Chapter 12
Duality

12.1 ▪ Motivation and Definition

The dual problem, which we will formally define later on, can be motivated as a way to find bounds on a given optimization problem. We will begin with an example.

Example 12.1. Consider the problem

$$(\mathrm{P}) \quad \begin{array}{ll} \min & x_1^2 + x_2^2 + 2x_1 \\ \text{s.t.} & x_1 + x_2 = 0. \end{array}$$

Problem (P) is of course not a difficult problem, and it can be solved by reducing it into a one-dimensional problem by eliminating x_2 via the relation $x_2 = -x_1$, thus transforming the objective function to $2x_1^2 + 2x_1$. The (unconstrained) minimizer of the latter function is $x_1 = -\frac{1}{2}$, and the optimal solution is thus $(-\frac{1}{2}, \frac{1}{2})$ with a corresponding optimal value of $f^* = -\frac{1}{2}$.

The theoretical exercise that we wish to make is to find lower bounds on the value of the problem by solving unconstrained problems. For example, the unconstrained problem derived by eliminating the single constraint is

$$(\mathrm{P}_0) \quad \min x_1^2 + x_2^2 + 2x_1.$$

The optimal value of (P_0) is a lower bound on the value of the optimal value of (P). We can write this fact by the following notation:

$$\mathrm{val}(\mathrm{P}_0) \le \mathrm{val}(\mathrm{P}).$$

The optimal solution of the convex problem (P_0) is attained at the stationary point $x_1 = -1, x_2 = 0$ with a corresponding optimal value of -1 (which is indeed a lower bound on f^*).

In order to find other lower bounds, we use the following trick. Take a real number μ and consider the following problem, which is equivalent to problem (P):

$$\begin{array}{ll} \min & x_1^2 + x_2^2 + 2x_1 + \mu(x_1 + x_2) \\ \text{s.t.} & x_1 + x_2 = 0. \end{array} \tag{12.1}$$

Now we can eliminate the equality constraint and obtain the unconstrained problem

$$(\mathrm{P}_\mu) \quad \min x_1^2 + x_2^2 + 2x_1 + \mu(x_1 + x_2).$$

We have that for all $\mu \in \mathbb{R}$

$$\mathrm{val}(\mathrm{P}_\mu) \le \mathrm{val}(\mathrm{P}).$$

The optimal solution of (P_μ) is attained at the stationary point $(x_1, x_2) = \left(-1 - \frac{\mu}{2}, -\frac{\mu}{2}\right)$, and the corresponding optimal value, which we denote by $q(\mu)$, is

$$q(\mu) \equiv \mathrm{val}(\mathrm{P}_\mu) = \left(-1 - \frac{\mu}{2}\right)^2 + \left(-\frac{\mu}{2}\right)^2 + 2\left(-1 - \frac{\mu}{2}\right) + \mu(-1-\mu) = -\frac{\mu^2}{2} - \mu - 1.$$

For example, $q(0) = -1$ is the lower bound obtained by (P_0). What interests us the most is the best (i.e., largest), lower bound obtained by this technique. The best lower bound is the solution of the problem

$$(\mathrm{D}) \quad \max\{q(\mu) : \mu \in \mathbb{R}\}.$$

This problem will be called *the dual problem*, and by its construction, its optimal value is a lower bound on the optimal value of the original problem (P), which we will call *the primal problem*:

$$\mathrm{val}(\mathrm{D}) \le \mathrm{val}(\mathrm{P}).$$

In this case the optimal solution of the dual problem is attained at $\mu = -1$, and the corresponding optimal value of the dual problem is $-\frac{1}{2}$, which is exactly f^*, meaning that the best lower bound obtained by the this technique is actually *equal* to the optimal value f^*. Later on, we will refer to this property as "strong duality" and discuss the conditions under which it holds. ∎

12.1.1 ▪ Definition of the Dual Problem

Consider the general model

$$\begin{array}{rll} f^* = \min & f(\mathbf{x}) & \\ \text{s.t.} & g_i(\mathbf{x}) \le 0, & i = 1, 2, \ldots, m, \\ & h_j(\mathbf{x}) = 0, & j = 1, 2, \ldots, p, \\ & \mathbf{x} \in X, & \end{array} \tag{12.2}$$

where $f, g_i, h_j (i = 1, 2, \ldots, m, j = 1, 2, \ldots, p)$ are functions defined on the set $X \subseteq \mathbb{R}^n$. Problem (12.2) will be referred to as the primal problem. At this point, we do not assume anything on the functions (they are not even assumed to be continuous). The Lagrangian of the problem is

$$L(\mathbf{x}, \lambda, \mu) = f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^{p} \mu_j h_j(\mathbf{x}) \quad (\mathbf{x} \in X, \lambda \in \mathbb{R}^m_+, \mu \in \mathbb{R}^p),$$

where $\lambda_1, \lambda_2, \ldots, \lambda_m$ are nonnegative Lagrange multipliers associated with the inequality constraints, and $\mu_1, \mu_2, \ldots, \mu_p$ are the Lagrange multipliers associated with the equality constraints. The *dual objective function* $q : \mathbb{R}^m_+ \times \mathbb{R}^p \to \mathbb{R} \cup \{-\infty\}$ is defined to be

$$q(\lambda, \mu) = \min_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \mu). \tag{12.3}$$

Note that we use the "min" notation even though the minimum is not necessarily attained. In addition, the optimal value of the minimization problem in (12.3) is not always finite;

there may be values of (λ,μ) for which $q(\lambda,\mu)=-\infty$. It is therefore natural to define the domain of the dual objective function as

$$\text{dom}(q)=\{(\lambda,\mu)\in\mathbb{R}^m_+\times\mathbb{R}^p : q(\lambda,\mu)>-\infty\}.$$

The *dual problem* is given by

$$\begin{array}{rll} q^* = & \max & q(\lambda,\mu) \\ & \text{s.t.} & (\lambda,\mu)\in\text{dom}(q). \end{array} \tag{12.4}$$

We begin by showing that the dual problem is always convex; it consists of maximizing a concave function over a convex feasible set.

Theorem 12.2 (convexity of the dual problem). *Consider problem* (12.2) *with f,g_i,h_j ($i=1,2,\ldots,m, j=1,2,\ldots,p$) being functions defined on the set $X\subseteq\mathbb{R}^n$, and let q be the function defined in* (12.3). *Then*

(a) $\text{dom}(q)$ *is a convex set,*

(b) q *is a concave function over* $\text{dom}(q)$.

Proof. (a) To establish the convexity of $\text{dom}(q)$, take $(\lambda_1,\mu_1),(\lambda_2,\mu_2)\in\text{dom}(q)$ and $\alpha\in[0,1]$. Then by the definition of $\text{dom}(q)$ we have that

$$\min_{\mathbf{x}\in X} L(\mathbf{x},\lambda_1,\mu_1)>-\infty, \tag{12.5}$$

$$\min_{\mathbf{x}\in X} L(\mathbf{x},\lambda_2,\mu_2)>-\infty. \tag{12.6}$$

Therefore, since the Lagrangian $L(\mathbf{x},\lambda,\mu)$ is affine with respect to λ,μ, we obtain that

$$\begin{aligned} q(\alpha\lambda_1+(1-\alpha)\lambda_2,\alpha\mu_1+(1-\alpha)\mu_2) &= \min_{\mathbf{x}\in X} L(\mathbf{x},\alpha\lambda_1+(1-\alpha)\lambda_2,\alpha\mu_1+(1-\alpha)\mu_2) \\ &= \min_{\mathbf{x}\in X}[\alpha L(\mathbf{x},\lambda_1,\mu_1)+(1-\alpha)L(\mathbf{x},\lambda_2,\mu_2)] \\ &\geq \alpha\min_{\mathbf{x}\in X} L(\mathbf{x},\lambda_1,\mu_1)+(1-\alpha)\min_{\mathbf{x}\in X} L(\mathbf{x},\lambda_2,\mu_2) \\ &= \alpha q(\lambda_1,\mu_1)+(1-\alpha)q(\lambda_2,\mu_2) \\ &> -\infty. \end{aligned}$$

Hence, $\alpha(\lambda_1,\mu_1)+(1-\alpha)(\lambda_2,\mu_2)\in\text{dom}(q)$, and the convexity of $\text{dom}(q)$ is established.

(b) As noted in the proof of part (a), $L(\mathbf{x},\lambda,\mu)$ is an affine function with respect to (λ,μ). In particular, it is a concave function with respect to (λ,μ). Hence, since $q(\lambda,\mu)$ is the minimum of concave functions, it must be concave. This follows immediately from the fact that the maximum of convex functions is a convex function (Theorem 7.38). ◻

Note that $-q$ is in fact an extended real-valued convex function over $\mathbb{R}^m_+\times\mathbb{R}^p$ as defined in Section 7.7, and the effective domain of $-q$ is exactly the domain defined in this section. The first important result is closely connected to the motivation of the construction of the dual problem: the optimal dual value is a lower bound on the optimal primal value. This result is called the *weak duality theorem*, and unsurprisingly, its proof is rather simple.

Theorem 12.3 (weak duality theorem). *Consider the primal problem* (12.2) *and its dual problem* (12.4). *Then*

$$q^* \le f^*,$$

where q^*, f^* are the optimal dual and primal values respectively.

Proof. Let us denote the feasible set of the primal problem by

$$S = \{\mathbf{x} \in X : g_i(\mathbf{x}) \le 0, h_j(\mathbf{x}) = 0, i = 1,2,\ldots,m, j = 1,2,\ldots,p\}.$$

Then for any $(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$ we have

$$\begin{aligned} q(\lambda, \mu) &= \min_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \mu) \\ &\le \min_{\mathbf{x} \in S} L(\mathbf{x}, \lambda, \mu) \\ &= \min_{\mathbf{x} \in S} \left[f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^p \mu_j h_j(\mathbf{x}) \right] \\ &\le \min_{\mathbf{x} \in S} f(\mathbf{x}), \end{aligned}$$

where the last inequality follows from the fact that $\lambda_i \ge 0$ and for any $\mathbf{x} \in S$, $g_i(\mathbf{x}) \le 0, h_j(\mathbf{x}) = 0$ $(i = 1,2,\ldots,m, j = 1,2,\ldots,p)$. We thus obtain that

$$q(\lambda, \mu) \le \min_{\mathbf{x} \in S} f(\mathbf{x})$$

for any $(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$. By taking the maximum over $(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p$, the result follows. □

The weak duality theorem states that the dual optimal value is a lower bound on the primal optimal value. Example 12.1 illustrated that the lower bound can be tight. However, the lower bound does not have to be tight, and the next example shows that it can be totally uninformative.

Example 12.4. Consider the problem

$$\begin{array}{ll} \min & x_1^2 - 3x_2^2 \\ \text{s.t.} & x_1 = x_2^3. \end{array}$$

It is not difficult to solve the problem. Substituting $x_1 = x_2^3$ into the objective function, we obtain that the problem is equivalent to the unconstrained one-dimensional minimization problem

$$\min_{x_2} x_2^6 - 3x_2^2.$$

An optimal solution exists since the objective function is coercive. The stationary points of the latter problem are the solutions to

$$6x_2^5 - 6x_2 = 0,$$

that is,

$$6x_2(x_2^4 - 1) = 0.$$

Hence, the stationary points are $x_2 = 0, \pm 1$, and thence the only candidates for the optimal solutions are $(0,0), (1,1), (-1,-1)$. Comparing the corresponding objective function values, it follows that the optimal solutions of the problem are $(1,1),(-1,-1)$ with an optimal value of $f^* = -2$.

Let us construct the dual problem. The Lagrangian function is

$$L(x_1, x_2, \mu) = x_1^2 - 3x_2^2 + \mu(x_1 - x_2^3) = x_1^2 + \mu x_1 - 3x_2^2 - \mu x_2^3.$$

Obviously, for any $\mu \in \mathbb{R}$

$$\min_{x_1, x_2} L(x_1, x_2, \mu) = -\infty,$$

and hence the dual optimal value is $q^* = -\infty$, which is an extremely poor lower bound on the primal optimal value $f^* = -2$. $\blacksquare$

12.2 ▪ Strong Duality in the Convex Case

In the convex case we can prove under rather mild conditions that strong duality holds; that is, the primal and dual optimal values coincide. Similarly to the derivation of the KKT conditions, we will rely on separation theorems in order to establish the result. The strict separation theorem (Theorem 10.1) from Section 10.1 states that a point can be strictly separated from any closed and convex set. We will require a variation of this result stating that a point can be separated from any convex set, not necessarily closed. Note that the separation is not strict, and in fact it also includes the case in which the point is on the boundary of the convex set and the theorem is hence called *the supporting hyperplane theorem*. Although the theorem holds for any convex set C, we will state and prove it only for convex sets with a nonempty interior.

Theorem 12.5 (supporting hyperplane theorem). *Let $C \subseteq \mathbb{R}^n$ be a convex set with a nonempty interior and let $\mathbf{y} \notin C$. Then there exists $\mathbf{0} \neq \mathbf{p} \in \mathbb{R}^n$ such that*

$$\mathbf{p}^T \mathbf{x} \le \mathbf{p}^T \mathbf{y} \text{ for any } \mathbf{x} \in C.$$

Proof. Since $\mathbf{y} \notin \text{int}(C)$, it follows that $\mathbf{y} \notin \text{int}(\text{cl}(C))$. (Recall that by Lemma 6.30 $\text{int}(C) = \text{int}(\text{cl}(C))$.) Therefore, there exists a sequence $\{\mathbf{y}_k\}_{k\ge 1}$ satisfying $\mathbf{y}_k \notin \text{cl}(C)$ such that $\mathbf{y}_k \to \mathbf{y}$. Since $\text{cl}(C)$ is convex by Theorem 6.27 and closed by its definition, it follows by the strict separation theorem (Theorem 10.1) that there exists $\mathbf{0} \neq \mathbf{p}_k \in \mathbb{R}^n$ such that

$$\mathbf{p}_k^T \mathbf{x} < \mathbf{p}_k^T \mathbf{y}_k$$

for all $\mathbf{x} \in \text{cl}(C)$. Dividing the latter inequality by $\|\mathbf{p}_k\| \neq 0$, we obtain

$$\frac{\mathbf{p}_k^T}{\|\mathbf{p}_k\|}(\mathbf{x} - \mathbf{y}_k) < 0 \text{ for any } \mathbf{x} \in \text{cl}(C). \tag{12.7}$$

Since the sequence $\{\frac{\mathbf{p}_k}{\|\mathbf{p}_k\|}\}_{k\ge 1}$ is bounded, it follows that there exists a subsequence $\{\frac{\mathbf{p}_k}{\|\mathbf{p}_k\|}\}_{k\in T}$ such that $\frac{\mathbf{p}_k}{\|\mathbf{p}_k\|} \to \mathbf{p}$ as $k \xrightarrow{T} \infty$ for some $\mathbf{p} \in \mathbb{R}^n$. Obviously, $\|\mathbf{p}\| = 1$ and hence in particular $\mathbf{p} \neq \mathbf{0}$. Taking the limit as $k \xrightarrow{T} \infty$ in inequality (12.7), we obtain that

$$\mathbf{p}^T(\mathbf{x} - \mathbf{y}) \le 0 \text{ for any } \mathbf{x} \in \text{cl}(C),$$

which readily implies the result since $C \subseteq \text{cl}(C)$. $\square$

We can now deduce a separation theorem between two disjoint convex sets.

Theorem 12.6 (separation of two convex sets). *Let $C_1, C_2 \subseteq \mathbb{R}^n$ be two convex sets with nonempty interiors such that $C_1 \cap C_2 = \emptyset$. Then there exists $0 \neq \mathbf{p} \in \mathbb{R}^n$ for which*

$$\mathbf{p}^T\mathbf{x} \leq \mathbf{p}^T\mathbf{y} \text{ for any } \mathbf{x} \in C_1, \mathbf{y} \in C_2.$$

Proof. The set $C_1 - C_2$ is a convex set (by part (a) of Theorem 6.8) with a nonempty interior, and since $C_1 \cap C_2 = \emptyset$, it follows that $0 \notin C_1 - C_2$. By the supporting hyperplane theorem (Theorem 12.5), it follows that there exists $0 \neq \mathbf{p} \in \mathbb{R}^n$ such that

$$\mathbf{p}^T(\mathbf{x} - \mathbf{y}) \leq \mathbf{p}^T 0 = 0 \text{ for any } \mathbf{x} \in C_1, \mathbf{y} \in C_2,$$

which is the same as the desired result. $\square$

We will now derive a result which is a nonlinear version of Farkas' lemma. The main difference is that a Slater-type condition must be assumed. Later on, this lemma will be the key in proving the strong duality result.

Theorem 12.7 (nonlinear Farkas' lemma). *Let $X \subseteq \mathbb{R}^n$ be a convex set and let $f, g_1, g_2, \ldots, g_m$ be convex functions over X. Assume that there exists $\hat{\mathbf{x}} \in X$ such that*

$$g_1(\hat{\mathbf{x}}) < 0, \quad g_2(\hat{\mathbf{x}}) < 0, \ldots, g_m(\hat{\mathbf{x}}) < 0.$$

Let $c \in \mathbb{R}$. Then the following two claims are equivalent.

(a) *The following implication holds:*

$$\mathbf{x} \in X, \quad g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m \Rightarrow f(\mathbf{x}) \geq c.$$

(b) *There exist $\lambda_1, \lambda_2, \ldots, \lambda_m \geq 0$ such that*

$$\min_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) \right\} \geq c. \tag{12.8}$$

Proof. The implication (b) $\Rightarrow$ (a) is rather straightforward. Indeed, suppose that there exist $\lambda_1, \lambda_2, \ldots, \lambda_m \geq 0$ such that (12.8) holds, and let $\mathbf{x} \in X$ satisfy $g_i(\mathbf{x}) \leq 0, i = 1, 2, \ldots, m$. Then by (12.8) we have that

$$f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) \geq c,$$

and hence, since $g_i(\mathbf{x}) \leq 0, \lambda_i \geq 0$,

$$f(\mathbf{x}) \geq c - \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) \geq c.$$

To prove that (a) $\Rightarrow$ (b), let us assume that the implication (a) holds. Consider the following two sets:

$$S = \{\mathbf{u} = (u_0, u_1, \ldots, u_m) : \exists \mathbf{x} \in X \text{ s.t. } f(\mathbf{x}) \leq u_0, g_i(\mathbf{x}) \leq u_i, i = 1, 2, \ldots, m\},$$
$$T = \{(u_0, u_1, \ldots, u_m) : u_0 < c, u_1 \leq 0, u_2 \leq 0, \ldots, u_m \leq 0\}.$$

Note that S,T are convex with nonempty interiors and in addition, by the validity of implication (a), $S \cap T = \emptyset$. Therefore, by Theorem 12.6 (separation of two convex sets), it follows that there exists a vector $\mathbf{a} = (a_0, a_1, \ldots, a_m) \neq 0$, such that

$$\min_{(u_0,u_1,\ldots,u_m)\in S} \sum_{j=0}^{m} a_j u_j \geq \max_{(u_0,u_1,\ldots,u_m)\in T} \sum_{j=0}^{m} a_j u_j. \tag{12.9}$$

First note that $\mathbf{a} \geq 0$. This is due to the fact that if there was a negative component, say $a_i < 0$, then by taking u_i to be a negative number tending to $-\infty$ while fixing all the other components as zeros, we obtain that the right-hand-side maximum in (12.9) is ∞, which is impossible. Since $\mathbf{a} \geq 0$, it follows that the right-hand side is $a_0 c$, and we thus obtain

$$\min_{(u_0,u_1,\ldots,u_m)\in S} \sum_{j=0}^{m} a_j u_j \geq a_0 c. \tag{12.10}$$

Now we will show that $a_0 > 0$. Suppose in contradiction that $a_0 = 0$. Then

$$\min_{(u_0,u_1,\ldots,u_m)\in S} \sum_{j=1}^{m} a_j u_j \geq 0.$$

However, since we can take $u_i = g_i(\hat{\mathbf{x}}), i = 1,2,\ldots,m$, we can deduce that

$$\sum_{j=1}^{m} a_j g_j(\hat{\mathbf{x}}) \geq 0,$$

which is impossible since $g_j(\hat{\mathbf{x}}) < 0$ for all j, and there exists at least one nonzero component in $(a_1, a_2, \ldots, a_m)$. Since $a_0 > 0$, we can divide (12.10) by a_0 to obtain

$$\min_{(u_0,u_1,\ldots,u_m)\in S} \left\{ u_0 + \sum_{j=1}^{m} \tilde{a}_j u_j \right\} \geq c, \tag{12.11}$$

where $\tilde{a}_j = \frac{a_j}{a_0}$. Define

$$\tilde{S} = \{\mathbf{u} = (u_0, u_1, \ldots, u_m) : \exists \mathbf{x} \in X \text{ s.t. } f(\mathbf{x}) = u_0, g_i(\mathbf{x}) = u_i, i = 1,2,\ldots,m\}.$$

Then obviously $\tilde{S} \subseteq S$. Therefore,

$$\begin{aligned}\min_{(u_0,u_1,\ldots,u_m)\in S} \left\{ u_0 + \sum_{j=1}^{m} \tilde{a}_j u_j \right\} &\leq \min_{(u_0,u_1,\ldots,u_m)\in \tilde{S}} \left\{ u_0 + \sum_{j=1}^{m} \tilde{a}_j u_j \right\} \\ &= \min_{\mathbf{x}\in X} \left\{ f(\mathbf{x}) + \sum_{j=1}^{m} \tilde{a}_j g_j(\mathbf{x}) \right\},\end{aligned}$$

which combined with (12.11) yields the desired result

$$\min_{\mathbf{x}\in X} \left\{ f(\mathbf{x}) + \sum_{j=1}^{m} \tilde{a}_j g_j(\mathbf{x}) \right\} \geq c. \quad \square$$

We are now able to show a strong duality result in the convex case under a Slater-type condition.

Theorem 12.8 (strong duality of convex problems with inequality constraints). *Consider the optimization problem*

$$\begin{aligned} f^* = \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \le 0, \quad i = 1, 2, \ldots, m, \\ & \mathbf{x} \in X, \end{aligned} \tag{12.12}$$

where X is a convex set and $f, g_i, i = 1, 2, \ldots, m$, are convex functions over X and suppose that there exists $\hat{\mathbf{x}} \in X$ for which $g_i(\hat{\mathbf{x}}) < 0, i = 1, 2, \ldots, m$ and suppose that problem (12.12) *has a finite optimal value. Then the optimal value of the dual problem*

$$q^* = \max\{q(\lambda) : \lambda \in \operatorname{dom}(q)\}, \tag{12.13}$$

where

$$q(\lambda) = \min_{\mathbf{x} \in X} L(\mathbf{x}, \lambda),$$

is attained, and the optimal values of the primal and dual problems are the same:

$$f^* = q^*.$$

Proof. Since $f^* > -\infty$ is the optimal value of (12.12), it follows that the following implication holds:

$$\mathbf{x} \in X, \quad g_i(\mathbf{x}) \le 0, \quad i = 1, 2, \ldots, m \Rightarrow f(\mathbf{x}) \ge f^*,$$

and hence by the nonlinear Farkas' lemma (Theorem 12.7) we have that there exist $\tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_m \ge 0$ such that

$$q(\tilde{\lambda}) = \min_{\mathbf{x} \in X} \left\{ f(\mathbf{x}) + \sum_{j=1}^{m} \tilde{\lambda}_j g_j(\mathbf{x}) \right\} \ge f^*,$$

which combined with the weak duality theorem (Theorem 12.3) yields

$$q^* \ge q(\tilde{\lambda}) \ge f^* \ge q^*.$$

Hence $f^* = q^*$ and $\tilde{\lambda}$ is an optimal solution of the dual problem. □

Example 12.9. Consider the problem

$$\begin{aligned} \min \quad & x_1^2 - x_2 \\ \text{s.t.} \quad & x_2^2 \le 0. \end{aligned}$$

The problem is convex but does not satisfy Slater's condition. The optimal solution is obviously $x_1 = x_2 = 0$ and hence $f^* = 0$. The Lagrangian is

$$L(x_1, x_2, \lambda) = x_1^2 - x_2 + \lambda x_2^2 \quad (\lambda \ge 0),$$

and the dual objective function is

$$q(\lambda) = \min_{x_1, x_2} L(x_1, x_2, \lambda) = \begin{cases} -\infty, & \lambda = 0, \\ -\frac{1}{4\lambda}, & \lambda > 0. \end{cases}$$

The dual problem is therefore

$$\max\left\{-\frac{1}{4\lambda} : \lambda > 0\right\}.$$

The dual optimal value is $q^* = 0$, so we do have the equality $f^* = q^*$. However, the strong duality theorem (Theorem 12.8) states that there exists an optimal solution to the dual problem, and this is obviously not the case in this example. The reason why this property does not hold is that in this example Slater's condition is not satisfied—there does not exist $\bar{x}_2$ for which $\bar{x}_2^2 < 0$. ∎

Example 12.10. ([9]) Consider the convex optimization problem

$$\begin{array}{ll} \min & e^{-x_2} \\ \text{s.t.} & \sqrt{x_1^2 + x_2^2} - x_1 \leq 0. \end{array}$$

Note that the feasible set is in fact

$$F = \{(x_1, x_2) : x_1 \geq 0, x_2 = 0\}.$$

The constraint is always satisfied as an equality constraint, and thus Slater's condition is not satisfied. Since x_2 is necessarily 0, it follows that the optimal value is $f^* = 1$. The Lagrangian of the problem is

$$L(x_1, x_2, \lambda) = e^{-x_2} + \lambda(\sqrt{x_1^2 + x_2^2} - x_1) \quad (\lambda \geq 0).$$

The dual objective function is

$$q(\lambda) = \min_{x_1, x_2} L(x_1, x_2, \lambda).$$

We will show that this minimum is 0, no matter what the value of λ is. First of all, $L(x_1, x_2, \lambda) \geq 0$ for all x_1, x_2 and hence $q(\lambda) \geq 0$. On the other hand, for any $\varepsilon > 0$, if we take $x_2 = -\ln \varepsilon, x_1 = \frac{x_2^2 - \varepsilon^2}{2\varepsilon}$, we have

$$\begin{aligned} \sqrt{x_1^2 + x_2^2} - x_1 &= \sqrt{\frac{(x_2^2 - \varepsilon^2)^2}{4\varepsilon^2} + x_2^2} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} \\ &= \sqrt{\frac{(x_2^2 + \varepsilon^2)^2}{4\varepsilon^2}} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} \\ &= \frac{x_2^2 + \varepsilon^2}{2\varepsilon} - \frac{x_2^2 - \varepsilon^2}{2\varepsilon} \\ &= \varepsilon. \end{aligned}$$

Therefore,

$$L(x_1, x_2, \lambda) = e^{-x_2} + \lambda(\sqrt{x_1^2 + x_2^2} - x_1) = \varepsilon + \lambda\varepsilon = (1 + \lambda)\varepsilon.$$

Consequently, $q(\lambda) = 0$ for all $\lambda \geq 0$. The dual problem is therefore the following "trivial" problem:

$$\max\{0 : \lambda \geq 0\},$$

whose optimal value is obviously $q^* = 0$. We thus obtained that there exists a duality gap $f^* - q^* = 1$, which is a result of the fact that Slater's condition is not satisfied. ∎

We can also derive the complementary slackness conditions under the sole assumption that $q^* = f^*$ (without any convexity assumptions).

Theorem 12.11 (complementary slackness conditions). *Consider problem* (12.12) *and assume that $f^* = q^*$, where q^* is the optimal value of the dual problem given by* (12.13)*. If $\mathbf{x}^*, \lambda^*$ are optimal solutions of the primal and dual problems respectively, then*

$$\mathbf{x}^* \in \operatorname{argmin} L(\mathbf{x}, \lambda^*),$$
$$\lambda_i^* g_i(\mathbf{x}^*) = 0, i = 1, 2, \ldots, m.$$

Proof. We have

$$q^* = q(\lambda^*) = \min_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*) \le L(\mathbf{x}^*, \lambda^*) = f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}^*) \le f(\mathbf{x}^*) = f^*,$$

where the last inequality follows from the fact that $\lambda_i^* \ge 0, g_i(\mathbf{x}^*) \le 0$. Therefore, since $q^* = f^*$, it follows that all the inequalities in the above chain of inequalities and equalities are satisfied as equalities, meaning that $\mathbf{x}^* \in \operatorname{argmin}_{\mathbf{x} \in X} L(\mathbf{x}, \lambda^*)$ and that $\sum_{i=1}^m \lambda_i^* g_i(\mathbf{x}^*) = 0$, which by the fact that $\lambda_i^* \ge 0, g_i(\mathbf{x}^*) \le 0$, implies that $\lambda_i^* g_i(\mathbf{x}^*) = 0$ for all $i = 1, 2, \ldots, m$. □

Finer analysis can show, for example, the following strong duality theorem that deals with linear equality and inequality constraints as well as nonlinear constraints.

Theorem 12.12. *Consider the optimization problem*

$$\begin{array}{rlll} f^* = \min & f(\mathbf{x}) & & \\ \text{s.t.} & g_i(\mathbf{x}) \le 0, & i = 1, 2, \ldots, m, & \\ & h_j(\mathbf{x}) \le 0, & j = 1, 2, \ldots, p, & \\ & s_k(\mathbf{x}) = 0, & k = 1, 2, \ldots, q, & \\ & \mathbf{x} \in X, & & \end{array} \tag{12.14}$$

where X is a convex set and $f, g_i, i = 1, 2, \ldots, m$, are convex functions over X. The functions $h_j, s_k, j = 1, 2, \ldots, p, k = 1, 2, \ldots, q$, are affine functions. Suppose that there exists $\hat{\mathbf{x}} \in \operatorname{int}(X)$ for which $g_i(\hat{\mathbf{x}}) < 0, i = 1, 2, \ldots, m, h_j(\hat{\mathbf{x}}) \le 0, j = 1, 2, \ldots, p, s_k(\hat{\mathbf{x}}) = 0, k = 1, 2, \ldots, q$. Then if problem (12.14) *has a finite optimal value, then the optimal value of the dual problem*

$$q^* = \max\{q(\lambda, \eta, \mu) : (\lambda, \eta, \mu) \in \operatorname{dom}(q)\},$$

where $q : \mathbb{R}_+^m \times \mathbb{R}_+^p \times \mathbb{R}^q \to \mathbb{R} \cup \{-\infty\}$ is given by

$$q(\lambda, \eta, \mu) = \min_{\mathbf{x} \in X} L(\mathbf{x}, \lambda, \eta, \mu) = \min_{\mathbf{x} \in X} \left[f(\mathbf{x}) + \sum_{i=1}^m \lambda_i g_i(\mathbf{x}) + \sum_{j=1}^p \eta_j h_j(\mathbf{x}) + \sum_{k=1}^q \mu_k s_k(\mathbf{x}) \right]$$

is attained, and the optimal values of the primal and dual problems are the same:

$$f^* = q^*.$$

Example 12.13. In this example we will demonstrate the fact that there is no "one" dual problem for a given primal problem, and in many cases there are several ways to construct the dual, which result in different dual problems and dual optimal values. Consider the simple two-dimensional problem

$$\begin{array}{ll} \min & x_1^3 + x_2^3 \\ \text{s.t.} & x_1 + x_2 \geq 1, \\ & x_1, x_2 \geq 0. \end{array}$$

It is not difficult to verify (for example, by finding all the KKT points) that the optimal solution of the problem is $(x_1, x_2) = (\frac{1}{2}, \frac{1}{2})$ and the optimal primal value is thus $f^* = (\frac{1}{2})^3 + (\frac{1}{2})^3 = \frac{1}{4}$. We will consider two possible options for constructing a dual problem. If we take the underlying set as

$$X = \{(x_1, x_2) : x_1, x_2 \geq 0\},$$

then the primal problem can be written as

$$\begin{array}{ll} \min & x_1^3 + x_2^3 \\ \text{s.t.} & x_1 + x_2 \geq 1, \\ & (x_1, x_2) \in X. \end{array}$$

Since the objective function is convex over X, it follows that this problem is in fact a convex optimization problem. Therefore, since Slater's condition is satisfied here (e.g., $(x_1, x_2) = (1, 1) \in X$ satisfy $x_1 + x_2 > 1$), we conclude from Theorem 12.8 that strong duality will hold. The dual problem in this case is constructed by associating a single dual variable λ to the linear inequality constraint $x_1 + x_2 \geq 1$. A second option for writing a dual problem is to take the underlying set X as the entire two-dimensional space and associate Lagrange multipliers with each of the three linear constraints. In this case, the assumptions of the strong duality theorem do not hold since $x_1^3 + x_2^3$ is not a convex function over $\mathbb{R}^2$. The Lagrangian is therefore $(\lambda, \eta_1, \eta_2 \in \mathbb{R}_+)$

$$L(x_1, x_2, \lambda, \eta_1, \eta_2) = x_1^3 + x_2^3 - \lambda(x_1 + x_2 - 1) - \eta_1 x_1 - \eta_2 x_2.$$

Since the minimization of a cubic function over the real line is always $-\infty$, it follows that

$$\begin{aligned} \min_{x_1, x_2} L(x_1, x_2, \lambda, \eta_1, \eta_2) &= \min_{x_1} \left[x_1^3 - (\lambda + \eta_1) x_1 \right] + \min_{x_2} \left[x_2^3 - (\lambda + \eta_2) x_2 \right] + \lambda \\ &= -\infty - \infty + \lambda = -\infty. \end{aligned}$$

Hence, $q^* = -\infty$ and the duality gap is infinite. The conclusion is that the way the dual problem is constructed is extremely important and may result in very different duality gaps. ∎

12.3 ▪ Examples

12.3.1 ▪ Linear Programming

Consider the linear programming problem

$$\begin{array}{ll} \min & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \end{array}$$

where $\mathbf{c} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m\times n}$, and $\mathbf{b} \in \mathbb{R}^m$. We will assume that the problem is feasible, and under this condition, Slater's condition given in Theorem 12.12 is satisfied so that strong duality holds. The Lagrangian is ($\lambda \geq \mathbf{0}$)

$$L(\mathbf{x}, \lambda) = \mathbf{c}^T\mathbf{x} + \lambda^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = (\mathbf{c} + \mathbf{A}^T\lambda)^T\mathbf{x} - \mathbf{b}^T\lambda,$$

and the dual objective function is given by

$$q(\lambda) = \min_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{x}, \lambda) = \min_{\mathbf{x}\in\mathbb{R}^n}(\mathbf{c} + \mathbf{A}^T\lambda)^T\mathbf{x} - \mathbf{b}^T\lambda = \begin{cases} -\mathbf{b}^T\lambda, & \mathbf{c} + \mathbf{A}^T\lambda = \mathbf{0}, \\ -\infty & \text{else.} \end{cases}$$

Therefore, the dual problem is

$$\begin{array}{ll} \max & -\mathbf{b}^T\lambda \\ \text{s.t.} & \mathbf{A}^T\lambda = -\mathbf{c}, \\ & \lambda \geq \mathbf{0}. \end{array}$$

As already mentioned, Slater's condition is satisfied if the primal problem is feasible and under the assumption that the optimal value is finite, strong duality holds.

12.3.2 ▪ Strictly Convex Quadratic Programming

Consider the following general strictly convex quadratic programming problem

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{f}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \end{array} \tag{12.15}$$

where $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is positive definite, $\mathbf{f} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m\times n}$, and $\mathbf{b} \in \mathbb{R}^m$. The Lagrangian of the problem is

$$(\lambda \in \mathbb{R}^m_+) \quad L(\mathbf{x}, \lambda) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2\mathbf{f}^T\mathbf{x} + 2\lambda^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = \mathbf{x}^T\mathbf{Q}\mathbf{x} + 2(\mathbf{A}^T\lambda + \mathbf{f})^T\mathbf{x} - 2\mathbf{b}^T\lambda.$$

To find the dual objective function we need to minimize the Lagrangian with respect to $\mathbf{x}$. The minimizer of the Lagrangian is attained at the stationary point of the Lagrangian which is the solution to

$$\nabla_{\mathbf{x}} L(\mathbf{x}^*, \lambda) = 2\mathbf{Q}\mathbf{x}^* + 2(\mathbf{A}^T\lambda + \mathbf{f}) = \mathbf{0},$$

and hence

$$\mathbf{x}^* = -\mathbf{Q}^{-1}(\mathbf{f} + \mathbf{A}^T\lambda). \tag{12.16}$$

Substituting this value back into the Lagrangian we obtain that

$$\begin{aligned} q(\lambda) &= L(\mathbf{x}^*, \lambda) \\ &= (\mathbf{f} + \mathbf{A}^T\lambda)^T\mathbf{Q}^{-1}\mathbf{Q}\mathbf{Q}^{-1}(\mathbf{f} + \mathbf{A}^T\lambda) - 2(\mathbf{f} + \mathbf{A}^T\lambda)^T\mathbf{Q}^{-1}(\mathbf{f} + \mathbf{A}^T\lambda) - 2\mathbf{b}^T\lambda \\ &= -(\mathbf{f} + \mathbf{A}^T\lambda)^T\mathbf{Q}^{-1}(\mathbf{f} + \mathbf{A}^T\lambda) - 2\mathbf{b}^T\lambda \\ &= -\lambda^T\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\lambda - 2\mathbf{f}^T\mathbf{Q}^{-1}\mathbf{A}^T\lambda - \mathbf{f}^T\mathbf{Q}^{-1}\mathbf{f} - 2\mathbf{b}^T\lambda \\ &= -\lambda^T\mathbf{A}\mathbf{Q}^{-1}\mathbf{A}^T\lambda - 2(\mathbf{A}\mathbf{Q}^{-1}\mathbf{f} + \mathbf{b})^T\lambda - \mathbf{f}^T\mathbf{Q}^{-1}\mathbf{f}. \end{aligned}$$

The dual problem is

$$\max\{q(\lambda) : \lambda \geq \mathbf{0}\}.$$

This problem is also a convex quadratic problem. However, its advantage over the primal problem (12.15) is that its feasible set is "simpler." In fact, we can develop a method for solving problem (12.15) which is based on an orthogonal projection method applied to the dual problem. As an illustration, we develop a method for computing the orthogonal projection onto a polytope defined by a set of inequalities.

Example 12.14. Given a polytope

$$S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} \leq \mathbf{b}\},$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$, we wish to compute the orthogonal projection of a given point $\mathbf{y}$. As opposed to affine spaces, on which the orthogonal projection can be computed by a simple formula like the one derived in Example 10.10 of Section 10.2, there is no simple expression for the orthogonal projection onto polytopes, but using duality we will show that a method finding the projection can be derived. For a given $\mathbf{y} \in \mathbb{R}^n$, the problem of computing $P_S(\mathbf{y})$ can be written as

$$\begin{array}{ll} \min & \|\mathbf{x}-\mathbf{y}\|^2 \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b}. \end{array}$$

This problem fits into the general model (12.15) with $\mathbf{Q} = \mathbf{I}$ and $\mathbf{f} = -\mathbf{y}$. Therefore, the dual problem is (omitting constants)

$$\begin{array}{ll} \max & -\lambda^T \mathbf{A}\mathbf{A}^T \lambda - 2(-\mathbf{A}\mathbf{y} + \mathbf{b})^T \lambda \\ \text{s.t.} & \lambda \geq 0. \end{array}$$

We assume that S is nonempty, and under this assumption, by Theorem 12.12, strong duality holds. We can solve the dual problem by the orthogonal projection method (presented in Section 9.4). If we use a constant stepsize, then it can be chosen as $\frac{1}{L}$, where L is the Lipschitz constant of the gradient of the objective function given by $L = 2\lambda_{\max}(\mathbf{A}\mathbf{A}^T)$. The general step of the method would then be

$$\lambda_{k+1} = \left[\lambda_k + \frac{2}{L}(-\mathbf{A}\mathbf{A}^T\lambda_k + \mathbf{A}\mathbf{y} - \mathbf{b})\right]_+ .$$

If the method stops at iteration N, then by (12.16) the primal optimal solution (up to a tolerance) is

$$\mathbf{x}^* = \mathbf{y} - \mathbf{A}^T \lambda_N.$$

A MATLAB function implemeting this dual-based method is described below.

```
function x=proj_polytope(y,A,b,N)
%INPUT
%================
% y ................. n-length vector
% A ................. mxn matrix
% b ................. m-length vector
% N ................. number of iterations
%OUTPUT
%================
% x ................. x=P_S(y), where S = {x: Ax<=b}

d=size(A);
m=d(1);
n=d(2);
```

```
lam=zeros(m,1);

L=2*max(eig(A*A'));
g=A*y-b;
for k=1:N
    lam=lam+2/L*(-A*(A'*lam)+g);
    lam=max(lam,0);
end
x=y-A'*lam;
```

Consider for example the set

$$S = \{(x_1, x_2) : x_1 + x_2 \leq 1, x_1 \geq 0, x_2 \geq 0\}.$$

This is the triangle in the plane with vertices $(0,0),(0,1),(1,0)$. Suppose that we wish to find the orthogonal projection of $(2,-1)$ onto S. Then taking 100 iterations of the dual-based method gives the solution, which is the vertex $(1,0)$:

```
>> A=[1,1;-1,0;0,-1];
>> b=[1;0;0];
>> y=[2;-1];
>> x=proj_polytope(y,A,b,100)
x =

     1.0000
    -0.0000
```

An interesting property of the method can be seen when applying only a few iterations. For example, after 5 iterations the estimated primal solution is

```
>> x=proj_polytope(y,A,b,5)

x =

     1.5926
    -0.3663
```

This is of course not a feasible solution of the primal problem. In fact, all the iterates are nonfeasible, as illustrated in Figure 12.1, but they converge to the optimal solution, which is of course feasible. This was a demonstration of the fact that dual-based methods generate nonfeasible points with respect to the primal problem. ∎

12.3.3 ▪ Dual of Convex Quadratic Problems

Now we will consider problem (12.15) when $\mathbf{Q}$ is positive semidefinite rather than positive definite. In this case, since $\mathbf{Q}$ is not necessarily invertible, the latter construction of the dual problem is not possible. To write a dual, we will use the fact that since $\mathbf{Q} \succeq 0$, it follows that there exists a matrix $\mathbf{D} \in \mathbb{R}^{n \times n}$ such that $\mathbf{Q} = \mathbf{D}^T \mathbf{D}$. Therefore, the problem can be recast as

$$\begin{array}{ll} \min & \mathbf{x}^T \mathbf{D}^T \mathbf{D} \mathbf{x} + 2\mathbf{f}^T \mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b}. \end{array}$$

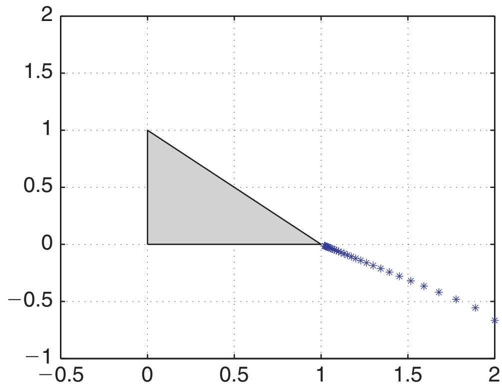

Figure 12.1. *First* 30 *iterations of the dual-based method (denoted by asterisks) for finding the orthogonal projection.*

We can now rewrite the problem by using an additional variables vector $\mathbf{z}$ as

$$\begin{array}{ll} \min_{\mathbf{x},\mathbf{z}} & \|\mathbf{z}\|^2 + 2\mathbf{f}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \le \mathbf{b}, \\ & \mathbf{z} = \mathbf{D}\mathbf{x}. \end{array}$$

The Lagrangian of the new reformulation is

$$\begin{aligned} (\lambda \in \mathbb{R}^m_+, \mu \in \mathbb{R}^n) \quad L(\mathbf{x},\mathbf{z},\lambda,\mu) &= \|\mathbf{z}\|^2 + 2\mathbf{f}^T\mathbf{x} + 2\lambda^T(\mathbf{A}\mathbf{x} - \mathbf{b}) + 2\mu^T(\mathbf{z} - \mathbf{D}\mathbf{x}) \\ &= \|\mathbf{z}\|^2 + 2\mu^T\mathbf{z} + 2(\mathbf{f} + \mathbf{A}^T\lambda - \mathbf{D}^T\mu)^T\mathbf{x} - 2\mathbf{b}^T\lambda. \end{aligned}$$

The Lagrangian is separable with respect to $\mathbf{x}$ and $\mathbf{z}$, and we can thus perform the minimizations with respect to $\mathbf{x}$ and $\mathbf{z}$ separately:

$$\begin{aligned} \min_{\mathbf{x}} (\mathbf{f} + \mathbf{A}^T\lambda - \mathbf{D}^T\mu)^T\mathbf{x} &= \begin{cases} 0, & \mathbf{f} + \mathbf{A}^T\lambda - \mathbf{D}^T\mu = 0, \\ -\infty & \text{else}, \end{cases} \\ \min_{\mathbf{z}} \|\mathbf{z}\|^2 + 2\mu^T\mathbf{z} &= -\|\mu\|^2. \end{aligned}$$

Hence,

$$q(\lambda,\mu) = \min_{\mathbf{x},\mathbf{z}} L(\mathbf{x},\mathbf{z},\lambda,\mu) = \begin{cases} -\|\mu\|^2 - 2\mathbf{b}^T\lambda, & \mathbf{f} + \mathbf{A}^T\lambda - \mathbf{D}^T\mu = 0, \\ -\infty & \text{else}. \end{cases}$$

The dual problem is thus

$$\begin{array}{ll} \max & -\|\mu\|^2 - 2\mathbf{b}^T\lambda \\ \text{s.t.} & \mathbf{f} + \mathbf{A}^T\lambda - \mathbf{D}^T\mu = 0, \\ & \lambda \in \mathbb{R}^m_+, \mu \in \mathbb{R}^n. \end{array}$$

12.3.4 ▪ Convex QCQPs

Consider the QCQP problem

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{A}_0\mathbf{x} + 2\mathbf{b}_0^T\mathbf{x} + c_0 \\ \text{s.t.} & \mathbf{x}^T\mathbf{A}_i\mathbf{x} + 2\mathbf{b}_i^T\mathbf{x} + c_i \le 0, \quad i = 1,2,\ldots,m, \end{array}$$

where $\mathbf{A}_i \succeq 0$ is an $n \times n$ matrix, $\mathbf{b}_i \in \mathbb{R}^n, c_i \in \mathbb{R}, i = 0, 1, \ldots, m$. We will consider two cases.

Case I: When $\mathbf{A}_0 \succ 0$, then the dual can be constructed as follows. The Lagrangian is

$$(\lambda \in \mathbb{R}^m_+) \quad L(\mathbf{x}, \lambda) = \mathbf{x}^T \mathbf{A}_0 \mathbf{x} + 2\mathbf{b}_0^T \mathbf{x} + c_0 + \sum_{i=1}^m \lambda_i (\mathbf{x}^T \mathbf{A}_i \mathbf{x} + 2\mathbf{b}_i^T \mathbf{x} + c_i)$$
$$= \mathbf{x}^T \left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i \right) \mathbf{x} + 2 \left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \right)^T \mathbf{x} + c_0 + \sum_{i=1}^m \lambda_i c_i.$$

The minimizer of the Lagrangian with respect to $\mathbf{x}$ is attained at the point in which its gradient is zero:

$$2 \left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i \right) \tilde{\mathbf{x}} = -2 \left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \right).$$

Thus,

$$\tilde{\mathbf{x}} = -\left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i \right)^{-1} \left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \right).$$

Plugging this expression back into the Lagrangian, we obtain the following expression for the dual objective function:

$$q(\lambda) = \min_{\mathbf{x}} L(\mathbf{x}, \lambda) = L(\tilde{\mathbf{x}}, \lambda)$$
$$= \tilde{\mathbf{x}}^T \left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i \right) \tilde{\mathbf{x}} + 2 \left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \right)^T \tilde{\mathbf{x}} + c_0 + \sum_{i=1}^m \lambda_i c_i$$
$$= -\left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \right)^T \left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i \right)^{-1} \left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \right) + c_0 + \sum_{i=1}^m \lambda_i c_i.$$

The dual problem is thus

$$\begin{array}{ll} \max & -\left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i\right)^T \left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i\right)^{-1} \left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i\right) + c_0 + \sum_{i=1}^m \lambda_i c_i \\ \text{s.t.} & \lambda_i \geq 0, \quad i = 1, 2, \ldots, m. \end{array}$$

Case II: When $\mathbf{A}_0$ is not positive definite but still positive semidefinite, the above dual is not well-defined since the matrix $\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i$ is not necessarily invertible. However, we can construct a different dual by decomposing the positive semidefinite matrices $\mathbf{A}_i$ as $\mathbf{A}_i = \mathbf{D}_i^T \mathbf{D}_i$ ($\mathbf{D}_i \in \mathbb{R}^{n \times n}$) and writing the equivalent formulation

$$\begin{array}{ll} \min & \mathbf{x}^T \mathbf{D}_0^T \mathbf{D}_0 \mathbf{x} + 2\mathbf{b}_0^T \mathbf{x} + c_0 \\ \text{s.t.} & \mathbf{x}^T \mathbf{D}_i^T \mathbf{D}_i \mathbf{x} + 2\mathbf{b}_i^T \mathbf{x} + c_i \leq 0, \quad i = 1, 2, \ldots, m, \end{array}$$

Now, we can add additional variables $\mathbf{z}_i \in \mathbb{R}^n (i = 0, 1, 2, \ldots, m)$ that will be defined to be $\mathbf{z}_i = \mathbf{D}_i \mathbf{x}$, giving rise to the formulation

$$\begin{array}{ll} \min & \|\mathbf{z}_0\|^2 + 2\mathbf{b}_0^T \mathbf{x} + c_0 \\ \text{s.t.} & \|\mathbf{z}_i\|^2 + 2\mathbf{b}_i^T \mathbf{x} + c_i \leq 0, \quad i = 1, 2, \ldots, m, \\ & \mathbf{z}_i = \mathbf{D}_i \mathbf{x}, \quad i = 0, 1, \ldots, m. \end{array}$$

The Lagrangian is ($\lambda \in \mathbb{R}^m_+, \mu_i \in \mathbb{R}^n, i = 0, 1, \ldots, m$)

$$
\begin{aligned}
& L(\mathbf{x}, \mathbf{z}_0, \ldots, \mathbf{z}_m, \lambda, \mu_0, \ldots, \mu_m) \\
&= \|\mathbf{z}_0\|^2 + 2\mathbf{b}_0^T \mathbf{x} + c_0 + \sum_{i=1}^m \lambda_i (\|\mathbf{z}_i\|^2 + 2\mathbf{b}_i^T \mathbf{x} + c_i) + 2\sum_{i=0}^m \mu_i^T (\mathbf{z}_i - \mathbf{D}_i \mathbf{x}) \\
&= \|\mathbf{z}_0\|^2 + 2\mu_0^T \mathbf{z}_0 + \sum_{i=1}^m (\lambda_i \|\mathbf{z}_i\|^2 + 2\mu_i^T \mathbf{z}_i) + 2\left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i - \sum_{i=0}^m \mathbf{D}_i^T \mu_i\right)^T \mathbf{x} \\
& + c_0 + \sum_{i=1}^m c_i \lambda_i .
\end{aligned}
$$

Note that for any $\lambda \in \mathbb{R}_+, \mu \in \mathbb{R}^n$ we have

$$
g(\lambda, \mu) \equiv \min_{\mathbf{z}} \lambda \|\mathbf{z}\|^2 + 2\mu^T \mathbf{z} = \begin{cases} -\frac{\|\mu\|^2}{\lambda}, & \lambda > 0, \\ 0, & \lambda = 0, \mu = 0, \\ -\infty, & \lambda = 0, \mu \neq 0. \end{cases}
$$

Since the Lagrangian is separable with respect to $\mathbf{z}_i$ and $\mathbf{x}$, we can perform the minimization with respect to each of the variables vectors,

$$
\begin{aligned}
\min_{\mathbf{z}_0} \left[\|\mathbf{z}_0\|^2 + 2\mu_0^T \mathbf{z}_0 \right] &= g(1, \mu_0) = -\|\mu_0\|^2, \\
\min_{\mathbf{z}_i} \left[\lambda_i \|\mathbf{z}_i\|^2 + 2\mu_i^T \mathbf{z}_i \right] &= g(\lambda_i, \mu_i), \\
\min_{\mathbf{x}} \left(\mathbf{b}_0 + \textstyle\sum_{i=1}^m \lambda_i \mathbf{b}_i - \sum_{i=0}^m \mathbf{D}_i^T \mu_i \right)^T \mathbf{x} &= \begin{cases} 0, & \mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i - \sum_{i=0}^m \mathbf{D}_i^T \mu_i = 0, \\ -\infty & \text{else,} \end{cases}
\end{aligned}
$$

and deduce that

$$
\begin{aligned}
& q(\lambda, \mu_0, \ldots, \mu_m) = \min_{\mathbf{x}, \mathbf{z}_0, \ldots, \mathbf{z}_m} L(\mathbf{x}, \mathbf{z}_0, \ldots, \mathbf{z}_m, \lambda, \mu_0, \ldots, \mu_m) \\
&= \begin{cases} g(1, \mu_0) + \sum_{i=1}^m g(\lambda_i, \mu_i) + c_0 + \sum_{i=1}^m c_i \lambda_i & \text{if } \mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i - \sum_{i=0}^m \mathbf{D}_i^T \mu_i = 0, \\ -\infty & \text{else.} \end{cases}
\end{aligned}
$$

The dual problem is therefore

$$
\begin{array}{ll}
\max & g(1, \mu_0) + \sum_{i=1}^m g(\lambda_i, \mu_i) + c_0 + \sum_{i=1}^m c_i \lambda_i \\
\text{s.t.} & \mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i - \sum_{i=0}^m \mathbf{D}_i^T \mu_i = 0, \\
& \lambda \in \mathbb{R}^m_+, \mu_0, \ldots, \mu_m \in \mathbb{R}^n.
\end{array}
$$

12.3.5 ▪ Nonconvex QCQPs

Consider now the QCQP problem

$$
\begin{array}{ll}
\min & \mathbf{x}^T \mathbf{A}_0 \mathbf{x} + 2\mathbf{b}_0^T \mathbf{x} + c_0 \\
\text{s.t.} & \mathbf{x}^T \mathbf{A}_i \mathbf{x} + 2\mathbf{b}_i^T \mathbf{x} + c_i \leq 0, \quad i = 1, 2, \ldots, m,
\end{array}
$$

where $\mathbf{A}_i = \mathbf{A}_i^T \in \mathbb{R}^{n \times n}, \mathbf{b}_i \in \mathbb{R}^n, c_i \in \mathbb{R}, i = 0, 1, \ldots, m$. We do not assume that $\mathbf{A}_i$ are positive semidefinite, and hence the problem is in general nonconvex, and the techniques used in the convex case are not applicable. We begin by forming the Lagrangian ($\lambda \in \mathbb{R}^m_+$):

$$\begin{aligned} L(\mathbf{x}, \lambda) &= \mathbf{x}^T \mathbf{A}_0 \mathbf{x} + 2\mathbf{b}_0^T \mathbf{x} + c_0 + \sum_{i=1}^m \lambda_i \left(\mathbf{x}^T \mathbf{A}_i \mathbf{x} + 2\mathbf{b}_i^T \mathbf{x} + c_i \right) \\ &= \mathbf{x}^T \left(\mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i \right) \mathbf{x} + 2 \left(\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \right)^T \mathbf{x} + c_0 + \sum_{i=1}^m \lambda_i c_i. \end{aligned}$$

The main idea is to use the following presentation of the dual objective function:

$$q(\lambda) = \min_{\mathbf{x}} L(\mathbf{x}, \lambda) = \max_t \{ t : L(\mathbf{x}, \lambda) \geq t \text{ for any } \mathbf{x} \in \mathbb{R}^n \}. \tag{12.17}$$

The above equation essentially states that the minimal value of $L(\mathbf{x}, \lambda)$ over $\mathbf{x} \in \mathbb{R}^n$ is in fact the largest lower bound on the function. We will now use Theorem 2.43 on the characterization of the nonnegativity of quadratic functions and deduce that the claim

$$L(\mathbf{x}, \lambda) \geq t \text{ for all } \mathbf{x} \in \mathbb{R}^n$$

is equivalent to

$$\begin{pmatrix} \mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i & \mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \\ (\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i)^T & c_0 + \sum_{i=1}^m \lambda_i c_i - t \end{pmatrix} \succeq \mathbf{0},$$

which combined with (12.17) yields that a dual problem is

$$\begin{array}{ll} \max_{t, \lambda_i} & t \\ \text{s.t.} & \begin{pmatrix} \mathbf{A}_0 + \sum_{i=1}^m \lambda_i \mathbf{A}_i & \mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i \\ (\mathbf{b}_0 + \sum_{i=1}^m \lambda_i \mathbf{b}_i)^T & c_0 + \sum_{i=1}^m \lambda_i c_i - t \end{pmatrix} \succeq \mathbf{0}, \\ & \lambda_i \geq 0, \quad i = 1, 2, \ldots, m. \end{array} \tag{12.18}$$

The above problem is convex as a dual problem, but since the primal problem is nonconvex, strong duality is of course not guaranteed. We also note that the form of the dual problem (12.18) is different from all the dual problems derived so far in the sense that not all the constraints are presented as inequality or equality constraints but instead are of the form $\mathbf{B}_0 + \sum_{i=1}^m \lambda_i \mathbf{B}_i \succeq \mathbf{0}$, where $\mathbf{B}_i$ are given matrices. This type of a constraint is called a *linear matrix inequality* (abbreviated LMI). Optimization problems consisting of minimization or maximization of a linear function subject to linear inequalities/ equalities and LMIs are called semidefinite programming (SDP) problems, and they are part of a larger class of problems called *conic problems*; see also Section 12.3.9.

12.3.6 ▪ Orthogonal Projection onto the Unit-Simplex

Given a vector $\mathbf{y} \in \mathbb{R}^n$, we would like to compute the orthogonal projection of the vector $\mathbf{y}$ onto Δ_n. The corresponding optimization problem is

$$\begin{array}{ll} \min & \|\mathbf{x} - \mathbf{y}\|^2 \\ \text{s.t.} & \mathbf{e}^T \mathbf{x} = 1, \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

We will associate a Lagrange multiplier $\lambda \in \mathbb{R}$ to the linear equality constraint $\mathbf{e}^T \mathbf{x} = 1$ and obtain the Lagrangian function

$$\begin{aligned} L(\mathbf{x}, \lambda) &= \|\mathbf{x} - \mathbf{y}\|^2 + 2\lambda(\mathbf{e}^T \mathbf{x} - 1) = \|\mathbf{x}\|^2 - 2(\mathbf{y} - \lambda \mathbf{e})^T \mathbf{x} + \|\mathbf{y}\|^2 - 2\lambda \\ &= \sum_{j=1}^n (x_j^2 - 2(y_j - \lambda) x_j) + \|\mathbf{y}\|^2 - 2\lambda. \end{aligned}$$

The arising problem is therefore saparable with respect to the variables x_j and hence the optimal x_j is the solution to the one-dimensional problem

$$\min_{x_j \geq 0}[x_j^2 - 2(y_j - \lambda)x_j].$$

The optimal solution to the above problem is given by

$$x_j = \begin{cases} y_j - \lambda, & y_j \geq \lambda \\ 0 & \text{else} \end{cases} = [y_j - \lambda]_+,$$

and the optimal value is $-[y_j - \lambda]_+^2$. The dual problem is therefore

$$\max_{\lambda \in \mathbb{R}} \left\{ g(\lambda) \equiv -\sum_{j=1}^{n} [y_j - \lambda]_+^2 - 2\lambda + \|\mathbf{y}\|^2 \right\}.$$

By the basic properties of dual problems, the dual objective function is concave. In order to actually solve the dual problem, we note that

$$\lim_{\lambda \to \infty} g(\lambda) = \lim_{\lambda \to -\infty} g(\lambda) = -\infty.$$

Therefore, since $-g$ is a coercive and differentiable function, it follows that there exists an optimal solution to the dual problem attained at a point λ in which

$$g'(\lambda) = 0,$$

meaning that

$$\sum_{j=1}^{n} [y_j - \lambda]_+ = 1.$$

The function $h(\lambda) = \sum_{j=1}^{n} [y_j - \lambda]_+ - 1$ is a nonincreasing function over $\mathbb{R}$ and is in fact strictly decreasing over $(-\infty, \max_j y_j]$. In addition, by denoting $y_{\max} = \max_{j=1,2,\ldots,n} y_j, y_{\min} = \min_{j=1,2,\ldots,n} y_j$, we have

$$h(y_{\max}) = -1,$$
$$h\left(y_{\min} - \frac{2}{n}\right) = \sum_{j=1}^{n} y_j - n y_{\min} + 2 - 1 > 0,$$

and we can therefore invoke a bisection procedure to find the unique root λ of the function h over the interval $[y_{\min} - \frac{2}{n}, y_{\max}]$ and then define $P_{\Delta_n}(\mathbf{y}) = [\mathbf{y} - \lambda \mathbf{e}]_+$. The MATLAB implementation follows.

```
function xp=proj_unit_simplex(y)
% INPUT
% ======================
% y ........a column vector
% OUTPUT
% ======================
% xp .......the orthogonal projection of y onto the unit simplex

f=@(lam)sum(max(y-lam,0))-1;
n=length(y);
lb=min(y)-2/n;
```

```
ub=max(y);
lam=bisection(f,lb,ub,1e-10);
xp=max(y-lam,0);
```

As a sanity check, let us compute the orthogonal projection of the vector $(-1,1,0.3)^T$ onto Δ_3,

```
>> xp=proj_unit_simplex([-1;1;0.3])
xp =
         0
    0.8500
    0.1500
```

and compare it with the result of CVX,

```
cvx_begin
variable x(3)
minimize(norm(x-[-1;1;0.3]))
sum(x)==1
x>=0
cvx_end
```

which unsurprisingly is the same:

```
>> x
x =
    0.0000
    0.8500
    0.1500
```

12.3.7 ▪ Dual of the Chebyshev Center Problem

Recall that in the Chebyshev center problem (see also Section 8.2.4) we are given a set of points $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \in \mathbb{R}^n$, and we seek to find a point $\mathbf{x} \in \mathbb{R}^n$, which is the center of the minimum radius ball containing the points

$$\begin{array}{ll} \min_{\mathbf{x},r} & r \\ \text{s.t.} & \|\mathbf{x}-\mathbf{a}_i\| \le r, \quad i=1,2,\ldots,m. \end{array}$$

Finding a dual problem to this formulation is not an easy task, but we can actually consider a different equivalent formulation to which a dual can be constructed in an easier way. The problem can be recast as

$$\begin{array}{ll} \min_{\mathbf{x},\gamma} & \gamma \\ \text{s.t.} & \|\mathbf{x}-\mathbf{a}_i\|^2 \le \gamma, \quad i=1,2,\ldots,m. \end{array}$$

where γ denotes the squared radius. The problems are equivalent since minimization of the radius is equivalent to minimization of the squared radius. The Lagrangian is ($\lambda \in \mathbb{R}^m_+$)

$$\begin{aligned} L(\mathbf{x},\gamma,\lambda) &= \gamma + \sum_{i=1}^m \lambda_i(\|\mathbf{x}-\mathbf{a}_i\|^2 - \gamma) \\ &= \gamma\left(1-\sum_{i=1}^m \lambda_i\right) + \sum_{i=1}^m \lambda_i \|\mathbf{x}-\mathbf{a}_i\|^2. \end{aligned}$$

The minimization of the above expression must be $-\infty$ unless $\sum_{i=1}^m \lambda_i = 1$, and in this case we have

$$\min_{\gamma} \gamma\left(1-\sum_{i=1}^m \lambda_i\right)=0.$$

We are therefore left with the task of finding the optimal value of

$$\min_{\mathbf{x}} \sum_{i=1}^m \lambda_i \|\mathbf{x}-\mathbf{a}_i\|^2$$

when $\lambda \in \Delta_m$. Since the objective function of the above minimization can be written as

$$\sum_{i=1}^m \lambda_i \|\mathbf{x}-\mathbf{a}_i\|^2 = \|\mathbf{x}\|^2 - 2\left(\sum_{i=1}^m \lambda_i \mathbf{a}_i\right)^T \mathbf{x} + \sum_{i=1}^m \lambda_i \|\mathbf{a}_i\|^2, \tag{12.19}$$

it follows that the minimum is attained at the point in which the gradient vanishes, meaning at

$$\mathbf{x}^* = \sum_{i=1}^m \lambda_i \mathbf{a}_i = \mathbf{A}\lambda, \tag{12.20}$$

where $\mathbf{A}$ is the $n \times m$ matrix whose columns are the vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m$. Substituting this expression back into (12.19), we have that the dual objective function is

$$q(\lambda) = \|\mathbf{A}\lambda\|^2 - 2(\mathbf{A}\lambda)^T(\mathbf{A}\lambda) + \sum_{i=1}^m \lambda_i \|\mathbf{a}_i\|^2 = -\|\mathbf{A}\lambda\|^2 + \sum_{i=1}^m \lambda_i \|\mathbf{a}_i\|^2.$$

The dual problem is therefore

$$\begin{array}{ll} \max & -\|\mathbf{A}\lambda\|^2 + \sum_{i=1}^m \lambda_i \|\mathbf{a}_i\|^2 \\ \text{s.t.} & \lambda \in \Delta_m. \end{array}$$

We can actually write a MATLAB function that solves this problem. For that, we will use the gradient projection method with a constant stepsize $\frac{1}{L}$, where $L = 2\lambda_{\max}(\mathbf{A}^T\mathbf{A})$ is the Lipschitz constant of the gradient of the objective function. At each iteration we will also use the MATLAB function `proj_unit_simplex` to find the orthogonal projection onto the unit-simplex. Note that the derived method is also a dual-based method and that it incorporates another dual-based method for computing the projection.

```
function [xp,r]=chebyshev_center(A)
% INPUT
% ======================
% A ........ a matrix whose columns are points in R^m
% OUTPUT
% ======================
% xp ........ the Chebyshev center of the points
% r ......... radius of the Chebyshev circle

d=size(A);
m=d(2);
Q=A'*A;
L=2*max(eig(Q));
b=sum(A.^2)';
%initialization with the uniform vector
lam=1/m*ones(m,1);
```

```
old_lam=zeros(m,1);
while (norm(lam-old_lam)>1e-5)
    old_lam=lam;
    lam=proj_unit_simplex(lam+1/L*(-2*Q*lam+b));
end
xp=A*lam;
r=0;
for i=1:m
    r=max(r,norm(xp-A(:,i)));
end
```

Example 12.15. Returning to Example 8.14, suppose that we wish to find the Chebyshev center of the 5 points

$$(-1,3),\quad (-3,10),\quad (-1,0),\quad (5,0),\quad (-1,-5).$$

For that, we can invoke the MATLAB function `chebyshev_center` that was just described:

```
A=[-1,-3,-1,5,-1;3,10,0,0,-5];
[xp,r]=chebyshev_center(A);
```

The Chebyshev center and radius are

```
 >> xp

xp =

   -2.0000
    2.5000

>> r

r =

    7.5664
```

These are the exact same results obtained by CVX in Example 8.14. ∎

12.3.8 ▪ Minimization of Sum of Norms

Consider the problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \sum_{i=1}^{m} \|\mathbf{A}_i\mathbf{x}+\mathbf{b}_i\|, \tag{12.21}$$

where $\mathbf{A}_i \in \mathbb{R}^{k_i\times n}, \mathbf{b}_i \in \mathbb{R}^{k_i}, i = 1,2,\ldots,m$. At a first glance, it seems that it is not possible to find a dual of an unconstrained problem. However, we can use a technique of variable decoupling to reformulate the problem as a problem with affine constraints. Specifically, problem (12.21) is the same as

$$\begin{array}{ll} \min_{\mathbf{x},\mathbf{y}_i} & \sum_{i=1}^{m} \|\mathbf{y}_i\| \\ \text{s.t.} & \mathbf{y}_i = \mathbf{A}_i\mathbf{x}+\mathbf{b}_i, \quad i=1,2,\ldots,m. \end{array}$$

Associating a Lagrange multiplier vector $\lambda_i \in \mathbb{R}^{k_i}$ with the ith set of constraints $\mathbf{y}_i = \mathbf{A}_i\mathbf{x}+\mathbf{b}_i$, we obtain the following Lagrangian:

$$\begin{aligned} L(\mathbf{x},\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_m,\lambda_1,\lambda_2,\ldots,\lambda_m) &= \sum_{i=1}^m \|\mathbf{y}_i\| + \sum_{i=1}^m \lambda_i^T(\mathbf{y}_i - \mathbf{A}_i\mathbf{x}-\mathbf{b}_i) \\ &= \sum_{i=1}^m \left[\|\mathbf{y}_i\| + \lambda_i^T\mathbf{y}_i\right] - \left(\sum_{i=1}^m \mathbf{A}_i^T\lambda_i\right)^T \mathbf{x} - \sum_{i=1}^m \mathbf{b}_i^T\lambda_i. \end{aligned}$$

By the separability of the Lagrangian with respect to $\mathbf{x},\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_m$, it follows that the dual objective function is given by

$$q(\lambda_1,\lambda_2,\ldots,\lambda_m) = \sum_{i=1}^m \min_{\mathbf{y}_i\in\mathbb{R}^{k_i}} \left[\|\mathbf{y}_i\| + \lambda_i^T\mathbf{y}_i\right] + \min_{\mathbf{x}\in\mathbb{R}^n}\left[-\left(\sum_{i=1}^m \mathbf{A}_i^T\lambda_i\right)^T\mathbf{x}\right] - \sum_{i=1}^m \mathbf{b}_i^T\lambda_i.$$

Obviously,

$$\min_{\mathbf{x}\in\mathbb{R}^n}\left[-\left(\sum_{i=1}^m \mathbf{A}_i^T\lambda_i\right)^T\mathbf{x}\right] = \begin{cases} 0, & \sum_{i=1}^m \mathbf{A}_i^T\lambda_i = \mathbf{0}, \\ -\infty & \text{else.}\end{cases} \tag{12.22}$$

In addition, we have for any $\mathbf{a}\in\mathbb{R}^k$

$$\min_{\mathbf{y}\in\mathbb{R}^k} \|\mathbf{y}\| + \mathbf{a}^T\mathbf{y} = \begin{cases} 0, & \|\mathbf{a}\|\le 1, \\ -\infty, & \|\mathbf{a}\| > 1.\end{cases} \tag{12.23}$$

To prove this result, note that when $\|\mathbf{a}\|\le 1$, we have by the Cauchy–Schwarz inequality that for any $\mathbf{y}\in\mathbb{R}^k$

$$\mathbf{a}^T\mathbf{y} \ge -\|\mathbf{a}\|\cdot\|\mathbf{y}\| \ge -\|\mathbf{y}\|$$

and hence $\|\mathbf{y}\| + \mathbf{a}^T\mathbf{y}\ge 0$ for any $\mathbf{y}\in\mathbb{R}^k$, and in addition $\|\mathbf{0}\| + \mathbf{a}^T\mathbf{0} = 0$, implying that $\min_{\mathbf{y}\in\mathbb{R}^k}\|\mathbf{y}\| + \mathbf{a}^T\mathbf{y} = 0$. If $\|\mathbf{a}\| > 1$, then taking $\mathbf{y}_\alpha = -\alpha\mathbf{a}$ we obtain that $\|\mathbf{y}_\alpha\| + \mathbf{a}^T\mathbf{y}_\alpha = \alpha\|\mathbf{a}\|(1-\|\mathbf{a}\|)\to -\infty$ as $\alpha\to\infty$, establishing the result (12.23). We thus conclude that for any $i = 1,2,\ldots,m$

$$\min_{\mathbf{y}_i\in\mathbb{R}^{k_i}}\left[\|\mathbf{y}_i\| + \lambda_i^T\mathbf{y}_i\right] = \begin{cases} 0, & \|\lambda_i\|\le 1, \\ -\infty & \text{else,}\end{cases}$$

which combined with (12.22) implies that the dual objective function is

$$q(\lambda_1,\lambda_2,\ldots,\lambda_m) = \begin{cases} -\sum_{i=1}^m \lambda_i^T\mathbf{b}_i, & \|\lambda_i\|\le 1, i = 1,2,\ldots,m, \\ & \sum_{i=1}^m \mathbf{A}_i^T\lambda_i = \mathbf{0}, \\ -\infty & \text{else.}\end{cases}$$

The dual problem is therefore

$$\begin{array}{ll} \max & -\sum_{i=1}^m \mathbf{b}_i^T\lambda_i \\ \text{s.t.} & \sum_{i=1}^m \mathbf{A}_i^T\lambda_i = \mathbf{0}, \\ & \|\lambda_i\|\le 1, \quad i = 1,2,\ldots,m. \end{array}$$

12.3.9 ▪ Conic Duality

A conic optimization problem is a convex optimization problem of the form

$$
(\text{C}) \quad \begin{array}{ll} \min & \mathbf{a}^T\mathbf{x} \\ \text{s.t.} & \mathbf{A}\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \in K, \end{array}
$$

where $K \subseteq \mathbb{R}^n$ is a closed and convex cone and $\mathbf{a} \in \mathbb{R}^n, \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$. Our aim is to find an expression for the dual problem. For that, let us associate a Lagrange multipliers vector $\mathbf{y} \in \mathbb{R}^m$ with the equality constraints $\mathbf{A}\mathbf{x} = \mathbf{b}$, leading to the Lagrangian

$$
L(\mathbf{x},\mathbf{y}) = \mathbf{a}^T\mathbf{x} - \mathbf{y}^T(\mathbf{A}\mathbf{x} - \mathbf{b}) = (\mathbf{a} - \mathbf{A}^T\mathbf{y})^T\mathbf{x} + \mathbf{b}^T\mathbf{y}.
$$

To construct the dual problem, we need to minimize the Lagrangian over $\mathbf{x} \in K$:

$$
q(\mathbf{y}) = \min_{\mathbf{x}\in K} L(\mathbf{x},\mathbf{y}) = \mathbf{b}^T\mathbf{y} + \min_{\mathbf{x}\in K} \left(\mathbf{a} - \mathbf{A}^T\mathbf{y}\right)^T \mathbf{x}.
$$

We will now use the following easily verifiable fact: for any $\mathbf{d} \in \mathbb{R}^n$ one has

$$
\min_{\mathbf{x}\in K} \mathbf{d}^T\mathbf{x} = \begin{cases} 0, & \mathbf{d} \in K^*, \\ -\infty, & \mathbf{d} \notin K^*. \end{cases}
$$

where K^* is the dual cone defined in Exercise 6.11 as

$$
K^* = \{\mathbf{z} \in \mathbb{R}^n : \mathbf{z}^T\mathbf{x} \geq 0 \text{ for any } \mathbf{x} \in K\}.
$$

The latter fact is almost a tautology. If $\mathbf{d} \in K^*$, then by the definition of the dual cone, $\mathbf{d}^T\mathbf{x} \geq 0$ for any $\mathbf{x} \in K$ and also $\mathbf{d}^T 0 = 0$ (recall that $0 \in K$ since K is a closed cone), and hence $\min_{\mathbf{x}\in K} \mathbf{d}^T\mathbf{x} = 0$. On the other hand, if $\mathbf{d} \notin K^*$, then it means that there exists $\mathbf{x}_0 \in K$ such that $\mathbf{d}^T\mathbf{x}_0 < 0$. Therefore, taking any $\alpha > 0$ we have that $\alpha\mathbf{x}_0 \in K$, and we obtain that $\mathbf{d}^T(\alpha\mathbf{x}_0) = \alpha(\mathbf{d}^T\mathbf{x}_0) \to -\infty$ as α tends to ∞, and hence $\min_{\mathbf{x}\in K} \mathbf{d}^T\mathbf{x} = -\infty$. To conclude, the dual objective function is

$$
q(\mathbf{y}) = \begin{cases} \mathbf{b}^T\mathbf{y}, & \mathbf{a} - \mathbf{A}^T\mathbf{y} \in K^*, \\ -\infty, & \mathbf{a} - \mathbf{A}^T\mathbf{y} \notin K^*. \end{cases}
$$

The dual problem is thus

$$
(\text{DC}) \quad \begin{array}{ll} \max & \mathbf{b}^T\mathbf{y} \\ \text{s.t.} & \mathbf{a} - \mathbf{A}^T\mathbf{y} \in K^*. \end{array}
$$

We can now invoke the strong duality theorem for convex problems (Theorem 12.12) and state one of the versions of the so-called *conic duality theorem*.

Theorem 12.16 (conic duality theorem). *Consider the primal and dual problems* (C) *and* (DC). *Suppose that there exists* $\mathbf{x} \in \text{int}(K)$ *such that* $\mathbf{A}\mathbf{x} = \mathbf{b}$ *and that problem* (C) *is bounded below. Then the dual problem* (DC) *has an optimal solution, and we have*

$$
val(C) = val(DC).
$$

12.3.10 ▪ Denoising

Suppose that we are given a signal contaminated with noise. In mathematical terms the model is

$$\mathbf{y} = \mathbf{x} + \mathbf{w},$$

where $\mathbf{x} \in \mathbb{R}^n$ is the noise-free signal, $\mathbf{w} \in \mathbb{R}^n$ is the unknown noise vector, and $\mathbf{y} \in \mathbb{R}^n$ is the observed and known vector. An example of "clean" and "noisy" signals can be found in Figure 12.2

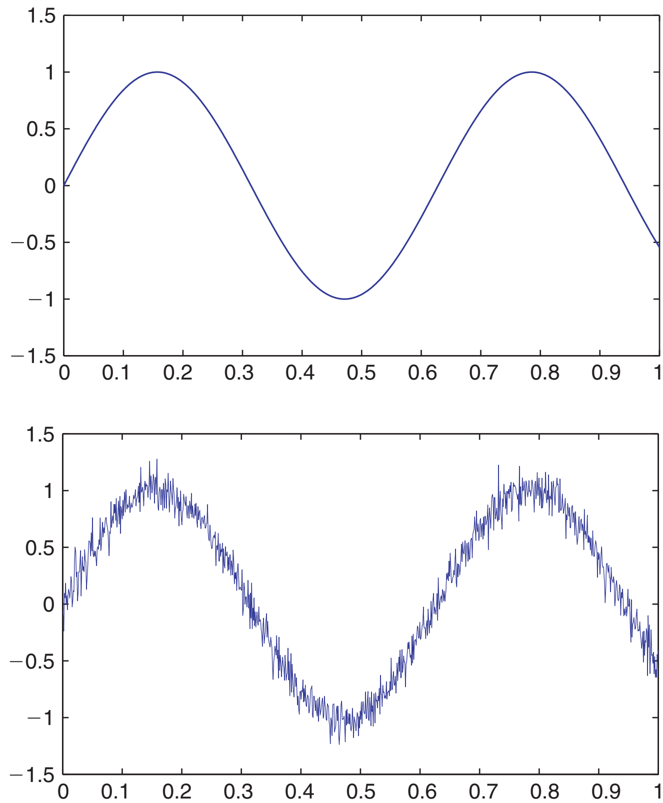

Figure 12.2. *True signal (top image) and noisy signal (bottom image).*

The plots were created by the MATLAB commands

```
randn('seed',314);
t=linspace(0,1,1000)';
n=length(t);
x=sin(10*t);
figure(1)
plot(t,x)
axis([0,1,-1.5,1.5]);
y=x+0.1*randn(size(t));
figure(2)
plot(t,y)
```

The objective is to reconstruct the true signal from the observed vector $\mathbf{y}$. A common approach for denoising is to use some prior information on the true image. A natural information is the smoothness of the signal. This information can be incorporated by adding a quadratic penalty function that measures in some sense the smoothness of the signal. For example, a standard approach is to solve the optimization problem

$$\min \|\mathbf{x}-\mathbf{y}\|^2 + \lambda \sum_{i=1}^{n-1} (x_i - x_{i+1})^2,$$

where $\lambda > 0$ is some predetermined regularization parameter. The problem can also be written as

$$\min \|\mathbf{x}-\mathbf{y}\|^2 + \lambda \|\mathbf{L}\mathbf{x}\|^2, \tag{12.24}$$

where

$$\mathbf{L} = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}.$$

Problem (12.24) is a regularized least squares problem (see Section 3.3), and its optimal solution can be derived by writing the stationarity condition

$$2(\mathbf{x}-\mathbf{y}) + 2\lambda \mathbf{L}^T \mathbf{L}\mathbf{x} = \mathbf{0}.$$

Thus,

$$\mathbf{x} = (\mathbf{I} + \lambda \mathbf{L}^T \mathbf{L})^{-1} \mathbf{y}. \tag{12.25}$$

The solution of the problem in the case of $\lambda = 1$ can thus be obtained by the MATLAB commands

```
L=sparse(n-1,n);
for i=1:n-1
    L(i,i)=1;
    L(i,i+1)=-1;
end

lambda=100;
xde=(speye(n)+lambda*L'*L)\y;
figure(3)
plot(t,xde);
```

resulting in the relatively good reconstruction given in Figure 12.3. The quadratic regularization method does not work so well for all types of signals. Suppose, for example, that we are given a noisy step signal generated by the MATLAB commands

```
randn('seed',314);
x=zeros(1000,1);
x(1:250)=1;
x(251:500)=3;
x(501:750)=0;
x(751:1000)=2;
```

```
figure(1)
plot(1:1000,x,'.')
axis([0,1000,-1,4]);
y=x+0.05*randn(size(x));
figure(2)
plot(1:1000,y,'.')
```

The "true" and "noisy" step signals are given in Figure 12.4.

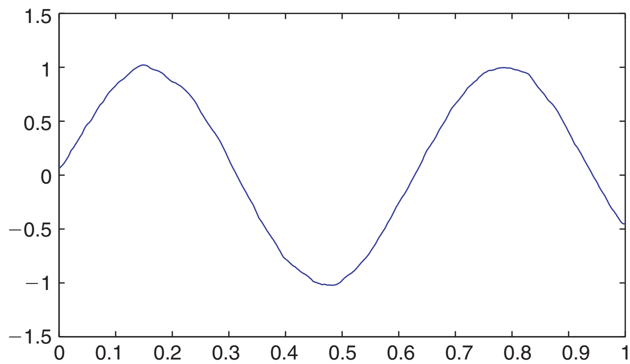

Figure 12.3. *Denoising of the sine signal via quadratic regularization.*

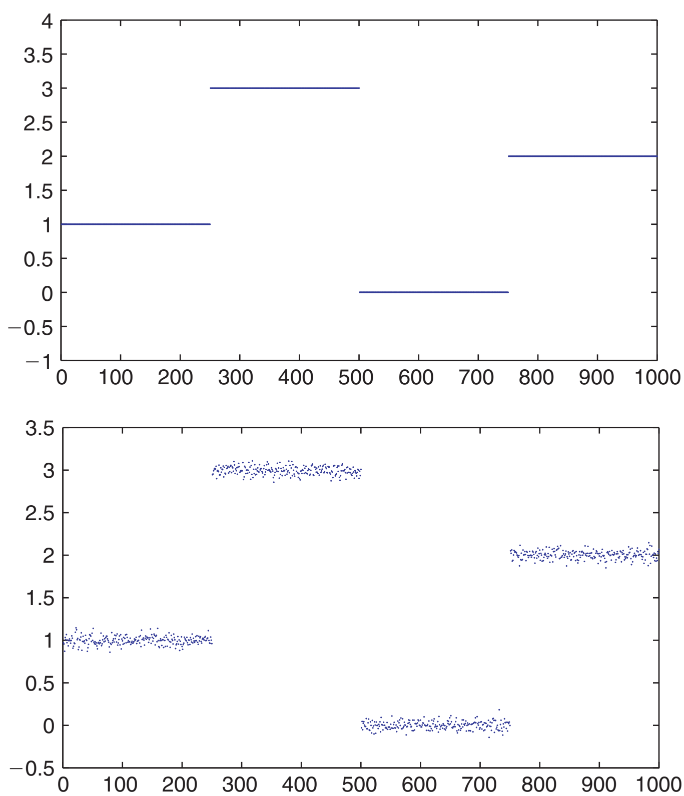

Figure 12.4. *True signal (top image) and noisy signal (bottom image).*

Unfortunately, the quadratic regularization approach does not yield good results, no matter what the value of the chosen regularization parameter λ is. Indeed, the regularized least squares solution (12.25) is not a good reconstruction since it is unable to deal correctly with the three breakpoints. The reason is that the jumps contribute large values to the penalty function $||\mathbf{Lx}||^2$ since their values are squared. Therefore, in a sense the regularized least squares solution tries to "smooth" the jumps, resulting in the reconstructions in Figure 12.5.

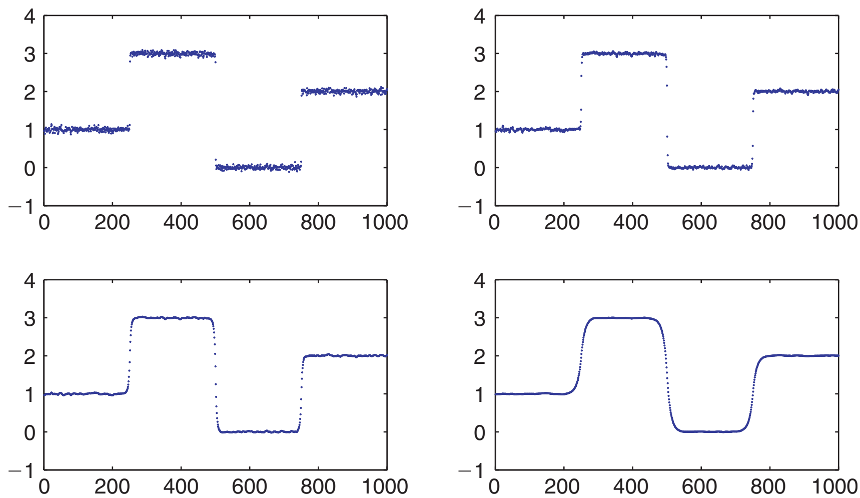

Figure 12.5. *Quadratic regularization with* $\lambda = 0.1$ *(top left),* $\lambda = 1$ *(top right),* $\lambda = 10$ *(bottom left),* $\lambda = 100$ *(bottom right).*

Another approach for denoising that is able to overcome this disadvantage is to solve the following problem in which the regularization term is in the l_1 norm:

$$\min ||\mathbf{x}-\mathbf{y}||^2 + \lambda||\mathbf{Lx}||_1. \tag{12.26}$$

We would like to construct a dual to problem (12.26). For that, note that the problem is equivalent to the optimization problem

$$\begin{array}{ll} \min_{\mathbf{x},\mathbf{z}} & ||\mathbf{x}-\mathbf{y}||^2 + \lambda||\mathbf{z}||_1 \\ \text{s.t.} & \mathbf{z} = \mathbf{Lx}. \end{array}$$

The Lagrangian of the problem is

$$\begin{aligned} L(\mathbf{x},\mathbf{z},\mu) &= ||\mathbf{x}-\mathbf{y}||^2 + \lambda||\mathbf{z}||_1 + \mu^T(\mathbf{Lx}-\mathbf{z}) \\ &= ||\mathbf{x}-\mathbf{y}||^2 + (\mathbf{L}^T\mu)^T\mathbf{x} + \lambda||\mathbf{z}||_1 - \mu^T\mathbf{z}. \end{aligned}$$

The Lagrangian is separable with respect to $\mathbf{x}$ and $\mathbf{z}$ and thus we can perform the minimization separately. The minimum of $||\mathbf{x}-\mathbf{y}||^2 + (\mathbf{L}^T\mu)^T\mathbf{x}$ over $\mathbf{x}$ is attained when the gradient vanishes,

$$2(\mathbf{x}-\mathbf{y}) + \mathbf{L}^T\mu = \mathbf{0},$$

and hence $\mathbf{x} = \mathbf{y} - \frac{1}{2}\mathbf{L}^T\mu$. Substituting this value back to the $\mathbf{x}$-part of the Lagrangian, we obtain

$$\min_{\mathbf{x}} ||\mathbf{x}-\mathbf{y}||^2 + (\mathbf{L}^T\mu)^T\mathbf{x} = -\frac{1}{4}\mu^T\mathbf{LL}^T\mu + \mu^T\mathbf{Ly}.$$

In addition,

$$\min_{\mathbf{z}} \lambda\|\mathbf{z}\|_1 - \mu^T \mathbf{z} = \begin{cases} 0, & \|\mu\|_\infty \le \lambda, \\ -\infty & \text{else.} \end{cases}$$

To conclude, the dual objective function is given by

$$q(\mu) = \min_{\mathbf{x},\mathbf{z}} L(\mathbf{x},\mathbf{z},\mu) = \begin{cases} -\frac{1}{4}\mu^T \mathbf{L}\mathbf{L}^T \mu + \mu^T \mathbf{L}\mathbf{y}, & \|\mu\|_\infty \le \lambda, \\ -\infty & \text{else.} \end{cases}$$

Therefore, the dual problem is

$$\begin{array}{ll} \max & -\frac{1}{4}\mu^T \mathbf{L}\mathbf{L}^T \mu + \mu^T \mathbf{L}\mathbf{y} \\ \text{s.t.} & \|\mu\|_\infty \le \lambda. \end{array} \tag{12.27}$$

Since the feasible set of the dual problem is a box, we can employ the gradient projection method in order to solve it. For that, we need to know an upper bound on the Lipschitz constant of its gradient. To find such an upper bound, note that

$$\|\mathbf{L}\mathbf{x}\|^2 = \sum_{i=1}^{n-1}(x_i - x_{i+1})^2 \le 2\left(\sum_{i=1}^{n-1} x_i^2 + \sum_{i=1}^{n-1} x_{i+1}^2\right) \le 4\|\mathbf{x}\|^2.$$

Therefore

$$\lambda_{\max}(\mathbf{L}\mathbf{L}^T) = \lambda_{\max}(\mathbf{L}^T\mathbf{L}) \le 4.$$

Hence, since the Lipschitz constant of the gradient of the objective function of (12.27) is $\frac{1}{2}\lambda_{\max}(\mathbf{L}\mathbf{L}^T)$, it follows that an upper bound on the Lipschitz constant is 2. The consequence is that we can employ the gradient projection method on problem (12.27) with constant stepsize $\frac{1}{2}$, and the convergence is guaranteed by Theorems 9.16 and 9.18. Explicitly, the method will read as

$$\mu_{k+1} = P_C\left(\mu_k - \frac{1}{4}\mathbf{L}\mathbf{L}^T \mu_k + \frac{1}{2}\mathbf{L}\mathbf{y}\right),$$

where

$$C = \left\{\mathbf{z} \in \mathbb{R}^{n-1} : -\lambda \le z_i \le \lambda, i = 1, 2\ldots, n-1\right\}.$$

If the result of the gradient projection method is μ^*, the primal optimal solution (up to some tolerance) will be $\mathbf{x}^* = \mathbf{y} - \frac{1}{2}\mathbf{L}^T\mu^*$. Following is a short MATLAB code that employs 1000 iterations of the gradient projection method:

```
lambda=1;
mu=zeros(n-1,1);
for i=1:1000
    mu=mu-0.25*L*(L'*mu)+0.5*(L*y);
    mu=lambda*mu./max(abs(mu),lambda);
    xde=y-0.5*L'*mu;
    end
figure(5)
plot(t,xde,'.');
axis([0,1,-1,4])
```

and the result is given in Figure 12.6. This result is much better than any of the quadratic regularization reconstructions, and it captures the breakpoints very well.

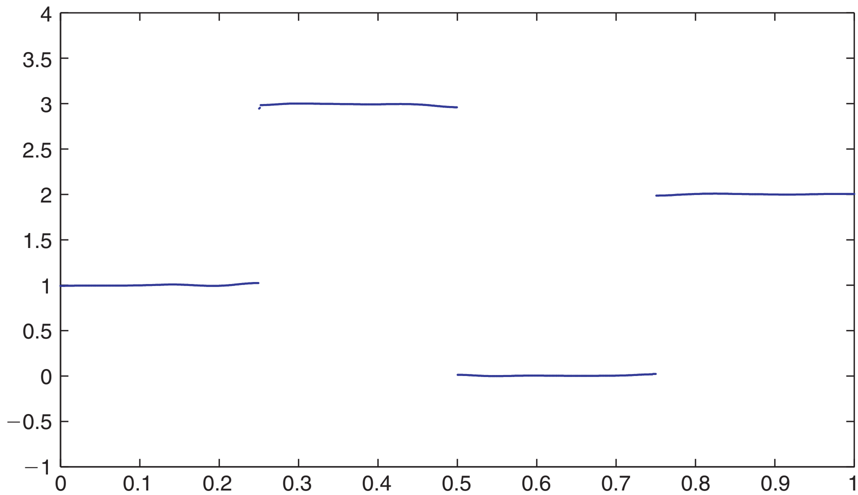

Figure 12.6. *Result of denoising via an l_1 norm regularization.*

12.3.11 ▪ Dual of the Linear Separation Problem

In Section 8.2.3 we considered the problem of finding a maximal margin hyperplane that separates two sets of points. We will assume that the given classified points are $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m \in \mathbb{R}^n$. For each i, we are given a scalar y_i which is equal to 1 if $\mathbf{x}_i$ is in class A or -1 if it is in class B. The linear separation problem is given by

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{s.t.} & y_i(\mathbf{w}^T\mathbf{x}_i + \beta) \geq 1, \quad i = 1, 2, \ldots, m. \end{array} \tag{12.28}$$

The disadvantage of the formulation (12.28) is that it is only relevant when the two classes of points are linearly separable. However, in many practical situations the two classes are not linearly separable, and in this case we need to find a formulation in which violation of the constraints is allowed and at the same time a penalty term is added to the objective function that is equal to the sum of the violations of the constraints. The new formulation is as follows:

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^m \xi_i \\ \text{s.t.} & y_i(\mathbf{w}^T\mathbf{x}_i + \beta) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, m, \\ & \xi_i \geq 0, \quad i = 1, 2, \ldots, m, \end{array}$$

where $C > 0$ is a given parameter. We will rewrite the problem in a slightly different form:

$$\begin{array}{ll} \min & \frac{1}{2}\|\mathbf{w}\|^2 + C(\mathbf{e}^T\xi) \\ \text{s.t.} & \mathbf{Y}(\mathbf{X}\mathbf{w} + \beta\mathbf{e}) \geq \mathbf{e} - \xi, \\ & \xi \geq 0, \end{array}$$

where $\mathbf{Y} = \text{diag}(y_1, y_2, \ldots, y_m)$ and $\mathbf{X}$ is the $m \times n$ matrix whose rows are $\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_m^T$. We begin by constructing the Lagrangian ($\alpha \in \mathbb{R}^m_+$)

$$\begin{aligned} L(\mathbf{w}, \beta, \xi, \alpha) &= \frac{1}{2}\|\mathbf{w}\|^2 + C(\mathbf{e}^T\xi) - \alpha^T[\mathbf{YXw} + \beta\mathbf{Ye} - \mathbf{e} + \xi] \\ &= \frac{1}{2}\|\mathbf{w}\|^2 - \mathbf{w}^T[\mathbf{X}^T\mathbf{Y}\alpha] - \beta(\alpha^T\mathbf{Ye}) + \xi^T(C\mathbf{e} - \alpha) + \alpha^T\mathbf{e}. \end{aligned}$$

The Lagrangian is separable with respect to $\mathbf{w}, \beta$ and ξ and therefore

$$q(\alpha) = \left[\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 - \mathbf{w}^T[\mathbf{X}^T\mathbf{Y}\alpha]\right] + \left[\min_{\beta}(-\beta(\alpha^T\mathbf{Y}\mathbf{e}))\right] + \left[\min_{\xi \geq 0} \xi^T(C\mathbf{e} - \alpha)\right] + \alpha^T\mathbf{e}.$$

Since

$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 - \mathbf{w}^T[\mathbf{X}^T\mathbf{Y}\alpha] = -\frac{1}{2}\alpha^T\mathbf{Y}\mathbf{X}\mathbf{X}^T\mathbf{Y}\alpha,$$

$$\min_{\beta}(-\beta(\alpha^T\mathbf{Y}\mathbf{e})) = \begin{cases} 0, & \alpha^T\mathbf{Y}\mathbf{e} = 0, \\ -\infty & \text{else,} \end{cases}$$

$$\min_{\xi \geq 0} \xi^T(C\mathbf{e} - \alpha) = \begin{cases} 0, & \alpha \leq C\mathbf{e}, \\ -\infty & \text{else,} \end{cases}$$

it follows that the dual objective function is given by

$$q(\alpha) = \begin{cases} \alpha^T\mathbf{e} - \frac{1}{2}\alpha^T\mathbf{Y}\mathbf{X}\mathbf{X}^T\mathbf{Y}\alpha, & \alpha^T\mathbf{Y}\mathbf{e} = 0, 0 \leq \alpha \leq C\mathbf{e}, \\ -\infty & \text{else.} \end{cases}$$

The dual problem is therefore

$$\begin{array}{ll} \max & \alpha^T\mathbf{e} - \frac{1}{2}\alpha^T\mathbf{Y}\mathbf{X}\mathbf{X}^T\mathbf{Y}\alpha \\ \text{s.t.} & \alpha^T\mathbf{Y}\mathbf{e} = 0, \\ & 0 \leq \alpha \leq C\mathbf{e}. \end{array}$$

We can also write the dual problem in the following way:

$$\begin{array}{ll} \max & \sum_{i=1}^m \alpha_i - \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^m \alpha_i\alpha_j y_i y_j (\mathbf{x}_i^T\mathbf{x}_j) \\ \text{s.t.} & \sum_{i=1}^m y_i\alpha_i = 0, \\ & 0 \leq \alpha_i \leq C, \quad i = 1, 2, \ldots, m. \end{array}$$

12.3.12 ▪ A Geometric Programming Example

A *geometric programming* problem is an optimization problem of the form

$$\begin{array}{lll} \min & f(\mathbf{x}) & \\ \text{s.t.} & g_i(\mathbf{x}) \leq 1, & i = 1, 2, \ldots, m, \\ & h_j(\mathbf{x}) = 1, & j = 1, 2, \ldots, p, \\ & \mathbf{x} \in \mathbb{R}^n_{++}, & \end{array}$$

where $f, g_1, g_2, \ldots, g_m$ are *posynomials* and $h_1, h_2, \ldots, h_p$ are *monomials*. In the context of geometric programming, a monomial is a function $\phi : \mathbb{R}^n_{++} \to \mathbb{R}$ of the form $\phi(\mathbf{x}) = c\Pi_{j=1}^n x_j^{\alpha_j}$, where $c > 0$ and $\alpha_1, \alpha_2, \ldots, \alpha_n \in \mathbb{R}$. A posynomial is a sum of monomials. Geometric programming problems are not convex, but they can be easily transformed into convex optimization problems. In addition, their dual can be explicitly derived by an elegant argument. Instead of showing the derivations in the most general setting, we will illustrate them using a simple example. Consider the geometric programming problem

$$\begin{array}{ll} \min & \frac{1}{t_1 t_2 t_3} + t_2 t_3 \\ \text{s.t.} & 2t_1 t_3 + t_1 t_2 \leq 4, \\ & t_1, t_2, t_3 > 0. \end{array}$$

We will make the transformation $t_i = e^{x_i}$, which transforms the problem into

$$\begin{array}{ll} \min_{x_1,x_2,x_3} & e^{-x_1-x_2-x_3} + e^{x_2+x_3} \\ \text{s.t.} & e^{\ln 2+x_1+x_3} + e^{x_1+x_2} \leq 4. \end{array}$$

The transformed problem is convex, and we have thus shown how to transform the non-convex geometric programming problem into a convex problem. To find a dual of this problem, we will consider the following equivalent problem:

$$\begin{array}{ll} \min & e^{y_1} + e^{y_2} \\ \text{s.t.} & e^{y_3} + e^{y_4} \leq 4, \\ & y_1 = -x_1 - x_2 - x_3, \\ & y_2 = x_2 + x_3, \\ & y_3 = x_1 + x_3 + \ln 2, \\ & y_4 = x_1 + x_2. \end{array} \tag{12.29}$$

We will now construct the Lagrangian. The first constraint will be associated with the nonnegative multiplier w, and with the four linear constraints, we associate the Lagrange multipliers u_1, u_2, u_3, u_4:

$$\begin{aligned} L(\mathbf{y}, \mathbf{x}, w, \mathbf{u}) &= e^{y_1} + e^{y_2} + w\,(e^{y_3} + e^{y_4} - 4) - u_1(y_1 + x_1 + x_2 + x_3) \\ &\quad - u_2(y_2 - x_2 - x_3) - u_3(y_3 - x_1 - x_3 - \ln 2) - u_4(y_4 - x_1 - x_2) \\ &= [e^{y_1} - u_1 y_1] + [e^{y_2} - u_2 y_2] \\ &\quad + [w e^{y_3} - u_3 y_3] + [w e^{y_4} - u_4 y_4] \\ &\quad - x_1(u_1 - u_3 - u_4) - x_2(u_1 - u_2 - u_4) - x_3(u_1 - u_2 - u_3) \\ &\quad + (\ln 2) u_3 - 4w. \end{aligned}$$

We will use the following simple and technical lemma. Note that we use the convention that $0 \ln 0 = 0$.

Lemma 12.17. *Let $\lambda \geq 0$ and $a \in \mathbb{R}$. Then*

$$\min_{y\in\mathbb{R}} [\lambda e^y - a y] = \begin{cases} a - a\ln\left(\frac{a}{\lambda}\right), & \lambda > 0, \quad a \geq 0, \\ 0, & \lambda = a = 0, \\ -\infty, & \lambda \geq 0, \quad a < 0, \\ -\infty, & \lambda = 0, \quad a > 0. \end{cases}$$

If $\lambda > 0$ and $a > 0$, then the optimal y is $y = \ln\left(\frac{a}{\lambda}\right)$.

Proof. If $\lambda = 0$, then obviously, the minimum is 0 if and only if $a = 0$, and otherwise it is $-\infty$. If $\lambda > 0$, then

$$\min_{y\in\mathbb{R}} [\lambda e^y - a y] = \lambda \min_{y\in\mathbb{R}} \left[e^y - \frac{a}{\lambda} y \right],$$

and the optimal solutions of both minimization problems are the same. If $a < 0$, then the minimum is $-\infty$ since taking $y \to -\infty$ we obtain that the objective function goes to $-\infty$. If $a = 0$ the (unattained) minimal value is 0. If $a > 0$, then the optimal solution is attained at the stationary point which is the solution to $e^y = \frac{a}{\lambda}$, that is, at $y = \ln\left(\frac{a}{\lambda}\right)$. Substituting this expression back to the objective function we obtain that the optimal value is

$$\lambda\left[\frac{a}{\lambda} - \frac{a}{\lambda}\ln\left(\frac{a}{\lambda}\right)\right] = a - a\ln\left(\frac{a}{\lambda}\right). \qquad \square$$

Based on Lemma 12.17 and the relations

$$\min_{x_1}[-x_1(u_1-u_3-u_4)] = \begin{cases} 0, & u_1-u_3-u_4=0, \\ -\infty & \text{else}, \end{cases}$$
$$\min_{x_2}[-x_2(u_1-u_2-u_4)] = \begin{cases} 0, & u_1-u_2-u_4=0, \\ -\infty & \text{else}, \end{cases}$$
$$\min_{x_3}[-x_3(u_1-u_2-u_3)] = \begin{cases} 0, & u_1-u_2-u_3=0, \\ -\infty & \text{else}, \end{cases}$$

we obtain that the dual problem is given by

$$\begin{array}{ll} \max & u_1 - u_1\ln u_1 + u_2 - u_2\ln u_2 + u_3 - u_3\ln\left(\frac{u_3}{w}\right) + u_4 - u_4\ln\left(\frac{u_4}{w}\right) + (\ln 2)u_3 - 4w \\ \text{s.t.} & u_1-u_3-u_4=0, \\ & u_1-u_2-u_4=0, \\ & u_1-u_2-u_3=0, \\ & u_1,u_2,u_3,u_4,w \geq 0. \end{array}$$

To make the problem well-defined and in order to be consistent with Lemma 12.17, the function $-u\ln\left(\frac{u}{w}\right)$ has the value 0 when $u=w=0$ and the value $-\infty$ when $u>0, w=0$.

It is interesting to note that we can actually solve the dual problem. Indeed, noting that the expression in the objective function that depends on w is

$$(u_3+u_4)\ln w - 4w,$$

we deduce that at an optimal solution $w = \frac{u_3+u_4}{4}$. The constraints of the dual problem imply that $u_2=u_3=u_4$ and that $u_1=2u_2$. Therefore, denoting the joint value of u_2,u_3,u_4 by $\alpha(\geq 0)$, we conclude that $u_1=2\alpha$ and $w=\frac{\alpha}{2}$. Plugging this into the dual, we obtain that the dual problem is reduced to the one-dimensional problem

$$\max\{3(1-\ln 2)\alpha - 3\alpha\ln(\alpha) : \alpha \geq 0\}.$$

The optimal solution of this problem is attained at the point at which the derivative vanishes,

$$3(1-\ln 2) - 3 - 3\ln\alpha = 0,$$

that is, at $\alpha = \frac{1}{2}$, and hence

$$u_1 = 1, \quad u_2=u_3=u_4=\frac{1}{2}, \quad w=\frac{1}{4}.$$

We can also find the optimal solution of the primal problem. For that, we will first compute the optimal y_1,y_2,y_3,y_4:

$$y_1 = \ln u_1 = 0, \quad y_2 = \ln u_2 = -\ln 2, \quad y_3 = \ln\left(\frac{u_3}{w}\right) = \ln 2, \quad y_4 = \ln\left(\frac{u_4}{w}\right) = \ln 2.$$

Hence, by the constraints of (12.29) we have

$$\begin{aligned} x_1+x_2+x_3 &= 0, \\ x_2+x_3 &= -\ln 2, \\ x_1+x_3 &= 0, \\ x_1+x_2 &= \ln 2, \end{aligned}$$

whose solution is $x_1 = \ln 2, x_2 = 0, x_3 = -\ln 2$. Therefore, the optimal solution of the primal problem is

$$t_1 = e^{x_1} = 2, \quad t_2 = e^{x_2} = 1, \quad t_3 = e^{x_3} = \frac{1}{2}.$$

Exercises

12.1. Find a dual problem to the convex problem

$$\begin{array}{ll} \min & x_1^2 + 0.5x_2^2 + x_1x_2 - 2x_1 - 3x_2 \\ \text{s.t.} & x_1 + x_2 \leq 1. \end{array}$$

Find the optimal solutions of both the dual and primal problems.

12.2. Write a dual problem to the problem

$$\begin{array}{ll} \min & x_1 - 4x_2 + x_3^4 \\ \text{s.t.} & x_1 + x_2 + x_3^2 \leq 2 \\ & x_1 \geq 0 \\ & x_2 \geq 0. \end{array}$$

Solve the dual problem.

12.3. Consider the problem

$$\begin{array}{ll} \min & x_1^2 + 2x_2^2 + 2x_1x_2 + x_1 - x_2 - x_3 \\ \text{s.t.} & x_1 + x_2 + x_3 \leq 1 \\ & x_3 \leq 1. \end{array}$$

(i) Is the problem convex?

(ii) Find an optimal solution of the problem.

(iii) Find a dual problem and solve it.

12.4. Consider the primal optimization problem

$$\begin{array}{ll} \min & x_1^4 - 2x_2^2 - x_2 \\ \text{s.t.} & x_1^2 + x_2^2 + x_2 \leq 0. \end{array}$$

(i) Is the problem convex?

(ii) Does there exist an optimal solution to the problem?

(iii) Write a dual problem. Solve it.

(iv) Is the optimal value of the dual problem equal to the optimal value of the primal problem? Find the optimal solution of the primal problem.

12.5. Consider the problem

$$\begin{array}{ll} \min & 3x_1^2 + x_1x_2 + 2x_2^2 \\ \text{s.t.} & 3x_1^2 + x_1x_2 + 2x_2^2 + x_1 - x_2 \geq 1 \\ & x_1 \geq 2x_2. \end{array}$$

(i) Is the problem convex?

(ii) Find a dual problem. Is the dual problem convex?

12.6. Find a dual to the convex optimization problem

$$\begin{array}{ll} \min & \sum_{i=1}^{n}(x_i \ln x_i - x_i) \\ \text{s.t.} & \mathbf{A}\mathbf{x} \leq \mathbf{b}, \\ & \mathbf{x} > \mathbf{0}, \end{array}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$.

12.7. Find a dual problem to the following convex minimization problem:

$$\begin{array}{ll} \min & \sum_{i=1}^{n}(a_i x_i^2 + 2b_i x_i + e^{\alpha_i x_i}) \\ \text{s.t.} & \sum_{i=1}^{n} x_i = 1, \end{array}$$

where $\mathbf{a}, \alpha \in \mathbb{R}^n_{++}, \mathbf{b} \in \mathbb{R}^n$.

12.8. Consider the convex optimization problem

$$\begin{array}{ll} \min & \sum_{j=1}^{n} x_j \ln \frac{x_j}{c_j} \\ \text{s.t.} & \mathbf{A}\mathbf{x} \geq \mathbf{b}, \\ & \sum_{j=1}^{n} x_j = 1 \\ & \mathbf{x} > \mathbf{0}, \end{array}$$

where $\mathbf{c} > \mathbf{0}, \mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{b} \in \mathbb{R}^m$. Find a dual problem.

12.9. Consider the following problem (also called *second order cone programming*):

$$\begin{array}{ll} \min & \mathbf{g}^T\mathbf{x} \\ \text{s.t.} & \|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\| \leq \mathbf{c}_i^T\mathbf{x} + d_i, \quad i = 1, 2, \ldots, k, \end{array}$$

where $\mathbf{g} \in \mathbb{R}^n, \mathbf{A}_i \in \mathbb{R}^{m_i\times n}, \mathbf{b}_i \in \mathbb{R}^{m_i}, \mathbf{c}_i \in \mathbb{R}^n, d_i \in \mathbb{R}, i = 1, 2, \ldots, k$. Find a dual problem.

12.10. Consider the primal optimization problem

$$\begin{array}{ll} \min & \sum_{j=1}^{n} \frac{c_j}{x_j} \\ \text{s.t.} & \mathbf{a}^T\mathbf{x} \leq b, \\ & \mathbf{x} > \mathbf{0}, \end{array}$$

where $\mathbf{a} \in \mathbb{R}^n_{++}, \mathbf{c} \in \mathbb{R}^n_{++}, b \in \mathbb{R}_{++}$.

(i) Find a dual problem with a single dual decision variable.

(ii) Solve the dual and primal problems.

12.11. Consider the following optimization problem in the variables $\alpha \in \mathbb{R}$ and $\mathbf{q} \in \mathbb{R}^n$:

$$\text{(P)} \quad \begin{array}{ll} \min & \alpha \\ \text{s.t.} & \mathbf{A}\mathbf{q} = \alpha\mathbf{f} \\ & \|\mathbf{q}\|^2 \leq \varepsilon, \end{array}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}, \mathbf{f} \in \mathbb{R}^m, \varepsilon \in \mathbb{R}_{++}$. Assume in addition that the rows of $\mathbf{A}$ are linearly independent.

(i) Explain why strong duality holds for problem (P).

(ii) Find a dual problem to problem (P). (Do not assign a Lagrange multiplier to the quadratic constraint.)

(iii) Solve the dual problem obtained in part (ii) and find the optimal solution of problem (P).

12.12. Let $\mathbf{a} \in \mathbb{R}^n_{++}$ and consider the problem

$$\begin{array}{ll} \min & \sum_{i=1}^n -\log(x_i + a_i) \\ \text{s.t.} & \sum_{i=1}^n x_i = 1 \\ & \mathbf{x} \geq \mathbf{0}. \end{array}$$

Find a dual problem with one dual decision variable. Is strong duality satisfied?

12.13. Let $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \in \mathbb{R}^n$, and consider the Fermat–Weber problem

$$\min \{\|\mathbf{x}-\mathbf{a}_1\| + \|\mathbf{x}-\mathbf{a}_2\| + \|\mathbf{x}-\mathbf{a}_3\|\}.$$

Find a dual problem.

12.14. Find a dual problem to the following primal one:

$$\begin{array}{ll} \min & \sum_{i=1}^n x_i \ln\left(\frac{x_i}{\alpha_i}\right) + \|\mathbf{x}\|^2 + 2\mathbf{a}^T\mathbf{x} \\ \text{s.t.} & \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c \leq 0, \end{array}$$

where $\alpha \in \mathbb{R}^n_{++}, \mathbf{a} \in \mathbb{R}^n, \mathbf{A} \succ \mathbf{0}, \mathbf{b} \in \mathbb{R}^n, c \in \mathbb{R}$. Under what condition is strong duality guaranteed to hold? Find a condition that is written explicitly in terms of the data.

12.15. Let $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_m \in \mathbb{R}^n$ and $b_1, b_2, \ldots, b_m \in \mathbb{R}$ and consider the problem of finding the so-called analytic center of the polytope $S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{a}_i^T\mathbf{x} < b_i, i = 1, 2, \ldots, m\}$ given by

$$\text{(A)} \quad \min \left\{ -\sum_{i=1}^m \ln(b_i - \mathbf{a}_i^T\mathbf{x}) : \mathbf{x} \in S \right\}.$$

Find a dual problem to (A).

12.16. Let $f : \mathbb{R}^n \to \mathbb{R}$ be defined as (k is a positive integer smaller than n)

$$f(\mathbf{x}) = \sum_{i=1}^k x_{[i]},$$

where $x_{[i]}$ is the ith largest values in the vector $\mathbf{x}$. We have seen in Example 7.27 that f is convex.

(i) Show that for any $\mathbf{x} \in \mathbb{R}^n$, we have that $f(\mathbf{x})$ is the optimal value of the problem

$$\text{(P)} \quad \begin{array}{ll} \max & \mathbf{x}^T\mathbf{y} \\ \text{s.t.} & \mathbf{e}^T\mathbf{y} = k \\ & \mathbf{0} \leq \mathbf{y} \leq \mathbf{e}. \end{array}$$

(ii) For any $\alpha \in \mathbb{R}$ show that $f(\mathbf{x}) \leq \alpha$ if and only if there exist $\lambda \in \mathbb{R}^n_+$ and $u \in \mathbb{R}$ such that

$$ku + \mathbf{e}^T\lambda \leq \alpha, u\mathbf{e} + \lambda \geq \mathbf{x}.$$

(iii) Let $\mathbf{Q} \in \mathbb{R}^{n \times n}$ be a positive definite matrix. Find a dual to the problem

$$\begin{array}{ll} \min & \mathbf{x}^T\mathbf{Q}\mathbf{x} \\ \text{s.t.} & f(\mathbf{x}) \leq \alpha. \end{array}$$

12.17. Consider the inequality constrained problem

$$\begin{array}{rll} f^* = & \min & f(\mathbf{x}) \\ & \text{s.t.} & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \ldots, m, \end{array}$$

where $f, g_1, g_2, \ldots, g_m$ are convex functions over $\mathbb{R}^n$. Suppose that there exists $\hat{\mathbf{x}} \in \mathbb{R}^n$ such that

$$g_i(\hat{\mathbf{x}}) < 0, \quad i = 1, 2, \ldots, m,$$

and that the problem is bounded below, i.e., $f^* > -\infty$. Consider also the dual problem given by

$$\max\{q(\lambda) : \lambda \in \text{dom}(q)\},$$

where $q(\lambda) = \min_{\mathbf{x}} L(\mathbf{x}, \lambda), \text{dom}(q) = \{\lambda \in \mathbb{R}^m_+ : \min L(\mathbf{x}, \lambda) > -\infty\}$. Let λ^* be an optimal solution of the dual problem. Prove that

$$\sum_{i=1}^m \lambda_i^* \leq \frac{f(\hat{\mathbf{x}}) - f^*}{\min_{i=1,2,\ldots,m}(-g_i(\hat{\mathbf{x}}))}.$$

12.18. Consider the optimization problem (with the convention that $0 \ln 0 = 0$)

$$\begin{array}{ll} \min & x_1 + 2x_2 + 3x_3 + 4x_4 + \sum_{i=1}^4 x_i \ln x_i \\ \text{s.t.} & \sum_{i=1}^4 x_i = 1 \\ & x_1, x_2, x_3, x_4 \geq 0. \end{array}$$

(i) Show that the problem cannot have more than one optimal solution.

(ii) Find a dual problem in one dual decision variable.

(iii) Solve the dual problem.

(iv) Find the optimal solution of the primal problem.

12.19. Find a dual problem for the optimization problem

$$\begin{array}{ll} \min & x_1 + 2x_2^2 + x_3^2 + \sqrt{4x_1^2 + x_1x_3 + x_3^2 + 2} \\ \text{s.t.} & x_1 + x_2^2 + 4x_3^2 \leq 5 \\ & x_1 + x_2 + x_3 \leq 15. \end{array}$$

12.20. Consider the problem

$$\begin{array}{ll} \min & x_1^2 - x_2^4 + x_3^2 + \sqrt{x_1^2 + x_3^2} \\ \text{s.t.} & x_1^2 + x_2^4 + x_3^2 + x_3 \leq 4. \end{array}$$

(i) Find a dual problem.

(ii) Is the dual problem convex?

12.21. Consider the optimization problem

$$\text{(P)} \quad \begin{array}{ll} \min & -6x_1 + 2x_2 + 4x_3^2 \\ \text{s.t.} & 2x_1 + 2x_2 + x_3 \leq 0 \\ & -2x_1 + 4x_2 + x_3^2 = 0 \\ & x_2 \geq 0 \end{array}$$

(i) Is the problem convex?

(ii) Find a dual problem for (P). Do not assign a Lagrange multiplier to the constraint $x_2 \geq 0$.

(iii) Find the optimal solution of the dual problem.

12.22. Let $\alpha \in \mathbb{R}_{++}$ and define the set

$$T_\alpha = \left\{ \mathbf{x} \in \mathbb{R}^n : \sum_{j=1}^{n} x_j = 1, 0 \leq x_j \leq \alpha \right\}.$$

(i) For which values of α is the set T_α nonempty?

(ii) Find a dual problem with one dual decision variable to the problem of finding the orthogonal projection of a given vector $\mathbf{y} \in \mathbb{R}^n$ onto T_α.

(iii) Write a MATLAB function for computing the orthogonal projection onto the set T_α based on the dual problem found in part (ii). The call to the function will be in the form

```
function xp=proj_bound(y,alpha)
```

where y is the point which should be projected onto T_α and xp is the resulting projection. The function should check whether the set T_α is nonempty.

(iv) Write a function for finding the orthogonal projection onto T_α which is based on CVX. The function call will be in the form

```
function xp=proj_bound_cvx(y,alpha)
```

(v) Compute the orthogonal projection of the vector $(0.5, 0.7, 0.1, 0.3, 0.1)^T$ onto $T_{0.3}$ using the MATLAB functions constructed in parts (iii) and (iv).

(vi) Compare the CPU times of the two functions on a problem with 10000 variables using the following commands:

```
rand('seed',314);
x=rand(10000,1);
tic,y=proj_bound(x,0.01);toc
tic,y=proj_bound_cvx(x,0.01);toc
```

12.23. Consider the following optimization problem:

$$\text{(P)} \quad \min\left\{ h(\mathbf{x}) \equiv \|\mathbf{x}-\mathbf{d}\|^2 + \sqrt{x_1^2+x_2^2} + \sqrt{x_2^2+x_3^2} + \sqrt{x_3^2+x_4^2} + \sqrt{x_4^2+x_5^2} : \mathbf{x} \in \mathbb{R}^5 \right\},$$

where $\mathbf{d} = (1,2,3,2,1)^T$.

(a) Find an explicit dual problem of (P) with a "simple" constraint set (meaning a set on which it is easy to compute the orthogonal projection operator).

(b) Run 10 iterations of the gradient projection method on the derived dual problem. Use a constant stepsize $\frac{1}{L}$, where L is the corresponding Lipschitz constant. You need to show at each iteration both the dual objective function value as well as the objective function of the primal problem (use the relations between the optimal primal and dual solutions to derive at each iteration a primal solution). Finally, write explicitly the optimal solution of the problem.