{\mathrm{Box}[\boldsymbol{\ell},\mathbf{u}]}(\mathbf{x}) = [\mathbf{x}]_+$, the result follows. $\square$

In order to expend the variety of sets on which we will be able to find simple expressions for the orthogonal projection mapping, in the next two subsections, we will discuss how to project onto level sets and epigraphs.

6.4.4 Projection onto Level Sets

Theorem 6.30 (orthogonal projection onto level sets). *Let $C = Lev(f, \alpha) = \{\mathbf{x} \in \mathbb{E} : f(\mathbf{x}) \leq \alpha\}$, where $f : \mathbb{E} \to (-\infty, \infty]$ is proper closed and convex, and $\alpha \in \mathbb{R}$. Assume that there exists $\hat{\mathbf{x}} \in \mathbb{E}$ for which $f(\hat{\mathbf{x}}) < \alpha$. Then*

$$P_C(\mathbf{x}) = \begin{cases} P_{\mathrm{dom}(f)}(\mathbf{x}), & f(P_{\mathrm{dom}(f)}(\mathbf{x})) \leq \alpha, \\ \mathrm{prox}_{\lambda^* f}(\mathbf{x}) & else, \end{cases} \tag{6.24}$$

where λ^* is any positive root of the equation

$$\varphi(\lambda) \equiv f(\mathrm{prox}_{\lambda f}(\mathbf{x})) - \alpha = 0.$$

In addition, the function φ is nonincreasing.

Proof. The orthogonal projection of $\mathbf{x}$ onto C is an optimal solution of the problem

$$\min_{\mathbf{y} \in \mathbb{E}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 : f(\mathbf{y}) \leq \alpha, \mathbf{y} \in X \right\},$$

where $X = \mathrm{dom}(f)$. A Lagrangian of the problem is ($\lambda \geq 0$)

$$L(\mathbf{y}; \lambda) = \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 + \lambda f(\mathbf{y}) - \alpha \lambda. \tag{6.25}$$

Since the problem is convex and satisfies Slater's condition, strong duality holds (see Theorem A.1), and therefore it follows by the optimality conditions in Theorem A.2 that $\mathbf{y}^*$ is an optimal solution of problem (6.25) if and only if there exists $\lambda^* \in \mathbb{R}_+$ for which

$$\mathbf{y}^* \in \mathrm{argmin}_{\mathbf{y} \in X} L(\mathbf{y}; \lambda^*), \tag{6.26}$$

$$f(\mathbf{y}^*) \leq \alpha, \tag{6.27}$$

$$\lambda^*(f(\mathbf{y}^*) - \alpha) = 0. \tag{6.28}$$

There are two cases. If $P_X(\mathbf{x})$ exists and $f(P_X(\mathbf{x})) \leq \alpha$, then $\mathbf{y}^* = P_X(\mathbf{x})$, and $\lambda^* = 0$ is a solution to the system (6.26), (6.27), (6.28). Otherwise, if $P_X(\mathbf{x})$ does not exist or $f(P_X(\mathbf{x})) > \alpha$, then $\lambda^* > 0$, and in this case the system (6.26), (6.27), (6.28) reduces to $\mathbf{y}^* = \mathrm{prox}_{\lambda^* f}(\mathbf{x})$ and $f(\mathrm{prox}_{\lambda^* f}(\mathbf{x})) = \alpha$, which yields the formula (6.24).

To prove that φ is nonincreasing, recall that

$$\mathrm{prox}_{\lambda f}(\mathbf{x}) = \mathrm{argmin}_{\mathbf{y} \in X} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|^2 + \lambda(f(\mathbf{y}) - \alpha) \right\}.$$

Take $0 \leq \lambda_1 < \lambda_2$. Then denoting $\mathbf{v}_1 = \mathrm{prox}_{\lambda_1 f}(\mathbf{x})$ and $\mathbf{v}_2 = \mathrm{prox}_{\lambda_2 f}(\mathbf{x})$, we have

$$
\frac{1}{2}\|\mathbf{v}_2 - \mathbf{x}\|^2 + \lambda_2(f(\mathbf{v}_2) - \alpha)
$$

$$
= \frac{1}{2}\|\mathbf{v}_2 - \mathbf{x}\|^2 + \lambda_1(f(\mathbf{v}_2) - \alpha) + (\lambda_2 - \lambda_1)(f(\mathbf{v}_2) - \alpha)
$$

$$
\geq \frac{1}{2}\|\mathbf{v}_1 - \mathbf{x}\|^2 + \lambda_1(f(\mathbf{v}_1) - \alpha) + (\lambda_2 - \lambda_1)(f(\mathbf{v}_2) - \alpha)
$$

$$
= \frac{1}{2}\|\mathbf{v}_1 - \mathbf{x}\|^2 + \lambda_2(f(\mathbf{v}_1) - \alpha) + (\lambda_2 - \lambda_1)(f(\mathbf{v}_2) - f(\mathbf{v}_1))
$$

$$
\geq \frac{1}{2}\|\mathbf{v}_2 - \mathbf{x}\|^2 + \lambda_2(f(\mathbf{v}_2) - \alpha) + (\lambda_2 - \lambda_1)(f(\mathbf{v}_2) - f(\mathbf{v}_1)).
$$

Therefore, $(\lambda_2 - \lambda_1)(f(\mathbf{v}_2) - f(\mathbf{v}_1)) \leq 0$. Since $\lambda_1 < \lambda_2$, we can conclude that $f(\mathbf{v}_2) \leq f(\mathbf{v}_1)$. Finally,

$$
\varphi(\lambda_2) = f(\mathbf{v}_2) - \alpha \leq f(\mathbf{v}_1) - \alpha = \varphi(\lambda_1),
$$

establishing the monotonicity of φ. $\square$

Remark 6.31. *Note that in Theorem* 6.30 *f is assumed to be closed, but this does not necessarily imply that* $\mathrm{dom}(f)$ *is closed. In cases where* $\mathrm{dom}(f)$ *is not closed, it might happen that* $P_{\mathrm{dom}(f)}(\mathbf{x})$ *does not exist and formula* (6.24) *amounts to* $P_C(\mathbf{x}) = \mathrm{prox}_{\lambda^* f}(\mathbf{x})$.

Example 6.32 (projection onto the intersection of a half-space and a box).
Consider the set

$$
C = H_{\mathbf{a},b}^- \cap \mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}] = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}^T \mathbf{x} \leq b, \boldsymbol{\ell} \leq \mathbf{x} \leq \mathbf{u}\},
$$

where $\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}, b \in \mathbb{R}, \boldsymbol{\ell} \in [-\infty, \infty)^n$ and $\mathbf{u} \in (-\infty, \infty]^n$. Assume that $C \neq \emptyset$. Then $C = \mathrm{Lev}(f, b)$, where $f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} + \delta_{\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x})$. For any $\lambda > 0$,

$$
\mathrm{prox}_{\lambda f}(\mathbf{x}) = \mathrm{prox}_{\lambda \mathbf{a}^T(\cdot) + \delta_{\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\cdot)}(\mathbf{x}) \stackrel{(*)}{=} \mathrm{prox}_{\delta_{\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]}}(\mathbf{x} - \lambda \mathbf{a}) = P_{\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x} - \lambda \mathbf{a}),
$$

where in the equality $(*)$ we used Theorem 6.13. Invoking Theorem 6.30, we obtain the following formula for the projection on C:

$$
P_C(\mathbf{x}) = \begin{cases} P_{\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x}), & \mathbf{a}^T P_{\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x}) \leq b, \\ P_{\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x} - \lambda^* \mathbf{a}), & \mathbf{a}^T P_{\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x}) > b, \end{cases}
$$

where λ^* is any positive root of the nonincreasing function

$$
\varphi(\lambda) = \mathbf{a}^T P_{\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]}(\mathbf{x} - \lambda \mathbf{a}) - b.
$$

■

Example 6.33 (projection onto the l_1 ball). Let $C = B_{\|\cdot\|_1}[\mathbf{0}, \alpha] = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_1 \leq \alpha\}$, where $\alpha > 0$. Then $C = \text{Lev}(f, \alpha)$ with $f(\mathbf{x}) = \|\mathbf{x}\|_1$. The prox of $\lambda f = \lambda \|\cdot\|_1$ for any $\lambda > 0$ was computed in Example 6.8, where it was shown that

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathcal{T}_\lambda(\mathbf{x}) \text{ for all } \mathbf{x} \in \mathbb{R}^n$$

with $\mathcal{T}_\lambda$ being the soft thresholding operator given by $\mathcal{T}_\lambda(\mathbf{x}) = [\mathbf{x} - \lambda\mathbf{e}]_+ \odot \text{sgn}(\mathbf{x})$. Invoking Theorem 6.30, we obtain that

$$P_{B_{\|\cdot\|_1}[\mathbf{0}, \alpha]}(\mathbf{x}) = \begin{cases} \mathbf{x}, & \|\mathbf{x}\|_1 \leq \alpha, \\ \mathcal{T}_{\lambda^*}(\mathbf{x}), & \|\mathbf{x}\|_1 > \alpha, \end{cases}$$

where λ^* is any positive root of the nonincreasing function

$$\varphi(\lambda) = \|\mathcal{T}_\lambda(\mathbf{x})\|_1 - \alpha.$$

∎

The next example uses a generalization of the soft thresholding mapping, which will be called the *two-sided soft thresholding* operator, and is defined for any $\mathbf{a}, \mathbf{b} \in (-\infty, \infty]^n$ as

$$\mathcal{S}_{\mathbf{a}, \mathbf{b}}(\mathbf{x}) = (\min\{\max\{|x_i| - a_i, 0\}, b_i\}\text{sgn}(x_i))_{i=1}^n, \quad \mathbf{x} \in \mathbb{R}^n.$$

Obviously,

$$\mathcal{S}_{\lambda\mathbf{e}, \infty\mathbf{e}} = \mathcal{T}_\lambda.$$

Here $\infty\mathbf{e}$ is the n-dimensional column vector whose elements are all ∞. A plot of the function $t \mapsto \mathcal{S}_{1,2}(t)$ is given in Figure 6.3.

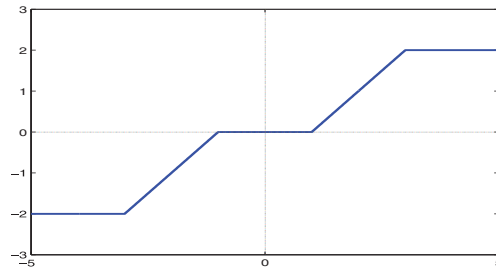**Figure 6.3.** *The two-sided soft thresholding function* $t \mapsto \mathcal{S}_{1,2}(t) = \min\{\max\{|t| - 1, 0\}, 2\}\text{sgn}(t)$.

Example 6.34 (projection onto the intersection of weighted l_1 ball and a box). Let $C \subseteq \mathbb{R}^n$ be given by

$$C = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n \omega_i |x_i| \leq \beta, -\boldsymbol{\alpha} \leq \mathbf{x} \leq \boldsymbol{\alpha} \right\},$$

where $\boldsymbol{\omega} \in \mathbb{R}^n_+$, $\boldsymbol{\alpha} \in [0, \infty]^n$, and $\beta \in \mathbb{R}_{++}$. Then obviously $C = \text{Lev}(f, \beta)$, where

$$f(\mathbf{x}) = \boldsymbol{\omega}^T |\mathbf{x}| + \delta_{\text{Box}[-\boldsymbol{\alpha},\boldsymbol{\alpha}]}(\mathbf{x}) = \begin{cases} \sum_{i=1}^n \omega_i |x_i|, & -\boldsymbol{\alpha} \le \mathbf{x} \le \boldsymbol{\alpha}, \\ \infty & \text{else} \end{cases}$$

for any $\mathbf{x} \in \mathbb{R}^n$. By Example 6.23, for any $\lambda > 0$ and $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = (\min\{\max\{|x_i| - \lambda\omega_i, 0\}, \alpha_i\}\text{sgn}(x_i))_{i=1}^n = \mathcal{S}_{\lambda\boldsymbol{\omega},\boldsymbol{\alpha}}(\mathbf{x}).$$

Therefore, invoking Theorem 6.30, we obtain that

$$P_C(\mathbf{x}) = \begin{cases} P_{\text{Box}[-\boldsymbol{\alpha},\boldsymbol{\alpha}]}(\mathbf{x}), & \boldsymbol{\omega}^T |P_{\text{Box}[-\boldsymbol{\alpha},\boldsymbol{\alpha}]}(\mathbf{x})| \le \beta, \\ \mathcal{S}_{\lambda^*\boldsymbol{\omega},\boldsymbol{\alpha}}(\mathbf{x}), & \boldsymbol{\omega}^T |P_{\text{Box}[-\boldsymbol{\alpha},\boldsymbol{\alpha}]}(\mathbf{x})| > \beta, \end{cases}$$

where λ^* is any positive root of the nonincreasing function

$$\varphi(\lambda) = \boldsymbol{\omega}^T |\mathcal{S}_{\lambda\boldsymbol{\omega},\boldsymbol{\alpha}}(\mathbf{x})| - \beta.$$

∎

As a final illustration of Theorem 6.30, we give an example in which the domain of f is not closed.

Example 6.35. Let

$$C = \left\{ \mathbf{x} \in \mathbb{R}^n_{++} : \Pi_{i=1}^n x_i \ge \alpha \right\},$$

where $\alpha > 0$. The key property that will allow us to compute the orthogonal projection onto C is the fact that it can be rewritten as

$$C = \left\{ \mathbf{x} \in \mathbb{R}^n_{++} : -\sum_{i=1}^n \log x_i \le -\log \alpha \right\}.$$

Thus, $C = \text{Lev}(f, -\log \alpha)$, where $f : \mathbb{R}^n \to (-\infty, \infty]$ is the negative sum of logs function:

$$f(\mathbf{x}) = \begin{cases} -\sum_{i=1}^n \log x_i, & \mathbf{x} \in \mathbb{R}^n_{++}, \\ \infty & \text{else.} \end{cases}$$

In Example 6.9 it was shown that for any $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \left(\frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2} \right)_{j=1}^n.$$

We can now invoke Theorem 6.30 to obtain a formula (up to a single parameter that can be found by a one-dimensional search) for the projection onto C, but there is one issue that needs to be treated delicately. If $\mathbf{x} \notin \mathbb{R}^n_{++}$, meaning that it has at least one nonpositive element, then $P_{\mathbb{R}^n_{++}}(\mathbf{x})$ does not exist. In this case only the second part of (6.24) is relevant, meaning that $P_C(\mathbf{x}) = \text{prox}_{\lambda^* f}(\mathbf{x})$. To conclude,

$$P_C(\mathbf{x}) = \begin{cases} \mathbf{x}, & \mathbf{x} \in C, \\ \left(\frac{x_j + \sqrt{x_j^2 + 4\lambda^*}}{2}\right)_{j=1}^n, & \mathbf{x} \notin C, \end{cases}$$

where λ^* is any positive root of the nonincreasing function

$$\varphi(\lambda) = -\sum_{j=1}^n \log\left(\frac{x_j + \sqrt{x_j^2 + 4\lambda}}{2}\right) + \log \alpha.$$

∎

6.4.5 Projection onto Epigraphs

We can use Theorem 6.30 to prove a theorem on the projection onto epigraphs of convex functions.

Theorem 6.36 (orthogonal projection onto epigraphs). *Let*

$$C = \mathrm{epi}(g) = \{(\mathbf{x}, t) \in \mathbb{E} \times \mathbb{R} : g(\mathbf{x}) \le t\},$$

where $g : \mathbb{E} \to \mathbb{R}$ *is convex. Then*

$$P_C((\mathbf{x}, s)) = \begin{cases} (\mathbf{x}, s), & g(\mathbf{x}) \le s, \\ (\mathrm{prox}_{\lambda^* g}(\mathbf{x}), s + \lambda^*), & g(\mathbf{x}) > s, \end{cases}$$

where λ^* *is any positive root of the function*

$$\psi(\lambda) = g(\mathrm{prox}_{\lambda g}(\mathbf{x})) - \lambda - s.$$

In addition, ψ *is nonincreasing.*

Proof. Define $f : \mathbb{E} \times \mathbb{R} \to \mathbb{R}$ as $f(\mathbf{x}, t) \equiv g(\mathbf{x}) - t$. By definition of the prox,

$$\mathrm{prox}_{\lambda f}(\mathbf{x}, s) = \mathrm{argmin}_{\mathbf{y},t} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2 + \frac{1}{2}(t - s)^2 + \lambda f(\mathbf{y}, t) \right\}$$

$$= \mathrm{argmin}_{\mathbf{y},t} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2 + \frac{1}{2}(t - s)^2 + \lambda g(\mathbf{y}) - \lambda t \right\}.$$

The above problem is separable in $\mathbf{y}$ and $\mathbf{t}$, and thus

$$\mathrm{prox}_{\lambda f}(\mathbf{x}, s) = \left(\mathrm{argmin}_{\mathbf{y}} \left\{ \frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2 + \lambda g(\mathbf{y}) \right\}, \mathrm{argmin}_t \left\{ \frac{1}{2}(t - s)^2 - \lambda t \right\} \right)$$

$$= \left(\mathrm{prox}_{\lambda g}(\mathbf{x}), \mathrm{prox}_{\lambda h}(s) \right),$$

where $h(t) \equiv -t$. Since λh is linear, then by Section 6.2.2, $\mathrm{prox}_{\lambda h}(z) = z + \lambda$ for any $z \in \mathbb{R}$. Thus,

$$\mathrm{prox}_{\lambda f}(\mathbf{x}, s) = \left(\mathrm{prox}_{\lambda g}(\mathbf{x}), s + \lambda \right).$$

Since $\mathrm{epi}(g) = \mathrm{Lev}(f,0)$, we can invoke Theorem 6.30 (noting that $\mathrm{dom}(f) = \mathbb{E}$) and obtain that

$$P_C((\mathbf{x},s)) = \begin{cases} (\mathbf{x},s), & g(\mathbf{x}) \leq s, \\ (\mathrm{prox}_{\lambda^* g}(\mathbf{x}), s + \lambda), & g(\mathbf{x}) > s, \end{cases}$$

where λ^* is any positive root of the function

$$\psi(\lambda) = g(\mathrm{prox}_{\lambda g}(\mathbf{x})) - \lambda - s,$$

which by Theorem 6.30 is nonincreasing. $\Box$

Example 6.37 (projection onto the Lorentz cone). Consider the *Lorentz cone*, which is given by $L^n = \{(\mathbf{x},t) \in \mathbb{R}^n \times \mathbb{R} : \|\mathbf{x}\|_2 \leq t\}$. We will show that for any $(\mathbf{x},s) \in \mathbb{R}^n \times \mathbb{R}$,

$$P_{L^n}(\mathbf{x},s) = \begin{cases} \left(\frac{\|\mathbf{x}\|_2 + s}{2\|\mathbf{x}\|_2}\mathbf{x}, \frac{\|\mathbf{x}\|_2 + s}{2} \right), & \|\mathbf{x}\|_2 \geq |s|, \\ (\mathbf{0},0), & s < \|\mathbf{x}\|_2 < -s, \\ (\mathbf{x},s), & \|\mathbf{x}\|_2 \leq s. \end{cases}$$

To show the above,[31] we invoke Theorem 6.36 to obtain the formula

$$P_{L^n}((\mathbf{x},s)) = \begin{cases} (\mathbf{x},s), & \|\mathbf{x}\|_2 \leq s, \\ (\mathrm{prox}_{\lambda^* \|\cdot\|_2}(\mathbf{x}), s + \lambda^*), & \|\mathbf{x}\|_2 > s, \end{cases}$$

where λ^* is any positive root of the nonincreasing function

$$\psi(\lambda) = \|\mathrm{prox}_{\lambda \|\cdot\|_2}(\mathbf{x})\|_2 - \lambda - s. \tag{6.29}$$

Let $(\mathbf{x},s) \in \mathbb{R}^n \times \mathbb{R}$ be such that $\|\mathbf{x}\|_2 > s$. Recall that by Example 6.19,

$$\mathrm{prox}_{\lambda \|\cdot\|_2}(\mathbf{x}) = \left[1 - \frac{\lambda}{\max\{\|\mathbf{x}\|_2, \lambda\}} \right]_+ \mathbf{x}.$$

Plugging the above into the expression of ψ in (6.29) yields

$$\psi(\lambda) = \begin{cases} \|\mathbf{x}\|_2 - 2\lambda - s, & \lambda \leq \|\mathbf{x}\|_2, \\ -\lambda - s, & \lambda \geq \|\mathbf{x}\|_2. \end{cases}$$

The unique positive root λ^* of the piecewise linear function ψ is

$$\lambda^* = \begin{cases} \frac{\|\mathbf{x}\|_2 - s}{2}, & \|\mathbf{x}\|_2 \geq -s, \\ -s, & \|\mathbf{x}\|_2 < -s. \end{cases}$$

---

[31] Actually, the formula for $P_C(\mathbf{x})$ when $\|\mathbf{x}\|_2 = s$ appears twice in the formula, but in both cases it amounts to $(\mathbf{x},s)$.

Thus, in the case $\|\mathbf{x}\|_2 > s$ (noting that $\|\mathbf{x}\|_2 \geq -s$ corresponds to the case where $\|\mathbf{x}\|_2 \geq \lambda^*$ and $\|\mathbf{x}\|_2 < -s$ corresponds to $\|\mathbf{x}\|_2 \leq \lambda^*$),

$$
\begin{aligned}
(\text{prox}_{\lambda^* \|\cdot\|_2}(\mathbf{x}), s + \lambda^*) &= \left(\left[1 - \frac{\lambda^*}{\max\{\|\mathbf{x}\|_2, \lambda^*\}} \right]_+ \mathbf{x}, s + \lambda^* \right), \\
&= \begin{cases} \left(\left[1 - \frac{\|\mathbf{x}\|_2 - s}{2\|\mathbf{x}\|_2} \right]_+ \mathbf{x}, \frac{\|\mathbf{x}\|_2 + s}{2} \right), & \|\mathbf{x}\|_2 \geq -s, \\ (\mathbf{0}, 0), & \|\mathbf{x}\|_2 < -s. \end{cases} \\
&= \begin{cases} \left(\frac{\|\mathbf{x}\|_2 + s}{2\|\mathbf{x}\|_2} \mathbf{x}, \frac{\|\mathbf{x}\|_2 + s}{2} \right), & \|\mathbf{x}\|_2 \geq -s, \\ (\mathbf{0}, 0), & \|\mathbf{x}\|_2 < -s. \end{cases}
\end{aligned}
$$

Recalling that $\|\mathbf{x}\|_2 > s$, we have thus established that $P_{L^n}(\mathbf{x}, s) = (\mathbf{0}, 0)$ when $s < \|\mathbf{x}\|_2 < -s$ and that whenever $(\mathbf{x}, s)$ satisfies $\|\mathbf{x}\|_2 > s$ and $\|\mathbf{x}\|_2 \geq -s$, the formula

$$
P_{L^n}(\mathbf{x}, s) = \left(\frac{\|\mathbf{x}\|_2 + s}{2\|\mathbf{x}\|_2} \mathbf{x}, \frac{\|\mathbf{x}\|_2 + s}{2} \right) \tag{6.30}
$$

holds. The result now follows by noting that

$$
\{(\mathbf{x}, s) : \|\mathbf{x}\|_2 \geq |s|\} = \{(\mathbf{x}, s) : \|\mathbf{x}\|_2 > s, \|\mathbf{x}\|_2 \geq -s\} \cup \{(\mathbf{x}, s) : \|\mathbf{x}\|_2 = s\},
$$

and that formula (6.30) is trivial for the case $\|\mathbf{x}\|_2 = s$ (amounts to $P_{L^n}(\mathbf{x}, s) = (\mathbf{x}, s)$). ∎

Example 6.38 (projection onto the epigraph of the l_1-norm). Let

$$
C = \{(\mathbf{y}, t) \in \mathbb{R}^n \times \mathbb{R} : \|\mathbf{y}\|_1 \leq t\}.
$$

Invoking Theorem 6.36 and recalling that for any $\lambda > 0$, $\text{prox}_{\lambda \|\cdot\|_1} = \mathcal{T}_\lambda$, where $\mathcal{T}_\lambda$ is the soft thresholding operator (see Example 6.8), it follows that

$$
P_C((\mathbf{x}, s)) = \begin{cases} (\mathbf{x}, s), & \|\mathbf{x}\|_1 \leq s, \\ (\mathcal{T}_{\lambda^*}(\mathbf{x}), s + \lambda^*), & \|\mathbf{x}\|_1 > s, \end{cases}
$$

where λ^* is any positive root of the nonincreasing function

$$
\varphi(\lambda) = \|\mathcal{T}_\lambda(\mathbf{x})\|_1 - \lambda - s.
$$

∎

6.4.6 Summary of Orthogonal Projection Computations

Table 6.1 describes all the examples of orthogonal projection computations onto subsets of $\mathbb{R}^n$ and $\mathbb{R}^n \times \mathbb{R}$ that were discussed so far.

Table 6.1. *The following notation is used in the table.* $[\mathbf{x}]_+$ *is the non-negative part of* $\mathbf{x}$, $\mathcal{T}_\lambda(\mathbf{y}) = ([|y_i| - \lambda]_+ \mathrm{sgn}(y_i))_{i=1}^n$, *and* $\mathcal{S}_{\mathbf{a},\mathbf{b}}(\mathbf{x}) = (\min\{\max\{|x_i| - a_i, 0\}, b_i\}\mathrm{sgn}(x_i))_{i=1}^n$.

set (C)	$P_C(\mathbf{x})$	Assumptions	Reference
$\mathbb{R}_+^n$	$[\mathbf{x}]_+$	$-$	Lemma 6.26
$\mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]$	$P_C(\mathbf{x})_i = \min\{\max\{x_i, \ell_i\}, u_i\}$	$\ell_i \le u_i$	Lemma 6.26
$B_{\|\cdot\|_2}[\mathbf{c}, r]$	$\mathbf{c} + \frac{r}{\max\{\|\mathbf{x}-\mathbf{c}\|_2, r\}}(\mathbf{x} - \mathbf{c})$	$\mathbf{c} \in \mathbb{R}^n, r > 0$	Lemma 6.26
$\{\mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}$	$\mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{A}\mathbf{x} - \mathbf{b})$	$\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{A}$ full row rank	Lemma 6.26
$\{\mathbf{x} : \mathbf{a}^T\mathbf{x} \le b\}$	$\mathbf{x} - \frac{[\mathbf{a}^T\mathbf{x}-b]_+}{\|\mathbf{a}\|^2}\mathbf{a}$	$\mathbf{0} \ne \mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}$	Lemma 6.26
Δ_n	$[\mathbf{x} - \mu^*\mathbf{e}]_+$ where $\mu^* \in \mathbb{R}$ satisfies $\mathbf{e}^T[\mathbf{x} - \mu^*\mathbf{e}]_+ = 1$		Corollary 6.29
$H_{\mathbf{a},b} \cap \mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]$	$P_{\mathrm{Box}[\boldsymbol{\ell},\mathbf{u}]}(\mathbf{x} - \mu^*\mathbf{a})$ where $\mu^* \in \mathbb{R}$ satisfies $\mathbf{a}^T P_{\mathrm{Box}[\boldsymbol{\ell},\mathbf{u}]}(\mathbf{x} - \mu^*\mathbf{a}) = b$	$\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $b \in \mathbb{R}$	Theorem 6.27
$H_{\mathbf{a},b}^- \cap \mathrm{Box}[\boldsymbol{\ell}, \mathbf{u}]$	$\begin{cases} P_{\mathrm{Box}[\boldsymbol{\ell},\mathbf{u}]}(\mathbf{x}), & \mathbf{a}^T\mathbf{v_x} \le b, \\ P_{\mathrm{Box}[\boldsymbol{\ell},\mathbf{u}]}(\mathbf{x} - \lambda^*\mathbf{a}), & \mathbf{a}^T\mathbf{v_x} > b, \end{cases}$ $\mathbf{v_x} = P_{\mathrm{Box}[\boldsymbol{\ell},\mathbf{u}]}(\mathbf{x})$, $\mathbf{a}^T P_{\mathrm{Box}[\boldsymbol{\ell},\mathbf{u}]}(\mathbf{x} - \lambda^*\mathbf{a}) = b, \lambda^* > 0$	$\mathbf{a} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, $b \in \mathbb{R}$	Example 6.32
$B_{\|\cdot\|_1}[\mathbf{0}, \alpha]$	$\begin{cases} \mathbf{x}, & \|\mathbf{x}\|_1 \le \alpha, \\ \mathcal{T}_{\lambda^*}(\mathbf{x}), & \|\mathbf{x}\|_1 > \alpha, \end{cases}$ $\|\mathcal{T}_{\lambda^*}(\mathbf{x})\|_1 = \alpha, \lambda^* > 0$	$\alpha > 0$	Example 6.33
$\{\mathbf{x} : \boldsymbol{\omega}^T\|\mathbf{x}\| \le \beta,$ $-\boldsymbol{\alpha} \le \mathbf{x} \le \boldsymbol{\alpha}\}$	$\begin{cases} \mathbf{v_x}, & \boldsymbol{\omega}^T\|\mathbf{v_x}\| \le \beta, \\ \mathcal{S}_{\lambda^*\boldsymbol{\omega},\boldsymbol{\alpha}}(\mathbf{x}), & \boldsymbol{\omega}^T\|\mathbf{v_x}\| > \beta, \end{cases}$ $\mathbf{v_x} = P_{\mathrm{Box}[-\boldsymbol{\alpha},\boldsymbol{\alpha}]}(\mathbf{x})$, $\boldsymbol{\omega}^T\|\mathcal{S}_{\lambda^*\boldsymbol{\omega},\boldsymbol{\alpha}}(\mathbf{x})\| = \beta, \lambda^* > 0$	$\boldsymbol{\omega} \in \mathbb{R}_+^n$, $\boldsymbol{\alpha} \in [0,\infty]^n$, $\beta \in \mathbb{R}_{++}$	Example 6.34
$\{\mathbf{x} > \mathbf{0} : \Pi x_i \ge \alpha\}$	$\begin{cases} \mathbf{x}, & \mathbf{x} \in C, \\ \left(\frac{x_j + \sqrt{x_j^2 + 4\lambda^*}}{2}\right)_{j=1}^n, & \mathbf{x} \notin C, \end{cases}$ $\Pi_{j=1}^n\left((x_j + \sqrt{x_j^2 + 4\lambda^*})/2\right) = \alpha, \lambda^* > 0$	$\alpha > 0$	Example 6.35
$\{(\mathbf{x}, s) : \|\mathbf{x}\|_2 \le s\}$	$\left(\frac{\|\mathbf{x}\|_2 + s}{2\|\mathbf{x}\|_2}\mathbf{x}, \frac{\|\mathbf{x}\|_2 + s}{2}\right)$ if $\|\mathbf{x}\|_2 \ge |s|$ $(\mathbf{0}, 0)$ if $s < \|\mathbf{x}\|_2 < -s$, $(\mathbf{x}, s)$ if $\|\mathbf{x}\|_2 \le s$.		Example 6.37
$\{(\mathbf{x}, s) : \|\mathbf{x}\|_1 \le s\}$	$\begin{cases} (\mathbf{x}, s), & \|\mathbf{x}\|_1 \le s, \\ (\mathcal{T}_{\lambda^*}(\mathbf{x}), s + \lambda^*), & \|\mathbf{x}\|_1 > s, \end{cases}$ $\|\mathcal{T}_{\lambda^*}(\mathbf{x})\|_1 - \lambda^* - s = 0, \lambda^* > 0$		Example 6.38

6.5 The Second Prox Theorem

We can use Fermat's optimality condition (Theorem 3.63) in order to prove the second prox theorem.

Theorem 6.39 (second prox theorem). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed and convex function. Then for any $\mathbf{x}, \mathbf{u} \in \mathbb{E}$, the following three claims are equivalent:*

(i) $\mathbf{u} = \mathrm{prox}_f(\mathbf{x})$.

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

(iii) $\langle \mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{u} \rangle \leq f(\mathbf{y}) - f(\mathbf{u})$ *for any $\mathbf{y} \in \mathbb{E}$.*

Proof. By definition, $\mathbf{u} = \mathrm{prox}_f(\mathbf{x})$ if and only if $\mathbf{u}$ is the minimizer of the problem

$$\min_{\mathbf{v}} \left\{ f(\mathbf{v}) + \frac{1}{2}\|\mathbf{v} - \mathbf{x}\|^2 \right\},$$

which, by Fermat's optimality condition (Theorem 3.63) and the sum rule of sub-differential calculus (Theorem 3.40), is equivalent to the relation

$$\mathbf{0} \in \partial f(\mathbf{u}) + \mathbf{u} - \mathbf{x}. \tag{6.31}$$

We have thus shown the equivalence between claims (i) and (ii). Finally, by the definition of the subgradient, the membership relation of claim (ii) is equivalent to (iii). $\quad\square$

A direct consequence of the second prox theorem is that for a proper closed and convex function, $\mathbf{x} = \mathrm{prox}_f(\mathbf{x})$ if and only $\mathbf{x}$ is a minimizer of f.

Corollary 6.40. *Let f be a proper closed and convex function. Then $\mathbf{x}$ is a minimizer of f if and only if $\mathbf{x} = \mathrm{prox}_f(\mathbf{x})$.*

Proof. $\mathbf{x}$ is a minimizer of f if and only if $\mathbf{0} \in \partial f(\mathbf{x})$, that is, if and only if $\mathbf{x} - \mathbf{x} \in \partial f(\mathbf{x})$, which by the second prox theorem (equivalence between (i) and (ii)) is the same as $\mathbf{x} = \mathrm{prox}_f(\mathbf{x})$. $\quad\square$

When $f = \delta_C$, with C being a nonempty closed and convex set, the equivalence between claims (i) and (iii) in the second prox theorem amounts to the second projection theorem.

Theorem 6.41 (second projection theorem). *Let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Let $\mathbf{u} \in C$. Then $\mathbf{u} = P_C(\mathbf{x})$ if and only if*

$$\langle \mathbf{x} - \mathbf{u}, \mathbf{y} - \mathbf{u} \rangle \leq 0 \text{ for any } \mathbf{y} \in C.$$

Another rather direct result of the second prox theorem is the firm nonexpansivity of the prox operator.

Theorem 6.42 (firm nonexpansivity of the prox operator). *Let f be a proper closed and convex function. Then for any $\mathbf{x}, \mathbf{y} \in \mathbb{E}$,*

(a) **(firm nonexpansivity)**

$$\langle \mathbf{x} - \mathbf{y}, \operatorname{prox}_f(\mathbf{x}) - \operatorname{prox}_f(\mathbf{y}) \rangle \geq \|\operatorname{prox}_f(\mathbf{x}) - \operatorname{prox}_f(\mathbf{y})\|^2.$$

(b) **(nonexpansivity)**

$$\|\operatorname{prox}_f(\mathbf{x}) - \operatorname{prox}_f(\mathbf{y})\| \leq \|\mathbf{x} - \mathbf{y}\|.$$

Proof. (a) Denoting $\mathbf{u} = \operatorname{prox}_f(\mathbf{x}), \mathbf{v} = \operatorname{prox}_f(\mathbf{y})$, by the equivalence of (i) and (ii) in the second prox theorem (Theorem 6.39), it follows that

$$\mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u}), \mathbf{y} - \mathbf{v} \in \partial f(\mathbf{v}).$$

Thus, by the subgradient inequality,

$$f(\mathbf{v}) \geq f(\mathbf{u}) + \langle \mathbf{x} - \mathbf{u}, \mathbf{v} - \mathbf{u} \rangle,$$
$$f(\mathbf{u}) \geq f(\mathbf{v}) + \langle \mathbf{y} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle.$$

Summing the above two inequalities, we obtain

$$0 \geq \langle \mathbf{y} - \mathbf{x} + \mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v} \rangle,$$

which is the same as

$$\langle \mathbf{x} - \mathbf{y}, \mathbf{u} - \mathbf{v} \rangle \geq \|\mathbf{u} - \mathbf{v}\|^2,$$

that is,

$$\langle \mathbf{x} - \mathbf{y}, \operatorname{prox}_f(\mathbf{x}) - \operatorname{prox}_f(\mathbf{y}) \rangle \geq \|\operatorname{prox}_f(\mathbf{x}) - \operatorname{prox}_f(\mathbf{y})\|^2.$$

(b) If $\operatorname{prox}_f(\mathbf{x}) = \operatorname{prox}_f(\mathbf{y})$, then the inequality is obvious. Assume that $\operatorname{prox}_f(\mathbf{x}) \neq \operatorname{prox}_f(\mathbf{y})$. Using (a) and the Cauchy–Schwarz inequality, it follows that

$$\|\operatorname{prox}_f(\mathbf{x}) - \operatorname{prox}_f(\mathbf{y})\|^2 \leq \langle \operatorname{prox}_h(\mathbf{x}) - \operatorname{prox}_h(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle$$
$$\leq \|\operatorname{prox}_h(\mathbf{x}) - \operatorname{prox}_h(\mathbf{y})\| \cdot \|\mathbf{x} - \mathbf{y}\|.$$

Dividing by $\|\operatorname{prox}_h(\mathbf{x}) - \operatorname{prox}_h(\mathbf{y})\|$, the desired result is established. $\quad\square$

The following result shows how to compute the prox of the distance function to a nonempty closed and convex set. The proof is heavily based on the second prox theorem.

Lemma 6.43 (prox of the distance function). *Let $C \subseteq \mathbb{E}$ be a nonempty, closed, and convex set. Let $\lambda > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\operatorname{prox}_{\lambda d_C}(\mathbf{x}) = \begin{cases} (1 - \theta)\mathbf{x} + \theta P_C(\mathbf{x}), & d_C(\mathbf{x}) > \lambda, \\ P_C(\mathbf{x}), & d_C(\mathbf{x}) \leq \lambda, \end{cases} \tag{6.32}$$

where[32]

$$\theta = \frac{\lambda}{d_C(\mathbf{x})}. \tag{6.33}$$

Proof. Let $\mathbf{u} = \text{prox}_{\lambda d_C}(\mathbf{x})$. By the second prox theorem (Theorem 6.39),

$$\mathbf{x} - \mathbf{u} \in \lambda \partial d_C(\mathbf{u}). \tag{6.34}$$

We will split the analysis into two cases.

Case I. $\mathbf{u} \notin C$. By Example 3.49, (6.34) is the same as

$$\mathbf{x} - \mathbf{u} = \lambda \frac{\mathbf{u} - P_C(\mathbf{u})}{d_C(\mathbf{u})}.$$

Denoting $\alpha = \frac{\lambda}{d_C(\mathbf{u})}$, the last equality can be rewritten as

$$\mathbf{u} = \frac{1}{\alpha+1}\mathbf{x} + \frac{\alpha}{\alpha+1}P_C(\mathbf{u}) \tag{6.35}$$

or as

$$\mathbf{x} - P_C(\mathbf{u}) = (\alpha+1)(\mathbf{u} - P_C(\mathbf{u})). \tag{6.36}$$

By the second projection theorem (Theorem 6.41), in order to show that $P_C(\mathbf{u}) = P_C(\mathbf{x})$, it is enough to show that

$$\langle \mathbf{x} - P_C(\mathbf{u}), \mathbf{y} - P_C(\mathbf{u}) \rangle \leq 0 \text{ for any } \mathbf{y} \in C. \tag{6.37}$$

Using (6.36), we can deduce that (6.37) is equivalent to

$$(\alpha+1)\langle \mathbf{u} - P_C(\mathbf{u}), \mathbf{y} - P_C(\mathbf{u}) \rangle \leq 0 \text{ for any } \mathbf{y} \in C,$$

which is a valid inequality by the second projection theorem, and hence $P_C(\mathbf{u}) = P_C(\mathbf{x})$. Using this fact and taking the norm in both sides of (6.36), we obtain that

$$d_C(\mathbf{x}) = (\alpha+1)d_C(\mathbf{u}) = d_C(\mathbf{u}) + \lambda,$$

which also shows that in this case $d_C(\mathbf{x}) > \lambda$ (since $d_C(\mathbf{u}) > 0$) and that

$$\frac{1}{\alpha+1} = \frac{d_C(\mathbf{u})}{\lambda + d_C(\mathbf{u})} = \frac{d_C(\mathbf{x}) - \lambda}{d_C(\mathbf{x})} = 1 - \theta,$$

where θ is given in (6.33). Therefore, (6.35) can also be written as (recalling also that $P_C(\mathbf{u}) = P_C(\mathbf{x})$)

$$\text{prox}_{\lambda d_C}(\mathbf{x}) = (1-\theta)\mathbf{x} + \theta P_C(\mathbf{x}). \tag{6.38}$$

Case II. If $\mathbf{u} \in C$, then $\mathbf{u} = P_C(\mathbf{x})$. To show this, let $\mathbf{v} \in C$. Since $\mathbf{u} = \text{prox}_{\lambda d_C}(\mathbf{x})$, it follows in particular that

$$\lambda d_C(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \leq \lambda d_C(\mathbf{v}) + \frac{1}{2}\|\mathbf{v} - \mathbf{x}\|^2,$$

---

[32]Since θ is used only when $\mathbf{x} \notin C$, it follows that $d_C(\mathbf{x}) > 0$, so that θ is well defined.

and hence, since $d_C(\mathbf{u}) = d_C(\mathbf{v}) = 0$,

$$\|\mathbf{u} - \mathbf{x}\| \leq \|\mathbf{v} - \mathbf{x}\|.$$

Therefore,

$$\mathbf{u} = \operatorname{argmin}_{\mathbf{v} \in C} \|\mathbf{v} - \mathbf{x}\| = P_C(\mathbf{x}).$$

By Example 3.49, the optimality condition (6.34) becomes

$$\frac{\mathbf{x} - P_C(\mathbf{x})}{\lambda} \in N_C(\mathbf{u}) \cap B[\mathbf{0}, 1],$$

which in particular implies that

$$\left\| \frac{\mathbf{x} - P_C(\mathbf{x})}{\lambda} \right\| \leq 1,$$

that is,

$$d_C(\mathbf{x}) = \|P_C(\mathbf{x}) - \mathbf{x}\| \leq \lambda.$$

Since the first case in which (6.38) holds corresponds to vectors satisfying $d_C(\mathbf{x}) > \lambda$, while the second case in which $\operatorname{prox}_{\lambda d_C}(\mathbf{x}) = P_C(\mathbf{x})$ corresponds to vectors satisfying $d_C(\mathbf{x}) \leq \lambda$, the desired result (6.32) is established.   □

6.6   Moreau Decomposition

A key property of the prox operator is the so-called Moreau decomposition theorem, which connects the prox operator of proper closed convex functions and their conjugates.

Theorem 6.44 (Moreau decomposition).   *Let $f : \mathbb{E} \to (-\infty, \infty]$ be proper closed and convex. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\operatorname{prox}_f(\mathbf{x}) + \operatorname{prox}_{f^*}(\mathbf{x}) = \mathbf{x}.$$

Proof.   Let $\mathbf{x} \in \mathbb{E}$ and denote $\mathbf{u} = \operatorname{prox}_f(\mathbf{x})$. Then by the equivalence between claims (i) and (ii) in the second prox theorem (Theorem 6.39), it follows that $\mathbf{x} - \mathbf{u} \in \partial f(\mathbf{u})$, which by the conjugate subgradient theorem (Theorem 4.20) is equivalent to $\mathbf{u} \in \partial f^*(\mathbf{x} - \mathbf{u})$. Using the second prox theorem again, we conclude that $\mathbf{x} - \mathbf{u} = \operatorname{prox}_{f^*}(\mathbf{x})$. Therefore,

$$\operatorname{prox}_f(\mathbf{x}) + \operatorname{prox}_{f^*}(\mathbf{x}) = \mathbf{u} + (\mathbf{x} - \mathbf{u}) = \mathbf{x}.   □$$

The next result is a useful extension of the Moreau decomposition theorem.

Theorem 6.45 (extended Moreau decomposition).   *Let $f : \mathbb{E} \to (-\infty, \infty]$ be proper closed and convex, and let $\lambda > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\operatorname{prox}_{\lambda f}(\mathbf{x}) + \lambda \operatorname{prox}_{\lambda^{-1} f^*}(\mathbf{x}/\lambda) = \mathbf{x}. \tag{6.39}$$

Proof. Using Moreau decomposition, for any $\mathbf{x} \in \mathbb{E}$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \text{prox}_{(\lambda f)^*}(\mathbf{x}) = \mathbf{x} - \text{prox}_{\lambda f^*(\cdot/\lambda)}(\mathbf{x}), \tag{6.40}$$

where the second equality follows by Theorem 4.14(a). By Theorem 6.12,

$$\text{prox}_{\lambda f^*(\cdot/\lambda)}(\mathbf{x}) = \lambda \text{prox}_{\lambda^{-1} f^*}(\mathbf{x}/\lambda),$$

which, combined with (6.40), yields (6.39). $\quad\square$

6.6.1 Support Functions

Using Moreau decomposition, we can develop a formula for computing the prox of a support function of a given nonempty closed and convex set in terms of the orthogonal projection operator.

Theorem 6.46 (prox of support functions). *Let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set, and let $\lambda > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\text{prox}_{\lambda \sigma_C}(\mathbf{x}) = \mathbf{x} - \lambda P_C(\mathbf{x}/\lambda). \tag{6.41}$$

Proof. A direct consequence of the extended Moreau decomposition formula (Theorem 6.45) along with the fact that $(\sigma_C)^* = \delta_C$ (Example 4.9). $\quad\square$

Following are several examples of prox computations using formula (6.41).

Example 6.47 (prox of norms). Let $f : \mathbb{E} \to \mathbb{R}$ be given by $f(\mathbf{x}) = \lambda \|\mathbf{x}\|_\alpha$, where $\lambda > 0$ and $\| \cdot \|_\alpha$ is any norm on $\mathbb{E}$. Note that $\| \cdot \|_\alpha$ is not necessarily the endowed norm on $\mathbb{E}$, which is denoted by $\| \cdot \|$ and in this chapter is always assumed to be the Euclidean norm. We know by Example 2.31 that

$$\|\mathbf{x}\|_\alpha = \sigma_C(\mathbf{x}),$$

where

$$C = B_{\|\cdot\|_{\alpha,*}}[\mathbf{0}, 1] = \{\mathbf{x} \in \mathbb{E} : \|\mathbf{x}\|_{\alpha,*} \le 1\}$$

with $\| \cdot \|_{\alpha,*}$ being the dual norm of $\| \cdot \|_\alpha$. Invoking Theorem 6.46, we obtain

$$\text{prox}_{\lambda \|\cdot\|_\alpha}(\mathbf{x}) = \mathbf{x} - \lambda P_{B_{\|\cdot\|_{\alpha,*}}[\mathbf{0},1]}(\mathbf{x}/\lambda).$$

$\blacksquare$

Example 6.48 (prox of l_∞-norm). By Example 6.47 we have for all $\lambda > 0$ and $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda \|\cdot\|_\infty}(\mathbf{x}) = \mathbf{x} - \lambda P_{B_{\|\cdot\|_1}[\mathbf{0},1]}(\mathbf{x}/\lambda).$$

The projection onto the l_1 unit ball can be easily computed by finding a root of a nonincreasing one-dimensional function; see Example 6.33. ∎

Example 6.49 (prox of the max function). Consider the max function $g : \mathbb{R}^n \to \mathbb{R}$ given by $g(\mathbf{x}) = \max(\mathbf{x}) \equiv \max\{x_1, x_2, \ldots, x_n\}$. It is easy to see that the max function is actually the support function of the unit simplex:

$$\max(\mathbf{x}) = \sigma_{\Delta_n}(\mathbf{x}).$$

Therefore, by Theorem 6.46, for any $\lambda > 0$ and $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda \max(\cdot)}(\mathbf{x}) = \mathbf{x} - \lambda P_{\Delta_n}(\mathbf{x}/\lambda).$$

The projection onto the unit simplex can be efficiently computed by finding a root of a nonincreasing one-dimensional function; see Corollary 6.29. ∎

Example 6.50 (prox of the sum-of-k-largest-values function). Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by

$$f(\mathbf{x}) = x_{[1]} + x_{[2]} + \cdots + x_{[k]},$$

where $k \in \{1, 2, \ldots, n\}$ and for any i, $x_{[i]}$ denotes ith largest value in the vector $\mathbf{x}$. It is not difficult to show that $f = \sigma_C$, where

$$C = \{\mathbf{y} \in \mathbb{R}^n : \mathbf{e}^T \mathbf{y} = k, \mathbf{0} \leq \mathbf{y} \leq \mathbf{e}\}.$$

Therefore, by Theorem 6.46,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \lambda P_C(\mathbf{x}/\lambda).$$

That is, for any $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \lambda P_{\{\mathbf{y}:\mathbf{e}^T \mathbf{y} = k, \mathbf{0} \leq \mathbf{y} \leq \mathbf{e}\}}(\mathbf{x}/\lambda).$$

As in the previous examples, computing the projection onto C amounts to finding a root of a monotone one-dimensional function; see Theorem 6.27. ∎

Example 6.51 (prox of the sum-of-k-largest-absolute-values function). Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by

$$f(\mathbf{x}) = \sum_{i=1}^{k} |x_{\langle i \rangle}|,$$

where $k \in \{1, 2, \ldots, n\}$ and $x_{\langle i \rangle}$ is the component of $\mathbf{x}$ with the ith largest absolute value, meaning in particular that $|x_{\langle 1 \rangle}| \geq |x_{\langle 2 \rangle}| \geq \cdots \geq |x_{\langle n \rangle}|$. Then

$$f(\mathbf{x}) = \max \left\{ \sum_{i=1}^{n} z_i x_i : \|\mathbf{z}\|_1 \leq k, -\mathbf{e} \leq \mathbf{z} \leq \mathbf{e} \right\}.$$

Therefore, $f = \sigma_C$, where

$$C = \{\mathbf{z} \in \mathbb{R}^n : \|\mathbf{z}\|_1 \leq k, -\mathbf{e} \leq \mathbf{z} \leq \mathbf{e}\},$$

and consequently, by Theorem 6.46,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \lambda P_C(\mathbf{x}/\lambda).$$

That is, for any $\mathbf{x} \in \mathbb{R}^n$,

$$\text{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \lambda P_{\{\mathbf{y}:\|\mathbf{y}\|_1 \leq k, -\mathbf{e} \leq \mathbf{y} \leq \mathbf{e}\}}(\mathbf{x}/\lambda).$$

The orthogonal projection in the above formula amounts to finding a root of a nonincreasing one-dimensional function; see Example 6.34. ∎

6.7 The Moreau Envelope

6.7.1 Definition and Basic Properties

Definition 6.52 (Moreau envelope). *Given a proper closed convex function* $f : \mathbb{E} \to (-\infty, \infty]$ *and* $\mu > 0$, *the* **Moreau envelope** *of* f *is the function*

$$M_f^\mu(\mathbf{x}) = \min_{\mathbf{u} \in \mathbb{E}} \left\{ f(\mathbf{u}) + \frac{1}{2\mu}\|\mathbf{x} - \mathbf{u}\|^2 \right\}. \tag{6.42}$$

The parameter μ is called the *smoothing parameter*. The explanation for this terminology will be given in Section 6.7.2. By the first prox theorem (Theorem 6.3), the minimization problem in (6.42) has a unique solution, given by $\text{prox}_{\mu f}(\mathbf{x})$. Therefore, $M_f^\mu(\mathbf{x})$ is always a real number and

$$M_f^\mu(\mathbf{x}) = f(\text{prox}_{\mu f}(\mathbf{x})) + \frac{1}{2\mu}\|\mathbf{x} - \text{prox}_{\mu f}(\mathbf{x})\|^2.$$

Example 6.53 (Moreau envelope of indicators). Let $f = \delta_C$, where $C \subseteq \mathbb{E}$ is a nonempty closed and convex set. By Theorem 6.24, $\text{prox}_{\mu f}(\mathbf{x}) = P_C(\mathbf{x})$. Thus, for any $\mathbf{x} \in \mathbb{E}$,

$$M_f^\mu(\mathbf{x}) = \delta_C(P_C(\mathbf{x})) + \frac{1}{2\mu}\|\mathbf{x} - P_C(\mathbf{x}))\|^2,$$

and hence

$$M_{\delta_C}^\mu = \frac{1}{2\mu}d_C^2.$$

∎

The next example will show that the Moreau envelope of the (Euclidean) norm is the so-called *Huber function* defined as

$$H_\mu(\mathbf{x}) = \begin{cases} \frac{1}{2\mu}\|\mathbf{x}\|^2, & \|\mathbf{x}\| \leq \mu, \\ \|\mathbf{x}\| - \frac{\mu}{2}, & \|\mathbf{x}\| > \mu. \end{cases} \tag{6.43}$$

The one-dimensional Huber function is plotted in Figure 6.4, where it is illustrated that the function becomes smoother as μ becomes larger.
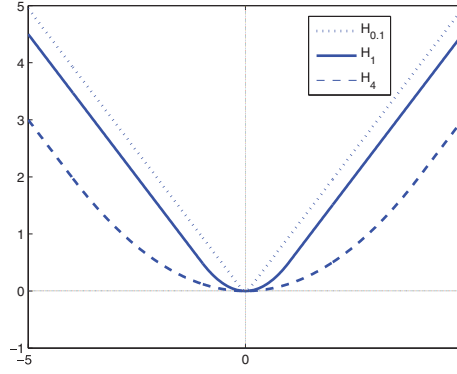
Figure 6.4. *The Huber function with parameters $\mu = 0.1, 1, 4$. The function becomes smoother as μ gets larger.*

Example 6.54 (Huber function). Let $f : \mathbb{E} \to \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|$. Then by Example 6.19, for any $\mathbf{x} \in \mathbb{E}$ and $\mu > 0$,

$$\mathrm{prox}_{\mu f}(\mathbf{x}) = \left(1 - \frac{\mu}{\max\{\|\mathbf{x}\|, \mu\}} \right) \mathbf{x}.$$

Therefore,

$$M_f^\mu(\mathbf{x}) = \|\mathrm{prox}_{\mu f}(\mathbf{x})\| + \frac{1}{2\mu}\|\mathbf{x} - \mathrm{prox}_{\mu f}(\mathbf{x})\|^2 = \left\{ \begin{array}{ll} \frac{1}{2\mu}\|\mathbf{x}\|^2, & \|\mathbf{x}\| \leq \mu, \\[2mm] \|\mathbf{x}\| - \frac{\mu}{2}, & \|\mathbf{x}\| > \mu. \end{array} \right.$$

Thus, for any $\mu > 0$,

$$M_{\|\cdot\|}^\mu = H_\mu.$$

∎

Note that the Moreau envelope function is actually a result of an infimal convolution operation between the function f and the function

$$\omega_\mu(\mathbf{x}) = \frac{1}{2\mu}\|\mathbf{x}\|^2. \tag{6.44}$$

That is,

$$M_f^\mu = f \square \omega_\mu.$$

One consequence of this observation is that by Theorem 2.19, if f is a proper closed[33] and convex function, then M_f^μ is convex. We summarize the above discussion in the following theorem.

---

[33]Actually, closedness is not necessary in order to establish the convexity of the Moreau envelope.

Theorem 6.55. *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed and convex function, and let ω_μ be given in (6.44), where $\mu > 0$. Then*

(a) $M_f^\mu = f \square \omega_\mu$;

(b) $M_f^\mu : \mathbb{E} \to \mathbb{R}$ *is real-valued and convex.*

We can immediately conclude from Theorem 6.55(a) along with the formula for the conjugate of the infimal convolution (Theorem 4.16) an expression for the conjugate of the Moreau envelope.

Corollary 6.56. *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed and convex function and let ω_μ be given in (6.44), where $\mu > 0$. Then*

$$(M_f^\mu)^* = f^* + \omega_{\frac{1}{\mu}}.$$

Another useful algebraic property of the Moreau envelope is described in the following result.

Lemma 6.57. *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed and convex function, and let $\lambda, \mu > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\lambda M_f^\mu(\mathbf{x}) = M_{\lambda f}^{\mu/\lambda}(\mathbf{x}). \tag{6.45}$$

Proof. For any $\mathbf{x} \in \mathbb{E}$,

$$\begin{aligned}
\lambda M_f^\mu(\mathbf{x}) &= \lambda \min_{\mathbf{u}} \left\{ f(\mathbf{u}) + \frac{1}{2\mu} \|\mathbf{u} - \mathbf{x}\|^2 \right\} \\
&= \min_{\mathbf{u}} \left\{ \lambda f(\mathbf{u}) + \frac{1}{2\mu/\lambda} \|\mathbf{u} - \mathbf{x}\|^2 \right\} \\
&= M_{\lambda f}^{\mu/\lambda}(\mathbf{x}). \quad \square
\end{aligned}$$

A simple calculus rule states that the Moreau envelope of a separable sum of functions is the sum of the corresponding Moreau envelopes.

Theorem 6.58 (Moreau envelope of separable functions). *Suppose that $\mathbb{E} = \mathbb{E}_1 \times \mathbb{E}_2 \times \cdots \times \mathbb{E}_m$, and let $f : \mathbb{E} \to (-\infty, \infty]$ be given by*

$$f(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) = \sum_{i=1}^m f_i(\mathbf{x}_i), \quad \mathbf{x}_1 \in \mathbb{E}_1, \mathbf{x}_2 \in \mathbb{E}_2, \ldots, \mathbf{x}_m \in \mathbb{E}_m,$$

with $f_i : \mathbb{E}_i \to (-\infty, \infty]$ being a proper closed and convex function for any i. Then given $\mu > 0$, for any $\mathbf{x}_1 \in \mathbb{E}_1, \mathbf{x}_2 \in \mathbb{E}_2, \ldots, \mathbf{x}_m \in \mathbb{E}_m$,

$$M_f^\mu(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m) = \sum_{i=1}^m M_{f_i}^\mu(\mathbf{x}_i).$$

Proof. For any $\mathbf{x}_1 \in \mathbb{E}_1, \mathbf{x}_2 \in \mathbb{E}_2, \ldots, \mathbf{x}_m \in \mathbb{E}_m$, denoting $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$, we have

$$M_f^\mu(\mathbf{x}) = \min_{\mathbf{u}_i \in \mathbb{E}_i, i=1,2,\ldots,m} \left\{ f(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m) + \frac{1}{2\mu} \|(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m) - \mathbf{x}\|^2 \right\}$$

$$= \min_{\mathbf{u}_i \in \mathbb{E}_i, i=1,2,\ldots,m} \left\{ \sum_{i=1}^m f_i(\mathbf{u}_i) + \frac{1}{2\mu} \sum_{i=1}^m \|\mathbf{u}_i - \mathbf{x}_i\|^2 \right\}$$

$$= \sum_{i=1}^m \min_{\mathbf{u}_i \in \mathbb{E}_i} \left\{ f_i(\mathbf{u}_i) + \frac{1}{2\mu} \|\mathbf{u}_i - \mathbf{x}_i\|^2 \right\}$$

$$= \sum_{i=1}^m M_{f_i}^\mu(\mathbf{x}_i). \quad \square$$

Example 6.59 (Moreau envelope of the l_1-norm). Consider the function $f : \mathbb{R}^n \to \mathbb{R}$ given by $f(\mathbf{x}) = \|\mathbf{x}\|_1$. Note that

$$f(\mathbf{x}) = \|\mathbf{x}\|_1 = \sum_{i=1}^n g(x_i),$$

where $g(t) = |t|$. By Example 6.54, $M_g^\mu = H_\mu$. Thus, invoking Theorem 6.58, we obtain that for any $\mathbf{x} \in \mathbb{R}^n$,

$$M_f^\mu(\mathbf{x}) = \sum_{i=1}^n M_g^\mu(x_i) = \sum_{i=1}^n H_\mu(x_i). \quad \blacksquare$$

6.7.2 Differentiability of the Moreau Envelope

The main differentiability properties of the Moreau envelope function are stated in the next result.

Theorem 6.60 (smoothness of the Moreau envelope). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed and convex function. Let $\mu > 0$. Then M_f^μ is $\frac{1}{\mu}$-smooth over $\mathbb{E}$, and for any $\mathbf{x} \in \mathbb{E}$,*

$$\nabla M_f^\mu(\mathbf{x}) = \frac{1}{\mu} \left(\mathbf{x} - \mathrm{prox}_{\mu f}(\mathbf{x}) \right).$$

Proof. By Theorem 6.55(a), $M_f^\mu = f \square \omega_\mu$, where $\omega_\mu = \frac{1}{2\mu} \| \cdot \|^2$. We can therefore invoke Theorem 5.30, whose assumptions are satisfied (taking $\omega = \omega_\mu$ and $L = \frac{1}{\mu}$), and conclude that M_f^μ is $\frac{1}{\mu}$-smooth. In addition, since

$$\mathrm{prox}_{\mu f}(\mathbf{x}) = \mathrm{argmin}_{\mathbf{u} \in \mathbb{E}} \left\{ f(\mathbf{u}) + \frac{1}{2\mu} \|\mathbf{u} - \mathbf{x}\|^2 \right\},$$

it follows that the vector $\mathbf{u}(\mathbf{x})$ defined in Theorem 5.30 is equal to $\mathrm{prox}_{\mu f}(\mathbf{x})$ and that

$$\nabla M_f^\mu(\mathbf{x}) = \nabla \omega_\mu(\mathbf{x} - \mathbf{u}(\mathbf{x})) = \frac{1}{\mu}(\mathbf{x} - \mathrm{prox}_{\mu f}(\mathbf{x})). \quad \square$$

Example 6.61 (1-smoothness of $\frac{1}{2}d_C^2$). Let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set. Recall that by Example 6.53, $\frac{1}{2}d_C^2 = M_{\delta_C}^1$. Then by Theorem 6.60, $\frac{1}{2}d_C^2$ is 1-smooth and

$$\nabla \left(\frac{1}{2}d_C^2 \right)(\mathbf{x}) = \mathbf{x} - \operatorname{prox}_{\delta_C}(\mathbf{x}) = \mathbf{x} - P_C(\mathbf{x}).$$

Note that the above expression for the gradient was already derived in Example 3.31 and that the 1-smoothness of $\frac{1}{2}d_C^2$ was already established twice in Examples 5.5 and 5.31. ∎

Example 6.62 (smoothness of the Huber function). Recall that the Huber function is given by

$$H_\mu(\mathbf{x}) = \begin{cases} \frac{1}{2\mu}\|\mathbf{x}\|^2, & \|\mathbf{x}\| \leq \mu, \\ \|\mathbf{x}\| - \frac{\mu}{2}, & \|\mathbf{x}\| > \mu. \end{cases}$$

By Example 6.54, $H_\mu = M_f^\mu$, where $f(\mathbf{x}) = \|\mathbf{x}\|$. Then, by Theorem 6.60, H_μ is $\frac{1}{\mu}$-smooth and

$$\begin{aligned} \nabla H_\mu(\mathbf{x}) &= \frac{1}{\mu}\left(\mathbf{x} - \operatorname{prox}_{\mu f}(\mathbf{x}) \right) \\ &\overset{(*)}{=} \frac{1}{\mu}\left(\mathbf{x} - \left(1 - \frac{\mu}{\max\{\|\mathbf{x}\|, \mu\}} \right)\mathbf{x} \right) \\ &= \begin{cases} \frac{1}{\mu}\mathbf{x}, & \|\mathbf{x}\| \leq \mu, \\ \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \|\mathbf{x}\| > \mu, \end{cases} \end{aligned}$$

where the equality $(*)$ uses the expression for $\operatorname{prox}_{\mu f}$ developed in Example 6.19. ∎

6.7.3 Prox of the Moreau Envelope

An interesting and important result states that if we can compute the prox of a proper closed and convex function f, then we can also compute the prox of its Moreau envelope.

Theorem 6.63 (prox of Moreau envelope). *Let* $f : \mathbb{E} \to (-\infty, \infty]$ *be a proper closed and convex function, and let* $\mu > 0$. *Then for any* $\mathbf{x} \in \mathbb{E}$,

$$\operatorname{prox}_{M_f^\mu}(\mathbf{x}) = \mathbf{x} + \frac{1}{\mu + 1}\left(\operatorname{prox}_{(\mu+1)f}(\mathbf{x}) - \mathbf{x} \right).$$

Proof. First note that

$$\min_{\mathbf{u}} \left\{ M_f^\mu(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\} = \min_{\mathbf{u}} \min_{\mathbf{y}} \left\{ f(\mathbf{y}) + \frac{1}{2\mu}\|\mathbf{u} - \mathbf{y}\|^2 + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\}. \tag{6.46}$$

Exchanging the order of minimizations, we obtain the following problem:

$$\min_{\mathbf{y}} \min_{\mathbf{u}} \left\{ f(\mathbf{y}) + \frac{1}{2\mu}\|\mathbf{u} - \mathbf{y}\|^2 + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|^2 \right\}. \tag{6.47}$$

The optimal solution of the inner minimization problem in $\mathbf{u}$ is attained when the gradient w.r.t. $\mathbf{u}$ vanishes:

$$\frac{1}{\mu}(\mathbf{u} - \mathbf{y}) + (\mathbf{u} - \mathbf{x}) = \mathbf{0},$$

that is, when

$$\mathbf{u} = \mathbf{u}_\mu \equiv \frac{\mu\mathbf{x} + \mathbf{y}}{\mu + 1}. \tag{6.48}$$

Therefore, the optimal value of the inner minimization problem in (6.47) is

$$f(\mathbf{y}) + \frac{1}{2\mu}\|\mathbf{u}_\mu - \mathbf{y}\|^2 + \frac{1}{2}\|\mathbf{u}_\mu - \mathbf{x}\|^2 = f(\mathbf{y}) + \frac{1}{2\mu}\left\|\frac{\mu\mathbf{x} - \mu\mathbf{y}}{\mu + 1}\right\|^2 + \frac{1}{2}\left\|\frac{\mathbf{y} - \mathbf{x}}{\mu + 1}\right\|^2$$

$$= f(\mathbf{y}) + \frac{1}{2(\mu + 1)}\|\mathbf{x} - \mathbf{y}\|^2.$$

Therefore, the optimal solution of (6.46) is given by (6.48), where $\mathbf{y}$ is the solution of

$$\min_{\mathbf{y}} \left\{ f(\mathbf{y}) + \frac{1}{2(\mu + 1)}\|\mathbf{x} - \mathbf{y}\|^2 \right\},$$

that is, $\mathbf{y} = \mathrm{prox}_{(\mu+1)f}(\mathbf{x})$. To summarize,

$$\mathrm{prox}_{M_f^\mu}(\mathbf{x}) = \frac{1}{\mu + 1}\left(\mu\mathbf{x} + \mathrm{prox}_{(\mu+1)f}(\mathbf{x}) \right). \quad \square$$

Combining Theorem 6.63 with Lemma 6.57 leads to the following corollary.

Corollary 6.64. *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed and convex function, and let $\lambda, \mu > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$\mathrm{prox}_{\lambda M_f^\mu}(\mathbf{x}) = \mathbf{x} + \frac{\lambda}{\mu + \lambda}\left(\mathrm{prox}_{(\mu+\lambda)f}(\mathbf{x}) - \mathbf{x} \right).$$

Proof. $\mathrm{prox}_{\lambda M_f^\mu}(\mathbf{x}) = \mathrm{prox}_{M_{\lambda f}^{\mu/\lambda}}(\mathbf{x}) = \mathbf{x} + \frac{\lambda}{\mu + \lambda}\left(\mathrm{prox}_{(\mu+\lambda)f}(\mathbf{x}) - \mathbf{x} \right). \quad \square$

Example 6.65 (prox of $\frac{\lambda}{2}d_C^2$). Let $C \subseteq \mathbb{E}$ be a nonempty closed and convex set, and let $\lambda > 0$. Consider the function $f = \frac{1}{2}d_C^2$. Then, by Example 6.53, $f = M_g^1$, where $g = \delta_C$. Recall that $\mathrm{prox}_g = P_C$. Therefore, invoking Corollary 6.64, we obtain that for any $\mathbf{x} \in \mathbb{E}$,

$$\mathrm{prox}_{\lambda f}(\mathbf{x}) = \mathrm{prox}_{\lambda M_g^1}(\mathbf{x}) = \mathbf{x} + \frac{\lambda}{\lambda + 1}\left(\mathrm{prox}_{(\lambda+1)g}(\mathbf{x}) - \mathbf{x} \right) = \mathbf{x} + \frac{\lambda}{\lambda + 1}\left(P_C(\mathbf{x}) - \mathbf{x} \right).$$

To conclude,

$$\mathrm{prox}_{\frac{\lambda}{2} d_C^2}(\mathbf{x}) = \frac{\lambda}{\lambda+1} P_C(\mathbf{x}) + \frac{1}{\lambda+1}\mathbf{x}.$$

∎

Example 6.66 (prox of the Huber function). Consider the function

$$f(\mathbf{x}) = \lambda H_\mu(\mathbf{x}),$$

where H_μ is the Huber function with a smoothing parameter $\mu > 0$ given in (6.43). By Example 6.54, $H_\mu = M_g^\mu$, where $g(\mathbf{x}) = \|\mathbf{x}\|$. Therefore, by Corollary 6.64, it follows that for any $\lambda > 0$ and $\mathbf{x} \in \mathbb{E}$ (recalling the expression for the prox of the Euclidean norm derived in Example 6.19),

$$\mathrm{prox}_{\lambda H_\mu}(\mathbf{x}) = \mathrm{prox}_{\lambda M_g^\mu}(\mathbf{x}) = \mathbf{x} + \frac{\lambda}{\mu+\lambda}\left(\mathrm{prox}_{(\mu+\lambda)g}(\mathbf{x}) - \mathbf{x}\right)$$

$$= \mathbf{x} + \frac{\lambda}{\mu+\lambda}\left(\left(1 - \frac{\mu+\lambda}{\max\{\|\mathbf{x}\|, \mu+\lambda\}}\right)\mathbf{x} - \mathbf{x}\right),$$

which, after some algebraic cancellations, reduces to

$$\mathrm{prox}_{\lambda H_\mu}(\mathbf{x}) = \left(1 - \frac{\lambda}{\max\{\|\mathbf{x}\|, \mu+\lambda\}}\right)\mathbf{x}.$$

∎

Similarly to the Moreau decomposition formula for the prox operator (Theorem 6.45), we can obtain a decomposition formula for the Moreau envelope function.

Theorem 6.67 (Moreau envelope decomposition). *Let $f : \mathbb{E} \to (-\infty, \infty]$ be a proper closed and convex function, and let $\mu > 0$. Then for any $\mathbf{x} \in \mathbb{E}$,*

$$M_f^\mu(\mathbf{x}) + M_{f^*}^{1/\mu}(\mathbf{x}/\mu) = \frac{1}{2\mu}\|\mathbf{x}\|^2.$$

Proof. Recall that for any $\mathbf{x} \in \mathbb{E}$,

$$M_f^\mu(\mathbf{x}) = \min_{\mathbf{u} \in \mathbb{E}}\{f(\mathbf{u}) + \psi(\mathbf{u})\},$$

where $\psi(\mathbf{u}) \equiv \frac{1}{2\mu}\|\mathbf{u} - \mathbf{x}\|^2$. By Fenchel's duality theorem (Theorem 4.15), we have

$$M_f^\mu(\mathbf{x}) = \max_{\mathbf{v} \in \mathbb{E}}\{-f^*(\mathbf{v}) - \psi^*(-\mathbf{v})\} = -\min_{\mathbf{v} \in \mathbb{E}}\{f^*(\mathbf{v}) + \psi^*(-\mathbf{v})\}.$$

Denote $\phi(\cdot) = \frac{1}{2}\|\cdot - \mathbf{x}\|^2$. Then

$$\phi^*(\mathbf{v}) = \frac{1}{2}\|\mathbf{v}\|^2 + \langle \mathbf{x}, \mathbf{v}\rangle.$$

Since $\psi = \frac{1}{\mu}\phi$, it follows by Theorem 4.14 that

$$\psi^*(\mathbf{v}) = \frac{1}{\mu}\phi^*(\mu\mathbf{v}) = \frac{\mu}{2}\|\mathbf{v}\|^2 + \langle\mathbf{x}, \mathbf{v}\rangle.$$

Therefore,

$$M_f^\mu(\mathbf{x}) = -\min_{\mathbf{v}\in\mathbb{E}}\left\{f^*(\mathbf{v}) + \frac{\mu}{2}\|\mathbf{v}\|^2 - \langle\mathbf{x}, \mathbf{v}\rangle\right\},$$

and hence

$$M_f^\mu(\mathbf{x}) = -\min_{\mathbf{v}\in\mathbb{E}}\left\{f^*(\mathbf{v}) + \frac{\mu}{2}\|\mathbf{v} - \mathbf{x}/\mu\|^2 - \frac{1}{2\mu}\|\mathbf{x}\|^2\right\} = \frac{1}{2\mu}\|\mathbf{x}\|^2 - M_{f^*}^{1/\mu}(\mathbf{x}/\mu),$$

establishing the desired result. $\square$

6.8   Miscellaneous Prox Computations

In this section we gather several examples of prox computations that are not linked to any specific result established in this chapter.

6.8.1   Norm of a Linear Transformation over $\mathbb{R}^n$

Lemma 6.68. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|_2$, where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is with full row rank, and let $\lambda > 0$. Then*

$$\text{prox}_{\lambda f}(\mathbf{x}) = \begin{cases} \mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}, & \|(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}\|_2 \leq \lambda, \\ \mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \alpha^*\mathbf{I})^{-1}\mathbf{A}\mathbf{x}, & \|(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}\|_2 > \lambda, \end{cases}$$

where α^* is the unique positive root of the decreasing function

$$g(\alpha) = \|(\mathbf{A}\mathbf{A}^T + \alpha\mathbf{I})^{-1}\mathbf{A}\mathbf{x}\|_2^2 - \lambda^2.$$

Proof. The vector $\text{prox}_{\lambda f}(\mathbf{x})$ is the unique optimal solution to

$$\min_{\mathbf{u}\in\mathbb{R}^n}\left\{\lambda\|\mathbf{A}\mathbf{u}\|_2 + \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2\right\},$$

which can also be rewritten as

$$\min_{\mathbf{u}\in\mathbb{R}^n, \mathbf{z}\in\mathbb{R}^m}\left\{\frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 + \lambda\|\mathbf{z}\|_2 : \mathbf{z} = \mathbf{A}\mathbf{u}\right\}. \tag{6.49}$$

To construct a Lagrangian dual problem, we first form the Lagrangian:

$$\begin{aligned} L(\mathbf{u}, \mathbf{z}; \mathbf{y}) &= \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 + \lambda\|\mathbf{z}\|_2 + \mathbf{y}^T(\mathbf{z} - \mathbf{A}\mathbf{u}) \\ &= \left[\frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 - (\mathbf{A}^T\mathbf{y})^T\mathbf{u}\right] + \left[\lambda\|\mathbf{z}\|_2 + \mathbf{y}^T\mathbf{z}\right]. \end{aligned}$$

Since the Lagrangian is separable w.r.t. $\mathbf{u}$ and $\mathbf{z}$, the dual objective function can be rewritten as

$$\min_{\mathbf{u},\mathbf{z}} L(\mathbf{u},\mathbf{z};\mathbf{y}) = \min_{\mathbf{u}} \left[\frac{1}{2}\|\mathbf{u}-\mathbf{x}\|_2^2 - (\mathbf{A}^T\mathbf{y})^T\mathbf{u}\right] + \min_{\mathbf{z}} \left[\lambda\|\mathbf{z}\|_2 + \mathbf{y}^T\mathbf{z}\right]. \qquad (6.50)$$

The minimizer of the minimization problem in $\mathbf{u}$ is $\tilde{\mathbf{u}} = \mathbf{x} + \mathbf{A}^T\mathbf{y}$ with a corresponding optimal value of

$$\min_{\mathbf{u}} \left[\frac{1}{2}\|\mathbf{u}-\mathbf{x}\|_2^2 - (\mathbf{A}^T\mathbf{y})^T\mathbf{u}\right] = \frac{1}{2}\|\tilde{\mathbf{u}}-\mathbf{x}\|_2^2 - (\mathbf{A}^T\mathbf{y})^T\tilde{\mathbf{u}}$$
$$= -\frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{A}^T\mathbf{y} - (\mathbf{A}\mathbf{x})^T\mathbf{y}. \qquad (6.51)$$

As for the second minimization problem, note that

$$\min_{\mathbf{z}} \left[\lambda\|\mathbf{z}\|_2 + \mathbf{y}^T\mathbf{z}\right] = -\max_{\mathbf{z}}[(-\mathbf{y})^T\mathbf{z} - \lambda\|\mathbf{z}\|_2] = -g^*(-\mathbf{y}),$$

where $g(\cdot) = \lambda\|\cdot\|_2$. Since $g^*(\mathbf{w}) = \lambda\delta_{B_{\|\cdot\|_2}[\mathbf{0},1]}(\mathbf{w}/\lambda) = \delta_{B_{\|\cdot\|_2}[\mathbf{0},\lambda]}$ (see Section 4.4.12 and Theorem 4.14), we can conclude that

$$\min_{\mathbf{z}} \left[\lambda\|\mathbf{z}\|_2 + \mathbf{y}^T\mathbf{z}\right] = \begin{cases} 0, & \|\mathbf{y}\|_2 \leq \lambda, \\ -\infty, & \|\mathbf{y}\|_2 > \lambda. \end{cases}$$

Combining this with (6.51), we obtain the following dual problem:

$$\max_{\mathbf{y}\in\mathbb{R}^m} \left\{-\frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{A}^T\mathbf{y} - (\mathbf{A}\mathbf{x})^T\mathbf{y} : \|\mathbf{y}\|_2 \leq \lambda\right\}. \qquad (6.52)$$

Note that strong duality holds for the primal-dual pair of problems (6.49) and (6.52) (see Theorem A.1). To solve problem (6.52), we will first rewrite it as a minimization problem:

$$\min_{\mathbf{y}\in\mathbb{R}^m} \left\{\frac{1}{2}\mathbf{y}^T\mathbf{A}\mathbf{A}^T\mathbf{y} + (\mathbf{A}\mathbf{x})^T\mathbf{y} : \|\mathbf{y}\|_2^2 \leq \lambda^2\right\}. \qquad (6.53)$$

So far we have shown that

$$\mathrm{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} + \mathbf{A}^T\mathbf{y}, \qquad (6.54)$$

where $\mathbf{y}$ is an optimal solution of problem (6.53). Since problem (6.53) is convex and satisfies Slater's condition, it follows by the KKT conditions that $\mathbf{y}$ is an optimal solution of (6.53) if and only if there exists α^* (optimal dual variable) for which

$$(\mathbf{A}\mathbf{A}^T + \alpha^*\mathbf{I})\mathbf{y} + \mathbf{A}\mathbf{x} = \mathbf{0}, \qquad (6.55)$$
$$\alpha^*(\|\mathbf{y}\|_2^2 - \lambda^2) = 0, \qquad (6.56)$$
$$\|\mathbf{y}\|_2^2 \leq \lambda^2, \qquad (6.57)$$
$$\alpha^* \geq 0. \qquad (6.58)$$

There are two options. In the first, $\alpha^* = 0$, and then by (6.55),

$$\mathbf{y} = -(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}. \qquad (6.59)$$

Since (6.56) and (6.58) are automatically satisfied for $\alpha^* = 0$, we can conclude that $\mathbf{y}$ given in (6.59) is the optimal solution of (6.53) if and only if (6.57) is satisfied, meaning if and only if $\|(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}\|_2 \leq \lambda$. In this case, by (6.54), $\operatorname{prox}_{\lambda f}(\mathbf{x}) = \mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}$.

On the other hand, if $\|(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{x}\|_2 > \lambda$, then $\alpha^* > 0$, and hence by the complementary slackness condition (6.56),

$$\|\mathbf{y}\|_2^2 = \lambda^2. \tag{6.60}$$

By (6.55),

$$\mathbf{y} = -(\mathbf{A}\mathbf{A}^T + \alpha^*\mathbf{I})^{-1}\mathbf{A}\mathbf{x}.$$

Using (6.60), we can conclude that α^* can be uniquely determined as the positive root of the function

$$g(\alpha) = \|(\mathbf{A}\mathbf{A}^T + \alpha\mathbf{I})^{-1}\mathbf{A}\mathbf{x}\|_2^2 - \lambda^2.$$

It is easy to see that g is strictly decreasing for $\alpha \geq 0$, and therefore g has a unique root. $\square$

6.8.2 Squared l_1-Norm

The prox of the l_1-norm has a simple formula. In this section we will show how to compute the prox of the squared l_1-norm—a task that will prove itself to be much more complicated. We will require the following lemma that expresses $\|\mathbf{x}\|_1^2$ as the optimal value of an optimization problem written in terms of the function

$$\varphi(s,t) = \begin{cases} \frac{s^2}{t}, & t > 0, \\ 0, & s = 0, t = 0, \\ \infty & \text{else.} \end{cases} \tag{6.61}$$

By Example 2.32, φ is closed and convex (even though it is not continuous at $(s,t) = (0,0)$).

Lemma 6.69 (variational representation of $\|\cdot\|_1^2$). *For any $\mathbf{x} \in \mathbb{R}^n$ the following holds:*

$$\min_{\boldsymbol{\lambda} \in \Delta_n} \sum_{j=1}^n \varphi(x_j, \lambda_j) = \|\mathbf{x}\|_1^2, \tag{6.62}$$

where φ is defined in (6.61). An optimal solution of the minimization problem in (6.62) is given by

$$\tilde{\lambda}_j = \begin{cases} \frac{|x_j|}{\|\mathbf{x}\|_1}, & \mathbf{x} \neq \mathbf{0}, \\ \frac{1}{n}, & \mathbf{x} = \mathbf{0}, \end{cases} \quad j = 1, 2, \ldots, n. \tag{6.63}$$

Proof. Since problem (6.62) consists of minimizing a closed and convex function (by Example 2.32) over a compact set, then by the Weierstrass theorem for closed

functions (Theorem 2.12), it possesses an optimal solution, which we denote by $\boldsymbol{\lambda}^* \in \Delta_n$. Define

$$I_0 = \{i \in \{1, 2, \ldots, n\} : \lambda_i^* = 0\},$$
$$I_1 = \{i \in \{1, 2, \ldots, n\} : \lambda_i^* > 0\}.$$

By the definitions of I_0 and I_1, we have

$$\sum_{i \in I_1} \lambda_j^* = \sum_{i=1}^n \lambda_j^* = 1. \tag{6.64}$$

It holds that $x_i = 0$ for any $i \in I_0$, since otherwise we will have that $\varphi(x_i, \lambda_i^*) = \infty$, which is a clear contradiction to the optimality of $\boldsymbol{\lambda}^*$. Therefore, using the Cauchy–Schwarz inequality,

$$\sum_{j=1}^n |x_j| = \sum_{j \in I_1} |x_j| = \sum_{j \in I_1} \frac{|x_j|}{\sqrt{\lambda_j^*}} \sqrt{\lambda_j^*} \le \sqrt{\sum_{j \in I_1} \frac{x_j^2}{\lambda_j^*}} \cdot \sqrt{\sum_{j \in I_1} \lambda_j^*} \overset{(6.64)}{=} \sqrt{\sum_{j \in I_1} \frac{x_j^2}{\lambda_j^*}}.$$

We can thus conclude that

$$\sum_{j=1}^n \varphi(x_j, \lambda_j^*) = \sum_{j \in I_1} \varphi(x_j, \lambda_j^*) = \sum_{j \in I_1} \frac{x_j^2}{\lambda_j^*} \ge \|\mathbf{x}\|_1^2. \tag{6.65}$$

On the other hand, since $\boldsymbol{\lambda}^*$ is an optimal solution of the problem in (6.62),

$$\sum_{j=1}^n \varphi(x_j, \lambda_j^*) \le \sum_{j=1}^n \varphi(x_j, \tilde{\lambda}_j) = \|\mathbf{x}\|_1^2, \tag{6.66}$$

where $\tilde{\boldsymbol{\lambda}}$ is given by (6.63). Combining (6.65) and (6.66), we finally conclude that the optimal value of the minimization problem in (6.62) is $\|\mathbf{x}\|_1^2$ and that $\tilde{\boldsymbol{\lambda}}$ is an optimal solution. $\square$

Lemma 6.70 (prox of $\|\cdot\|_1^2$).[34] *Let $f : \mathbb{R}^n \to \mathbb{R}$ be given by $f(\mathbf{x}) = \|\mathbf{x}\|_1^2$, and let $\rho > 0$. Then*

$$\mathrm{prox}_{\rho f}(\mathbf{x}) = \begin{cases} \left(\frac{\lambda_i x_i}{\lambda_i + 2\rho}\right)_{i=1}^n, & \mathbf{x} \ne \mathbf{0}, \\ \mathbf{0}, & \mathbf{x} = \mathbf{0}, \end{cases}$$

where $\lambda_i = \left[\frac{\sqrt{\rho}|x_i|}{\sqrt{\mu^*}} - 2\rho\right]_+$ with μ^* being any positive root of the nonincreasing function

$$\psi(\mu) = \sum_{i=1}^n \left[\frac{\sqrt{\rho}|x_i|}{\sqrt{\mu}} - 2\rho\right]_+ - 1.$$

---

[34] The computation of the prox of the squared l_1-norm is due to Evgeniou, Pontil, Spinellis, and Nassuphis [54].

Proof. If $\mathbf{x} = \mathbf{0}$, then obviously $\operatorname{prox}_{\rho f}(\mathbf{x}) = \operatorname{argmin}_{\mathbf{u}} \left\{ \frac{1}{2}\|\mathbf{u}\|_2^2 + \rho\|\mathbf{u}\|_1^2 \right\} = \mathbf{0}$. Assume that $\mathbf{x} \neq \mathbf{0}$. By Lemma 6.69, $\mathbf{u} = \operatorname{prox}_{\rho f}(\mathbf{x})$ if and only if it is the $\mathbf{u}$-part of the optimal solution of

$$\min_{\mathbf{u} \in \mathbb{R}^n, \boldsymbol{\lambda} \in \Delta_n} \left\{ \frac{1}{2}\|\mathbf{u} - \mathbf{x}\|_2^2 + \rho \sum_{i=1}^n \varphi(u_i, \lambda_i) \right\},$$

where φ is defined in (6.61). Minimizing first with respect to $\mathbf{u}$, we obtain that $u_i = \frac{\lambda_i x_i}{\lambda_i + 2\rho}$, and the problem thus reduces to

$$
\begin{aligned}
\min_{\boldsymbol{\lambda}} \quad & \sum_{i=1}^n \frac{\rho x_i^2}{\lambda_i + 2\rho} \\
\text{s.t.} \quad & \mathbf{e}^T \boldsymbol{\lambda} = 1, \\
& \boldsymbol{\lambda} \geq \mathbf{0}.
\end{aligned}
\tag{6.67}
$$

By Theorem A.1, strong duality holds for problem (6.67) (taking the underlying set as $X = \mathbb{R}_+^n$). Associating a Lagrange multiplier μ to the equality constraint, the Lagrangian is

$$L(\boldsymbol{\lambda}; \mu) = \sum_{i=1}^n \left(\frac{\rho x_i^2}{\lambda_i + 2\rho} + \lambda_i \mu \right) - \mu.$$

By Theorem A.2, $\boldsymbol{\lambda}^*$ is an optimal solution of (6.67) if and only if there exists μ^* for which

$$\boldsymbol{\lambda}^* \in \operatorname{argmin}_{\boldsymbol{\lambda} \geq \mathbf{0}} L(\boldsymbol{\lambda}; \mu^*), \tag{6.68}$$

$$\mathbf{e}^T \boldsymbol{\lambda}^* = 1. \tag{6.69}$$

Since the minimum in (6.68) is finite and attained, and since $\mathbf{x} \neq \mathbf{0}$, it follows that $\mu^* > 0$ (otherwise, if $\mu^* = 0$, the minimum in (6.68) would not be attained). Exploiting the separability of the Lagrangian, it follows that (6.68) is the same as

$$\lambda_i^* = \left[\frac{\sqrt{\rho}|x_i|}{\sqrt{\mu}} - 2\rho \right]_+.$$

The dual optimal variable μ^* is chosen to satisfy (6.69):

$$\sum_{i=1}^n \left[\frac{\sqrt{\rho}|x_i|}{\sqrt{\mu}} - 2\rho \right]_+ = 1. \quad \square$$

6.8.3 Projection onto the Set of s-Sparse Vectors

Let $s \in \{1, 2, \ldots, n\}$ and consider the set

$$C_s = \{ \mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|_0 \leq s \}.$$

The set C_s comprises all *s-sparse vectors*, meaning all vectors with at most s nonzero elements. Obviously C_s is not convex; for example, for $n = 2$, $(0,1)^T, (1,0)^T \in C_1$,

but $(0.5, 0.5)^T = 0.5(0, 1)^T + 0.5(1, 0)^T \notin C_1$. The set C_s is closed as a level set of the closed function $\|\cdot\|_0$ (see Example 2.11). Therefore, by Theorem 6.4, $P_{C_s} = \operatorname{prox}_{\delta_{C_s}}$ is nonempty; however, the nonconvexity of C_s implies that $P_{C_s}(\mathbf{x})$ is not necessarily a singleton.

The set $P_{C_s}(\mathbf{x})$ is described in Lemma 6.71 below. The description requires some additional notation. For a vector $\mathbf{x} \in \mathbb{R}^n$ and a set of indices $S \subseteq \{1, 2, \ldots, n\}$, the vector $\mathbf{x}_S$ is the subvector of $\mathbf{x}$ that corresponds to the indices in S. For example, for $n = 4$, if $\mathbf{x} = (4, 3, 5, -1)^T$, then $\mathbf{x}_{\{1,4\}} = (4, -1)^T, \mathbf{x}_{\{2,3\}} = (3, 5)^T$. For a given indices set $S \subseteq \{1, 2, \ldots, n\}$, the matrix $\mathbf{U}_S$ is the submatrix of the identity matrix $\mathbf{I}_n$ comprising the columns corresponding to the indices in S. For example, for $n = 3$,

$$
\mathbf{U}_{\{1,3\}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \mathbf{U}_{\{2\}} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.
$$

For a given indices set $S \subseteq \{1, 2, \ldots, n\}$, the complement set S^c is given by $S^c = \{1, 2, \ldots, n\} \setminus S$.

Finally, we recall our notation (that was also used in Example 6.51) that for a given $\mathbf{x} \in \mathbb{R}^n$, $x_{\langle i \rangle}$ is the ith largest value among $|x_1|, |x_2|, \ldots, |x_n|$. Therefore, in particular, $|x_{\langle 1 \rangle}| \geq |x_{\langle 2 \rangle}| \geq \cdots \geq |x_{\langle n \rangle}|$. Lemma 6.71 shows that $P_{C_s}(\mathbf{x})$ comprises all vectors consisting of the s components of $\mathbf{x}$ with the largest absolute values and with zeros elsewhere. There may be several choices for the s components with largest absolute values, and this is why $P_{C_s}(\mathbf{x})$ might consist of several vectors. Note that in the statement of the lemma, we characterize the property of an index set S to "comprise s indices corresponding to the s largest absolute values in $\mathbf{x}$" by the relation

$$
S \subseteq \{1, 2, \ldots, n\}, \quad |S| = s, \quad \sum_{i \in S} |x_i| = \sum_{i=1}^{s} |x_{\langle i \rangle}|.
$$

Lemma 6.71 (projection onto C_s). *Let $s \in \{1, 2, \ldots, n\}$ and $\mathbf{x} \in \mathbb{R}^n$. Then*

$$
P_{C_s}(\mathbf{x}) = \left\{ \mathbf{U}_S \mathbf{x}_S : |S| = s, S \subseteq \{1, 2, \ldots, n\}, \sum_{i \in S} |x_i| = \sum_{i=1}^{s} |x_{\langle i \rangle}| \right\}.
$$

Proof. Since C_s consists of all s-sparse vectors, it can be represented as the following union:

$$
C_s = \bigcup_{S \subseteq \{1, 2, \ldots, n\}, |S| = s} A_S,
$$

where $A_S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x}_{S^c} = \mathbf{0}\}$. Therefore,[35]

$$
P_{C_s}(\mathbf{x}) \subseteq \bigcup_{S \subseteq \{1, 2, \ldots, n\}, |S| = s} \{P_{A_S}(\mathbf{x})\}. \tag{6.70}
$$

---

[35]Since A_S is convex, we treat $P_{A_S}(\mathbf{x})$ as a vector and not as a singleton set. The inclusion (6.70) holds since if $B_1, B_2, \ldots, B_m$ are closed convex sets, then $P_{\cup_{i=1}^m B_i}(\mathbf{x}) \subseteq \cup_{i=1}^m \{P_{B_i}(\mathbf{x})\}$ for any $\mathbf{x}$.

The vectors in $P_{C_s}(\mathbf{x})$ will be the vectors $P_{A_S}(\mathbf{x})$ with the smallest possible value of $\|P_{A_S}(\mathbf{x}) - \mathbf{x}\|^2$. The vector $P_{A_S}(\mathbf{x})$ is the optimal solution of the problem

$$\min_{\mathbf{y} \in \mathbb{R}^n} \left\{ \|\mathbf{y} - \mathbf{x}\|_2^2 : \mathbf{y}_{S^c} = \mathbf{0} \right\},$$

which can be rewritten as

$$\min_{\mathbf{y} \in \mathbb{R}^n} \left\{ \|\mathbf{y}_S - \mathbf{x}_S\|_2^2 + \|\mathbf{x}_{S^c}\|_2^2 : \mathbf{y}_{S^c} = \mathbf{0} \right\}.$$

The optimal solution of the above problem is obviously given by $\mathbf{y}_S = \mathbf{x}_S, \mathbf{y}_{S^c} = \mathbf{0}$, that is, $\mathbf{y} = \mathbf{U}_S \mathbf{x}_S$, and the optimal value is $\|\mathbf{x}_{S^c}\|_2^2$. The vectors in $P_{C_s}(\mathbf{x})$ will therefore be of the form $\mathbf{U}_S \mathbf{x}_S$, with indices sets S with cardinality s and with minimal value $\|\mathbf{x}_{S^c}\|_2^2$, which is equivalent to the condition that S consists of s indices corresponding to the s largest absolute values in $\mathbf{x}$. $\square$

Example 6.72. Suppose that $n = 4$. Then

$$P_{C_2}[(2, 3, -2, 1)^T] = \{(2, 3, 0, 0)^T, (0, 3, -2, 0)^T\}.   \blacksquare$$

6.9 Summary of Prox Computations

$f(\mathbf{x})$	$\operatorname{dom}(f)$	$\operatorname{prox}_f(\mathbf{x})$	Assumptions	Reference
$\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x} + c$	$\mathbb{R}^n$	$(\mathbf{A}+\mathbf{I})^{-1}(\mathbf{x}-\mathbf{b})$	$\mathbf{A} \in \mathbb{S}_+^n$, $\mathbf{b} \in \mathbb{R}^n$, $c \in \mathbb{R}$	Section 6.2.3
λx^3	$\mathbb{R}_+$	$\frac{-1+\sqrt{1+12\lambda[x]_+}}{6\lambda}$	$\lambda > 0$	Lemma 6.5
μx	$[0,\alpha] \cap \mathbb{R}$	$\min\{\max\{x-\mu,0\},\alpha\}$	$\mu \in \mathbb{R}$, $\alpha \in [0,\infty]$	Example 6.14
$\lambda\|\mathbf{x}\|$	$\mathbb{E}$	$\left(1-\frac{\lambda}{\max\{\|\mathbf{x}\|,\lambda\}}\right)\mathbf{x}$	$\|\cdot\|$—Euclidean norm, $\lambda > 0$	Example 6.19
$-\lambda\|\mathbf{x}\|$	$\mathbb{E}$	$\left(1+\frac{\lambda}{\|\mathbf{x}\|}\right)\mathbf{x}, \quad \mathbf{x} \neq \mathbf{0},$ $\{\mathbf{u} : \|\mathbf{u}\| = \lambda\}, \quad \mathbf{x} = \mathbf{0}.$	$\|\cdot\|$—Euclidean norm, $\lambda > 0$	Example 6.21
$\lambda\|\mathbf{x}\|_1$	$\mathbb{R}^n$	$\mathcal{T}_\lambda(\mathbf{x}) = [|\mathbf{x}| - \lambda\mathbf{e}]_+ \odot \operatorname{sgn}(\mathbf{x})$	$\lambda > 0$	Example 6.8
$\|\boldsymbol{\omega} \odot \mathbf{x}\|_1$	$\operatorname{Box}[-\boldsymbol{\alpha},\boldsymbol{\alpha}]$	$\mathcal{S}_{\boldsymbol{\omega},\boldsymbol{\alpha}}(\mathbf{x})$	$\boldsymbol{\alpha} \in [0,\infty]^n$, $\boldsymbol{\omega} \in \mathbb{R}_+^n$	Example 6.23
$\lambda\|\mathbf{x}\|_\infty$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_{B_{\|\cdot\|_1}[\mathbf{0},1]}(\mathbf{x}/\lambda)$	$\lambda > 0$	Example 6.48
$\lambda\|\mathbf{x}\|_a$	$\mathbb{E}$	$\mathbf{x} - \lambda P_{B_{\|\cdot\|_{a,*}}[\mathbf{0},1]}(\mathbf{x}/\lambda)$	$\|\mathbf{x}\|_a$—arbitrary norm, $\lambda > 0$	Example 6.47
$\lambda\|\mathbf{x}\|_0$	$\mathbb{R}^n$	$\mathcal{H}_{\sqrt{2\lambda}}(x_1) \times \cdots \times \mathcal{H}_{\sqrt{2\lambda}}(x_n)$	$\lambda > 0$	Example 6.10
$\lambda\|\mathbf{x}\|^3$	$\mathbb{E}$	$\frac{2}{1+\sqrt{1+12\lambda\|\mathbf{x}\|}}\mathbf{x}$	$\|\cdot\|$—Euclidean norm, $\lambda > 0$,	Example 6.20
$-\lambda\sum_{j=1}^n \log x_j$	$\mathbb{R}_{++}^n$	$\left(\frac{x_j+\sqrt{x_j^2+4\lambda}}{2}\right)_{j=1}^n$	$\lambda > 0$	Example 6.9
$\delta_C(\mathbf{x})$	$\mathbb{E}$	$P_C(\mathbf{x})$	$\emptyset \neq C \subseteq \mathbb{E}$	Theorem 6.24
$\lambda\sigma_C(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda)$	$\lambda > 0$, $C \neq \emptyset$ closed convex	Theorem 6.46
$\lambda\max\{x_i\}$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_{\Delta_n}(\mathbf{x}/\lambda)$	$\lambda > 0$	Example 6.49
$\lambda\sum_{i=1}^k x_{[i]}$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda)$, $C = H_{\mathbf{e},k} \cap \operatorname{Box}[\mathbf{0},\mathbf{e}]$	$\lambda > 0$	Example 6.50
$\lambda\sum_{i=1}^k |x_{\langle i\rangle}|$	$\mathbb{R}^n$	$\mathbf{x} - \lambda P_C(\mathbf{x}/\lambda)$, $C = B_{\|\cdot\|_1}[\mathbf{0},k] \cap \operatorname{Box}[-\mathbf{e},\mathbf{e}]$	$\lambda > 0$	Example 6.51
$\lambda M_f^\mu(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x} + \frac{\lambda}{\mu+\lambda}\left(\operatorname{prox}_{(\mu+\lambda)f}(\mathbf{x}) - \mathbf{x}\right)$	$\lambda,\mu > 0$, f proper closed convex	Corollary 6.64
$\lambda d_C(\mathbf{x})$	$\mathbb{E}$	$\mathbf{x} + \min\left\{\frac{\lambda}{d_C(\mathbf{x})},1\right\}(P_C(\mathbf{x}) - \mathbf{x})$	$\emptyset \neq C$ closed convex, $\lambda > 0$	Lemma 6.43
$\frac{\lambda}{2}d_C^2(\mathbf{x})$	$\mathbb{E}$	$\frac{\lambda}{\lambda+1}P_C(\mathbf{x}) + \frac{1}{\lambda+1}\mathbf{x}$	$\emptyset \neq C$ closed convex, $\lambda > 0$	Example 6.65
$\lambda H_\mu(\mathbf{x})$	$\mathbb{E}$	$\left(1-\frac{\lambda}{\max\{\|\mathbf{x}\|,\mu+\lambda\}}\right)\mathbf{x}$	$\lambda,\mu > 0$	Example 6.66
$\rho\|\mathbf{x}\|_1^2$	$\mathbb{R}^n$	$\left(\frac{v_i x_i}{v_i+2\rho}\right)_{i=1}^n$, $\mathbf{v} = \left[\sqrt{\frac{\rho}{\mu}}|\mathbf{x}| - 2\rho\right]_+$, $\mathbf{e}^T\mathbf{v} = 1$ ($\mathbf{0}$ when $\mathbf{x} = \mathbf{0}$)	$\rho > 0$	Lemma 6.70
$\lambda\|\mathbf{A}\mathbf{x}\|_2$	$\mathbb{R}^n$	$\mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T + \alpha^*\mathbf{I})^{-1}\mathbf{A}\mathbf{x}$, $\alpha^* = 0$ if $\|\mathbf{v}_0\|_2 \leq \lambda$; otherwise, $\|\mathbf{v}_{\alpha^*}\|_2 = \lambda$; $\mathbf{v}_\alpha \equiv (\mathbf{A}\mathbf{A}^T + \alpha\mathbf{I})^{-1}\mathbf{A}\mathbf{x}$	$\mathbf{A} \in \mathbb{R}^{m \times n}$ with full row rank, $\lambda > 0$	Lemma 6.68