1
Convex Optimization Models: An Overview

Contents

In this chapter we provide an overview of some broad classes of convex optimization models. Our primary focus will be on large challenging problems, often connected in some way to duality. We will consider two types of duality. The first is *Lagrange duality* for constrained optimization, which is obtained by assigning dual variables to the constraints. The second is *Fenchel duality* together with its special case, conic duality, which involves a cost function that is the sum of two convex function components. Both of these duality structures arise often in applications, and in Sections 1.1 and 1.2 we provide an overview, and discuss some examples.†

In Sections 1.3 and 1.4, we discuss additional model structures involving a large number of additive terms in the cost, or a large number of constraints. These types of problems also arise often in the context of duality, as well as in other contexts such as machine learning and signal processing with large amounts of data. In Section 1.5, we discuss the exact penalty function technique, whereby we can transform a convex constrained optimization problem to an equivalent unconstrained problem.

1.1 LAGRANGE DUALITY

We start our overview of Lagrange duality with the basic case of nonlinear inequality constraints, and then consider extensions involving linear inequality and equality constraints. Consider the problem‡

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad g(x) \leq 0, \end{aligned} \tag{1.1}$$

where X is a nonempty set,

$$g(x) = \big(g_1(x), \ldots, g_r(x)\big)',$$

and $f : X \mapsto \Re$ and $g_j : X \mapsto \Re$, $j = 1, \ldots, r$, are given functions. We refer to this as the *primal problem*, and we denote its optimal value by f^*. A vector x satisfying the constraints of the problem is referred to as *feasible*. The *dual* of problem (1.1) is given by

$$\begin{aligned} &\text{maximize} \quad q(\mu) \\ &\text{subject to} \quad \mu \in \Re^r, \end{aligned} \tag{1.2}$$

† Consistent with its overview character, this chapter contains few proofs, and refers frequently to the literature, and to Appendix B, which contains a full list of definitions and propositions (without proofs) relating to nonalgorithmic aspects of convex optimization. This list reflects and summarizes the content of the author's "Convex Optimization Theory" book [Ber09]. The proposition numbers of [Ber09] have been preserved, so all omitted proofs of propositions in Appendix B can be readily accessed from [Ber09].

‡ Appendix A contains an overview of the mathematical notation, terminology, and results from linear algebra and real analysis that we will be using.

where the dual function q is

$$q(\mu) = \begin{cases} \inf_{x \in X} L(x, \mu) & \text{if } \mu \geq 0, \\ -\infty & \text{otherwise,} \end{cases}$$

and L is the Lagrangian function defined by

$$L(x, \mu) = f(x) + \mu' g(x), \qquad x \in X, \ \mu \in \Re^r;$$

(cf. Section 5.3 of Appendix B).

Note that the dual function is extended real-valued, and that the effective constraint set of the dual problem is

$$\left\{ \mu \geq 0 \;\middle|\; \inf_{x \in X} L(x, \mu) > -\infty \right\}.$$

The optimal value of the dual problem is denoted by q^*.

The *weak duality* relation, $q^* \leq f^*$, always holds. It is easily shown by writing for all $\mu \geq 0$, and $x \in X$ with $g(x) \leq 0$,

$$q(\mu) = \inf_{z \in X} L(z, \mu) \leq L(x, \mu) = f(x) + \sum_{j=1}^{r} \mu_j g_j(x) \leq f(x),$$

so that

$$q^* = \sup_{\mu \in \Re^r} q(\mu) = \sup_{\mu \geq 0} q(\mu) \leq \inf_{x \in X,\, g(x) \leq 0} f(x) = f^*.$$

We state this formally as follows (cf. Prop. 4.1.2 in Appendix B).

Proposition 1.1.1: (Weak Duality Theorem) Consider problem (1.1). For any feasible solution x and any $\mu \in \Re^r$, we have $q(\mu) \leq f(x)$. Moreover, $q^* \leq f^*$.

When $q^* = f^*$, we say that *strong duality* holds. The following proposition gives necessary and sufficient conditions for strong duality, and primal and dual optimality (see Prop. 5.3.2 in Appendix B).

Proposition 1.1.2: (Optimality Conditions) Consider problem (1.1). There holds $q^* = f^*$, and (x^*, μ^*) are a primal and dual optimal solution pair if and only if x^* is feasible, $\mu^* \geq 0$, and

$$x^* \in \arg\min_{x \in X} L(x, \mu^*), \qquad \mu_j^* g_j(x^*) = 0, \quad j = 1, \ldots, r.$$

Both of the preceding propositions do not require any convexity assumptions on f, g, and X. However, generally the analytical and algorithmic solution process is simplified when strong duality ($q^* = f^*$) holds. This typically requires convexity assumptions, and in some cases conditions on $\text{ri}(X)$, the relative interior of X, as exemplified by the following result, given in Prop. 5.3.1 in Appendix B. The result delineates the two principal cases where there is no duality gap in an inequality-constrained problem.

Proposition 1.1.3: (Strong Duality – Existence of Dual Optimal Solutions) Consider problem (1.1) under the assumption that the set X is convex, and the functions f, and $g_1, \ldots, g_r$ are convex. Assume further that f^* is finite, and that one of the following two conditions holds:

(1) There exists $\overline{x} \in X$ such that $g_j(\overline{x}) < 0$ for all $j = 1, \ldots, r$.

(2) The functions g_j, $j = 1, \ldots, r$, are affine, and there exists $\overline{x} \in \text{ri}(X)$ such that $g(\overline{x}) \leq 0$.

Then $q^* = f^*$ and there exists at least one dual optimal solution. Under condition (1) the set of dual optimal solutions is also compact.

Convex Programming with Inequality and Equality Constraints

Let us consider an extension of problem (1.1), with additional linear equality constraints. It is our principal constrained optimization model under convexity assumptions, and it will be referred to as the *convex programming problem*. It is given by

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \ \ x \in X, \quad g(x) \leq 0, \quad Ax = b, \end{aligned} \tag{1.3}$$

where X is a convex set, $g(x) = \big(g_1(x), \ldots, g_r(x)\big)'$, $f : X \mapsto \Re$ and $g_j : X \mapsto \Re$, $j = 1, \ldots, r$, are given convex functions, A is an $m \times n$ matrix, and $b \in \Re^m$.

The preceding duality framework may be applied to this problem by converting the constraint $Ax = b$ to the equivalent set of linear inequality constraints

$$Ax \leq b, \qquad -Ax \leq -b,$$

with corresponding dual variables $\lambda^+ \geq 0$ and $\lambda^- \geq 0$. The Lagrangian function is

$$f(x) + \mu' g(x) + (\lambda^+ - \lambda^-)'(Ax - b),$$

and by introducing a dual variable

$$\lambda = \lambda^+ - \lambda^-$$

with no sign restriction, it can be written as

$$L(x, \mu, \lambda) = f(x) + \mu' g(x) + \lambda'(Ax - b).$$

The dual problem is

$$\begin{aligned} &\text{maximize} && \inf_{x \in X} L(x, \mu, \lambda) \\ &\text{subject to} && \mu \geq 0, \ \lambda \in \Re^m. \end{aligned}$$

In this manner, Prop. 1.1.3 under condition (2), together with Prop. 1.1.2, yield the following for the case where all constraint functions are linear.

Proposition 1.1.4: (Convex Programming – Linear Equality and Inequality Constraints) Consider problem (1.3).

(a) Assume that f^* is finite, that the functions g_j are affine, and that there exists $\overline{x} \in \text{ri}(X)$ such that $A\overline{x} = b$ and $g(\overline{x}) \leq 0$. Then $q^* = f^*$ and there exists at least one dual optimal solution.

(b) There holds $f^* = q^*$, and (x^*, μ^*, λ^*) are a primal and dual optimal solution pair if and only if x^* is feasible, $\mu^* \geq 0$, and

$$x^* \in \arg\min_{x \in X} L(x, \mu^*, \lambda^*), \qquad \mu_j^* g_j(x^*) = 0, \quad j = 1, \ldots, r.$$

In the special case where there are no inequality constraints:

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && x \in X, \quad Ax = b, \end{aligned} \tag{1.4}$$

the Lagrangian function is

$$L(x, \lambda) = f(x) + \lambda'(Ax - b),$$

and the dual problem is

$$\begin{aligned} &\text{maximize} && \inf_{x \in X} L(x, \lambda) \\ &\text{subject to} && \lambda \in \Re^m. \end{aligned}$$

The corresponding result, a simpler special case of Prop. 1.1.4, is given in the following proposition.

Proposition 1.1.5: (Convex Programming – Linear Equality Constraints) Consider problem (1.4).

(a) Assume that f^* is finite and that there exists $\overline{x} \in \text{ri}(X)$ such that $A\overline{x} = b$. Then $f^* = q^*$ and there exists at least one dual optimal solution.

(b) There holds $f^* = q^*$, and (x^*, λ^*) are a primal and dual optimal solution pair if and only if x^* is feasible and

$$x^* \in \arg\min_{x \in X} L(x, \lambda^*).$$

The following is an extension of Prop. 1.1.4(a) to the case where the inequality constraints may be nonlinear. It is the most general convex programming result relating to duality in this section (see Prop. 5.3.5 in Appendix B).

Proposition 1.1.6: (Convex Programming – Linear Equality and Nonlinear Inequality Constraints) Consider problem (1.3). Assume that f^* is finite, that there exists $\overline{x} \in X$ such that $A\overline{x} = b$ and $g(\overline{x}) < 0$, and that there exists $\tilde{x} \in \text{ri}(X)$ such that $A\tilde{x} = b$. Then $q^* = f^*$ and there exists at least one dual optimal solution.

Aside from the preceding results, there are alternative optimality conditions for convex and nonconvex optimization problems, which are based on extended versions of the Fritz John theorem; see [BeO02] and [BOT06], and the textbooks [Ber99] and [BNO03]. These conditions are derived using a somewhat different line of analysis and supplement the ones given here, but we will not have occasion to use them in this book.

Discrete Optimization and Lower Bounds

The preceding propositions deal mostly with situations where strong duality holds ($q^* = f^*$). However, duality can be useful even when there is duality gap, as often occurs in problems that have a finite constraint set X. An example is *integer programming*, where the components of x must be integers from a bounded range (usually 0 or 1). An important special case is the linear 0-1 integer programming problem

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \quad Ax \le b, \quad x_i = 0 \text{ or } 1, \quad i = 1, \ldots, n, \end{aligned}$$

where $x = (x_1, \ldots, x_n)$.

A principal approach for solving discrete optimization problems with a finite constraint set is the *branch-and-bound method*, which is described in many sources; see e.g., one of the original works [LaD60], the survey [BaT85], and the book [NeW88]. The general idea of the method is that bounds on the cost function can be used to exclude from consideration portions of the feasible set. To illustrate, consider minimizing $F(x)$ over $x \in X$, and let Y_1, Y_2 be two subsets of X. Suppose that we have bounds

$$\underline{F}_1 \leq \min_{x \in Y_1} f(x), \qquad \overline{F}_2 \geq \min_{x \in Y_2} f(x).$$

Then, if $\overline{F}_2 \leq \underline{F}_1$, the solutions in Y_1 may be disregarded since their cost cannot be smaller than the cost of the best solution in Y_2. The lower bound $\underline{F}_1$ can often be conveniently obtained by minimizing f over a suitably enlarged version of Y_1, while for the upper bound $\overline{F}_2$, a value $f(x)$, where $x \in Y_2$, may be used.

Branch-and-bound is often based on weak duality (cf. Prop. 1.1.1) to obtain lower bounds to the optimal cost of restricted problems of the form

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in \tilde{X}, \qquad g(x) \leq 0, \end{aligned} \tag{1.5}$$

where $\tilde{X}$ is a subset of X; for example in the 0-1 integer case where X specifies that all x_i should be 0 or 1, $\tilde{X}$ may be the set of all 0-1 vectors x such that one or more components x_i are fixed at either 0 or 1 (i.e., are restricted to satisfy $x_i = 0$ for all $x \in \tilde{X}$ or $x_i = 1$ for all $x \in \tilde{X}$). These lower bounds can often be obtained by finding a dual-feasible (possibly dual-optimal) solution $\mu \geq 0$ of this problem and the corresponding dual value

$$q(\mu) = \inf_{x \in \tilde{X}} \big\{ f(x) + \mu' g(x) \big\}, \tag{1.6}$$

which by weak duality, is a lower bound to the optimal value of the restricted problem (1.5). In a strengthened version of this approach, the given inequality constraints $g(x) \leq 0$ may be augmented by additional inequalities that are known to be satisfied by optimal solutions of the original problem.

An important point here is that when $\tilde{X}$ is finite, the dual function q of Eq. (1.6) is concave and polyhedral. Thus solving the dual problem amounts to minimizing the polyhedral function $-q$ over the nonnegative orthant. This is a major context within which polyhedral functions arise in convex optimization.

1.1.1 Separable Problems – Decomposition

Let us now discuss an important problem structure that involves Lagrange duality and arises frequently in applications. Here x has m components,

$x = (x_1, \ldots, x_m)$, with each x_i being a vector of dimension n_i (often $n_i = 1$). The problem has the form

$$\begin{aligned} &\text{minimize} \quad \sum_{i=1}^{m} f_i(x_i) \\ &\text{subject to} \quad \sum_{i=1}^{m} g_{ij}(x_i) \leq 0, \quad x_i \in X_i, \quad i = 1, \ldots, m, \ j = 1, \ldots, r, \end{aligned} \tag{1.7}$$

where $f_i : \Re^{n_i} \mapsto \Re$ and $g_{ij} : \Re^{n_i} \mapsto \Re^r$ are given functions, and X_i are given subsets of $\Re^{n_i}$. By assigning a dual variable μ_j to the jth constraint, we obtain the dual problem [cf. Eq. (1.2)]

$$\begin{aligned} &\text{maximize} \quad \sum_{i=1}^{m} q_i(\mu) \\ &\text{subject to} \quad \mu \geq 0, \end{aligned} \tag{1.8}$$

where

$$q_i(\mu) = \inf_{x_i \in X_i} \left\{ f_i(x_i) + \sum_{j=1}^{r} \mu_j g_{ij}(x_i) \right\},$$

and $\mu = (\mu_1, \ldots, \mu_r)$.

Note that the minimization involved in the calculation of the dual function has been decomposed into m simpler minimizations. These minimizations are often conveniently done either analytically or computationally, in which case the dual function can be easily evaluated. This is the key advantageous structure of separable problems: it facilitates computation of dual function values (as well as subgradients as we will see in Section 3.1), and it is amenable to decomposition and distributed computation.

Let us also note that in the special case where the components x_i are one-dimensional, and the functions f_i and sets X_i are convex, there is a particularly favorable duality result for the separable problem (1.7): essentially, strong duality holds without any qualifications such as the linearity of the constraint functions, or the Slater condition of Prop. 1.1.3; see [Tse09].

Duality Gap Estimates for Nonconvex Separable Problems

The separable structure is additionally helpful when the cost and/or the constraints are not convex, and there is a duality gap. In particular, in this case *the duality gap turns out to be relatively small and can often be shown to diminish to zero relative to the optimal primal value as the number m of separable terms increases*. As a result, one can often obtain a near-optimal primal solution, starting from a dual-optimal solution, without resorting to costly branch-and-bound procedures.

The small duality gap size is a consequence of the structure of the set S of constraint-cost pairs of problem (1.7), which in the case of a separable problem, can be written as a vector sum of m sets, one for each separable term, i.e.,

$$S = S_1 + \cdots + S_m,$$

where

$$S_i = \big\{ \big(g_i(x_i), f_i(x_i)\big) \mid x_i \in X_i \big\},$$

and $g_i : \Re^{n_i} \mapsto \Re^r$ is the function $g_i(x_i) = \big(g_{i1}(x_i), \ldots, g_{im}(x_i)\big)$. It can be shown that the duality gap is related to how much S "differs" from its convex hull (a geometric explanation is given in [Ber99], Section 5.1.6, and [Ber09], Section 5.7). Generally, a set that is the vector sum of a large number of possibly nonconvex but roughly similar sets "tends to be convex" in the sense that any vector in its convex hull can be closely approximated by a vector in the set. As a result, the duality gap tends to be relatively small. The analytical substantiation is based on a theorem by Shapley and Folkman (see [Ber99], Section 5.1, or [Ber09], Prop. 5.7.1, for a statement and proof of this theorem). In particular, it is shown in [AuE76], and also [BeS82], [Ber82a], Section 5.6.1, under various reasonable assumptions, that the duality gap satisfies

$$f^* - q^* \le (r+1) \max_{i=1,\ldots,m} \rho_i,$$

where for each i, ρ_i is a nonnegative scalar that depends on the structure of the functions f_i, g_{ij}, $j = 1, \ldots, r$, and the set X_i (the paper [AuE76] focuses on the case where the problem is nonconvex but continuous, while [BeS82] and [Ber82a] focus on an important class of mixed integer programming problems). This estimate suggests that as $m \to \infty$ and $|f^*| \to \infty$, the duality gap is bounded, while the "relative" duality gap $(f^* - q^*)/|f^*|$ diminishes to 0 as $m \to \infty$.

The duality gap has also been investigated in the author's book [Ber09] within the more general min common-max crossing framework (Section 4.1 of Appendix B). This framework includes as special cases minimax and zero-sum game problems. In particular, consider a function $\phi : X \times Z \mapsto \Re$ defined over nonempty subsets $X \subset \Re^n$ and $Z \subset \Re^m$. Then it can be shown that the gap between "infsup" and "supinf" of ϕ can be decomposed into the sum of two terms that can be computed separately: one term can be attributed to the lack of convexity and/or closure of ϕ with respect to x, and the other can be attributed to the lack of concavity and/or upper semicontinuity of ϕ with respect to z. We refer to [Ber09], Section 5.7.2, for the analysis.

1.1.2 Partitioning

It is important to note that there are several different ways to introduce duality in the solution of large-scale optimization problems. For example a strategy, often called *partitioning*, is to divide the variables in two subsets, and minimize first with respect to one subset while taking advantage of whatever simplification may arise by fixing the variables in the other subset.

As an example, the problem

$$\begin{aligned} &\text{minimize} \;\; F(x) + G(y) \\ &\text{subject to} \;\; Ax + By = c, \quad x \in X, \;\; y \in Y, \end{aligned}$$

can be written as

$$\begin{aligned} &\text{minimize} \;\; F(x) + \inf_{By=c-Ax,\, y\in Y} G(y) \\ &\text{subject to} \;\; x \in X, \end{aligned}$$

or

$$\begin{aligned} &\text{minimize} \;\; F(x) + p(c - Ax) \\ &\text{subject to} \;\; x \in X, \end{aligned}$$

where p is given by

$$p(u) = \inf_{By=u,\, y\in Y} G(y).$$

In favorable cases, p can be dealt with conveniently (see e.g., the book [Las70] and the paper [Geo72]).

Strategies of splitting or transforming the variables to facilitate algorithmic solution will be frequently encountered in what follows, and in a variety of contexts, including duality. The next section describes some significant contexts of this type.

1.2 FENCHEL DUALITY AND CONIC PROGRAMMING

Let us consider the Fenchel duality framework (see Section 5.3.5 of Appendix B). It involves the problem

$$\begin{aligned} &\text{minimize} \quad f_1(x) + f_2(Ax) \\ &\text{subject to} \;\; x \in \Re^n, \end{aligned} \tag{1.9}$$

where A is an $m \times n$ matrix, $f_1 : \Re^n \mapsto (-\infty, \infty]$ and $f_2 : \Re^m \mapsto (-\infty, \infty]$ are closed proper convex functions, and we assume that there exists a feasible solution, i.e., an $x \in \Re^n$ such that $x \in \text{dom}(f_1)$ and $Ax \in \text{dom}(f_2)$.†

The problem is equivalent to the following constrained optimization problem in the variables $x_1 \in \Re^n$ and $x_2 \in \Re^m$:

$$\begin{aligned} &\text{minimize} \quad f_1(x_1) + f_2(x_2) \\ &\text{subject to} \;\; x_1 \in \text{dom}(f_1), \;\; x_2 \in \text{dom}(f_2), \qquad x_2 = Ax_1. \end{aligned} \tag{1.10}$$

† We remind the reader that our convex analysis notation, terminology, and nonalgorithmic theory are summarized in Appendix B.

Viewing this as a convex programming problem with the linear equality constraint $x_2 = Ax_1$, we obtain the dual function as

$$\begin{aligned} q(\lambda) &= \inf_{x_1 \in \mathrm{dom}(f_1),\ x_2 \in \mathrm{dom}(f_2)} \big\{ f_1(x_1) + f_2(x_2) + \lambda'(x_2 - Ax_1) \big\} \\ &= \inf_{x_1 \in \Re^n} \big\{ f_1(x_1) - \lambda' A x_1 \big\} + \inf_{x_2 \in \Re^m} \big\{ f_2(x_2) + \lambda' x_2 \big\}. \end{aligned}$$

The dual problem of maximizing q over $\lambda \in \Re^m$, after a sign change to convert it to a minimization problem, takes the form

$$\begin{aligned} &\text{minimize} \quad f_1^\star(A'\lambda) + f_2^\star(-\lambda) \\ &\text{subject to} \quad \lambda \in \Re^m, \end{aligned} \tag{1.11}$$

where $f_1^\star$ and $f_2^\star$ are the conjugate functions of f_1 and f_2. We denote by f^* and q^* the corresponding optimal primal and dual values [q^* is the negative of the optimal value of problem (1.11)].

The following Fenchel duality result is given as Prop. 5.3.8 in Appendix B. Parts (a) and (b) are obtained by applying Prop. 1.1.5(a) to problem (1.10), viewed as a problem with $x_2 = Ax_1$ as the only linear equality constraint. The first equation of part (c) is a consequence of Prop. 1.1.5(b). Its equivalence with the last two equations is a consequence of the Conjugate Subgradient Theorem (Prop. 5.4.3, App. B), which states that for a closed proper convex function f, its conjugate $f^\star$, and any pair of vectors (x, y), we have

$$x \in \arg\min_{z \in \Re^n} \big\{ f(z) - z'y \big\} \quad \text{iff} \quad y \in \partial f(x) \quad \text{iff} \quad x \in \partial f^\star(y),$$

with all of these three relations being equivalent to $x'y = f(x) + f^\star(y)$. Here $\partial f(x)$ denotes the subdifferential of f at x (the set of all subgradients of f at x); see Section 5.4 of Appendix B.

Proposition 1.2.1: (Fenchel Duality) Consider problem (1.9).

(a) If f^* is finite and $\big(A \cdot \mathrm{ri}\big(\mathrm{dom}(f_1)\big)\big) \cap \mathrm{ri}\big(\mathrm{dom}(f_2)\big) \neq \emptyset$, then $f^* = q^*$ and there exists at least one dual optimal solution.

(b) If q^* is finite and $\mathrm{ri}\big(\mathrm{dom}(f_1^\star)\big) \cap \big(A' \cdot \mathrm{ri}\big(-\mathrm{dom}(f_2^\star)\big)\big) \neq \emptyset$, then $f^* = q^*$ and there exists at least one primal optimal solution.

(c) There holds $f^* = q^*$, and (x^*, λ^*) is a primal and dual optimal solution pair if and only if any one of the following three equivalent conditions hold:

$$x^* \in \arg\min_{x \in \Re^n} \big\{ f_1(x) - x'A'\lambda^* \big\} \quad \text{and} \quad Ax^* \in \arg\min_{z \in \Re^m} \big\{ f_2(z) + z'\lambda^* \big\}, \tag{1.12}$$

$$A'\lambda^* \in \partial f_1(x^*) \quad \text{and} \quad -\lambda^* \in \partial f_2(Ax^*), \tag{1.13}$$

$$x^* \in \partial f_1^\star(A'\lambda^*) \quad \text{and} \quad Ax^* \in \partial f_2^\star(-\lambda^*). \tag{1.14}$$

Minimax Problems

Minimax problems involve minimization over a set X of a function $\overline{F}$ of the form

$$\overline{F}(x) = \sup_{z \in Z} \phi(x, z),$$

where X and Z are subsets of $\Re^n$ and $\Re^m$, respectively, and $\phi : \Re^n \times \Re^m \mapsto \Re$ is a given function. Some (but not all) problems of this type are related to constrained optimization and Fenchel duality.

Example 1.2.1: (Connection with Constrained Optimization)

Let ϕ and Z have the form

$$\phi(x, z) = f(x) + z' g(x), \qquad Z = \{z \mid z \geq 0\},$$

where $f : \Re^n \mapsto \Re$ and $g : \Re^n \mapsto \Re^m$ are given functions. Then it is seen that

$$\overline{F}(x) = \sup_{z \in Z} \phi(x, z) = \begin{cases} f(x) & \text{if } g(x) \leq 0, \\ \infty & \text{otherwise.} \end{cases}$$

Thus minimization of $\overline{F}$ over $x \in X$ is equivalent to solving the constrained optimization problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in X, \quad g(x) \leq 0. \end{aligned} \tag{1.15}$$

The dual problem is to maximize over $z \geq 0$ the function

$$\underline{F}(z) = \inf_{x \in X} \big\{ f(x) + z' g(x) \big\} = \inf_{x \in X} \phi(x, z),$$

and the minimax equality

$$\inf_{x \in X} \sup_{z \in Z} \phi(x, z) = \sup_{z \in Z} \inf_{x \in X} \phi(x, z) \tag{1.16}$$

is equivalent to problem (1.15) having no duality gap.

Example 1.2.2: (Connection with Fenchel Duality)

Let ϕ have the special form

$$\phi(x, z) = f(x) + z' A x - g(z),$$

where $f : \Re^n \mapsto \Re$ and $g : \Re^m \mapsto \Re$ are given functions, and A is a given $m \times n$ matrix. Then we have

$$\overline{F}(x) = \sup_{z \in Z} \phi(x, z) = f(x) + \sup_{z \in Z} \big\{ (Ax)' z - g(z) \big\} = f(x) + \hat{g}^\star(Ax),$$

where $\hat{g}^\star$ is the conjugate of the function

$$\hat{g}(z) = \begin{cases} g(z) & \text{if } z \in Z, \\ \infty & \text{otherwise.} \end{cases}$$

Thus the minimax problem of minimizing $\overline{F}$ over $x \in X$ comes under the Fenchel framework (1.9) with $f_2 = \hat{g}^\star$ and f_1 given by

$$f_1(x) = \begin{cases} f(x) & \text{if } x \in X, \\ \infty & \text{if } x \notin X. \end{cases}$$

It can also be verified that the Fenchel dual problem (1.11) is equivalent to maximizing over $z \in Z$ the function $\underline{F}(z) = \inf_{x \in X} \phi(x, z)$. Again having no duality gap is equivalent to the minimax equality (1.16) holding.

Finally note that strong duality theory is connected with minimax problems primarily when X and Z are convex sets, and ϕ is convex in x and concave in z. When Z is a finite set, there is a different connection with constrained optimization that does not involve Fenchel duality and applies without any convexity conditions. In particular, the problem

$$\begin{aligned} &\text{minimize} \quad \max\big\{g_1(x), \ldots, g_r(x)\big\} \\ &\text{subject to} \quad x \in X, \end{aligned}$$

where $g_j : \Re^n \mapsto \Re$ are any real-valued functions, is equivalent to the constrained optimization problem

$$\begin{aligned} &\text{minimize} \quad y \\ &\text{subject to} \quad x \in X, \qquad g_j(x) \le y, \quad j = 1, \ldots, r, \end{aligned}$$

where y is an additional scalar optimization variable. Minimax problems will be discussed further later, in Section 1.4, as an example of problems that may involve a large number of constraints.

Conic Programming

An important problem structure, which can be analyzed as a special case of the Fenchel duality framework is *conic programming*. This is the problem

$$\begin{aligned} &\text{minimize} \quad f(x) \\ &\text{subject to} \quad x \in C, \end{aligned} \tag{1.17}$$

where $f : \Re^n \mapsto (-\infty, \infty]$ is a closed proper convex function and C is a closed convex cone in $\Re^n$.

Indeed, let us apply Fenchel duality with A equal to the identity and the definitions

$$f_1(x) = f(x), \qquad f_2(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{if } x \notin C. \end{cases}$$

The corresponding conjugates are

$$f_1^\star(\lambda) = \sup_{x\in\Re^n}\big\{\lambda'x - f(x)\big\}, \qquad f_2^\star(\lambda) = \sup_{x\in C}\lambda'x = \begin{cases} 0 & \text{if } \lambda \in C^*, \\ \infty & \text{if } \lambda \notin C^*, \end{cases}$$

where

$$C^* = \{\lambda \mid \lambda'x \le 0,\ \forall\ x \in C\}$$

is the polar cone of C (note that $f_2^\star$ is the support function of C; cf. Section 1.6 of Appendix B). The dual problem is

$$\begin{aligned} &\text{minimize} \quad f^\star(\lambda) \\ &\text{subject to} \quad \lambda \in \hat{C}, \end{aligned} \tag{1.18}$$

where $f^\star$ is the conjugate of f and $\hat{C}$ is the negative polar cone (also called the *dual cone* of C):

$$\hat{C} = -C^* = \{\lambda \mid \lambda'x \ge 0,\ \forall\ x \in C\}.$$

Note the symmetry between primal and dual problems. The strong duality relation $f^* = q^*$ can be written as

$$\inf_{x\in C} f(x) = -\inf_{\lambda\in\hat{C}} f^\star(\lambda).$$

The following proposition translates the conditions of Prop. 1.2.1(a), which guarantees that there is no duality gap and that the dual problem has an optimal solution.

Proposition 1.2.2: (Conic Duality Theorem) Assume that the primal conic problem (1.17) has finite optimal value, and moreover $\text{ri}\big(\text{dom}(f)\big) \cap \text{ri}(C) \neq \emptyset$. Then, there is no duality gap and the dual problem (1.18) has an optimal solution.

Using the symmetry of the primal and dual problems, we also obtain that there is no duality gap and the primal problem (1.17) has an optimal solution if the optimal value of the dual conic problem (1.18) is finite and $\text{ri}\big(\text{dom}(f^\star)\big) \cap \text{ri}(\hat{C}) \neq \emptyset$. It is also possible to derive primal and dual optimality conditions by translating the optimality conditions of the Fenchel duality framework [Prop. 1.2.1(c)].

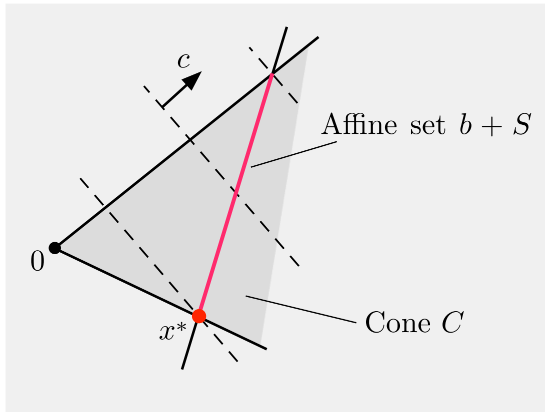

Figure 1.2.1. Illustration of a linear-conic problem: minimizing a linear function $c'x$ over the intersection of an affine set $b + S$ and a convex cone C.

1.2.1 Linear-Conic Problems

An important special case of conic programming, called *linear-conic problem*, arises when $\text{dom}(f)$ is an affine set and f is linear over $\text{dom}(f)$, i.e.,

$$f(x) = \begin{cases} c'x & \text{if } x \in b + S, \\ \infty & \text{if } x \notin b + S, \end{cases}$$

where b and c are given vectors, and S is a subspace. Then the primal problem can be written as

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \quad x - b \in S, \quad x \in C; \end{aligned} \tag{1.19}$$

see Fig. 1.2.1.

To derive the dual problem, we note that

$$\begin{aligned} f^\star(\lambda) &= \sup_{x-b\in S} (\lambda - c)'x \\ &= \sup_{y\in S} (\lambda - c)'(y + b) \\ &= \begin{cases} (\lambda - c)'b & \text{if } \lambda - c \in S^\perp, \\ \infty & \text{if } \lambda - c \notin S^\perp. \end{cases} \end{aligned}$$

It can be seen that the dual problem $\min_{\lambda \in \hat{C}} f^\star(\lambda)$ [cf. Eq. (1.18)], after discarding the superfluous term $c'b$ from the cost, can be written as

$$\begin{aligned} &\text{minimize} \quad b'\lambda \\ &\text{subject to} \quad \lambda - c \in S^\perp, \quad \lambda \in \hat{C}, \end{aligned} \tag{1.20}$$

where $\hat{C}$ is the dual cone:

$$\hat{C} = \{\lambda \mid \lambda' x \geq 0,\ \forall\ x \in C\}.$$

By specializing the conditions of the Conic Duality Theorem (Prop. 1.2.2) to the linear-conic duality context, we obtain the following.

Proposition 1.2.3: (Linear-Conic Duality Theorem) Assume that the primal problem (1.19) has finite optimal value, and moreover $(b+S)\cap \text{ri}(C) \neq \emptyset$. Then, there is no duality gap and the dual problem has an optimal solution.

Special Forms of Linear-Conic Problems

The primal and dual linear-conic problems (1.19) and (1.20) have been placed in an elegant symmetric form. There are also other useful formats that parallel and generalize similar formats in linear programming. For example, we have the following dual problem pairs:

$$\min_{Ax=b,\ x\in C} c'x \qquad \Longleftrightarrow \qquad \max_{c-A'\lambda\in \hat{C}} b'\lambda, \tag{1.21}$$

$$\min_{Ax-b\in C} c'x \qquad \Longleftrightarrow \qquad \max_{A'\lambda=c,\ \lambda\in \hat{C}} b'\lambda, \tag{1.22}$$

where A is an $m \times n$ matrix, and $x \in \Re^n$, $\lambda \in \Re^m$, $c \in \Re^n$, $b \in \Re^m$.

To verify the duality relation (1.21), let $\overline{x}$ be any vector such that $A\overline{x} = b$, and let us write the primal problem on the left in the primal conic form (1.19) as

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \quad x - \overline{x} \in \text{N}(A), \quad x \in C, \end{aligned}$$

where $\text{N}(A)$ is the nullspace of A. The corresponding dual conic problem (1.20) is to solve for μ the problem

$$\begin{aligned} &\text{minimize} \quad \overline{x}'\mu \\ &\text{subject to} \quad \mu - c \in \text{N}(A)^\perp, \quad \mu \in \hat{C}. \end{aligned} \tag{1.23}$$

Since $\text{N}(A)^\perp$ is equal to $\text{Ra}(A')$, the range of A', the constraints of problem (1.23) can be equivalently written as $c - \mu \in -\text{Ra}(A') = \text{Ra}(A')$, $\mu \in \hat{C}$, or

$$c - \mu = A'\lambda, \qquad \mu \in \hat{C},$$

for some $\lambda \in \Re^m$. Making the change of variables $\mu = c - A'\lambda$, the dual problem (1.23) can be written as

$$\begin{array}{ll} \text{minimize} & \overline{x}'(c - A'\lambda) \\ \text{subject to} & c - A'\lambda \in \hat{C}. \end{array}$$

By discarding the constant $\overline{x}'c$ from the cost function, using the fact $A\overline{x} = b$, and changing from minimization to maximization, we see that this dual problem is equivalent to the one in the right-hand side of the duality pair (1.21). The duality relation (1.22) is proved similarly.

We next discuss two important special cases of conic programming: *second order cone programming* and *semidefinite programming*. These problems involve two different special cones, and an explicit definition of the affine set constraint. They arise in a variety of applications, and their computational difficulty in practice tends to lie between that of linear and quadratic programming on one hand, and general convex programming on the other hand.

1.2.2 Second Order Cone Programming

In this section we consider the linear-conic problem (1.22), with the cone

$$C = \left\{ (x_1, \ldots, x_n) \;\middle|\; x_n \geq \sqrt{x_1^2 + \cdots + x_{n-1}^2} \right\},$$

which is known as the *second order cone* (see Fig. 1.2.2). The dual cone is

$$\hat{C} = \{ y \mid 0 \leq y'x, \ \forall\, x \in C \} = \left\{ y \;\middle|\; 0 \leq \inf_{\|(x_1,\ldots,x_{n-1})\| \leq x_n} y'x \right\},$$

and it can be shown that $\hat{C} = C$. This property is referred to as *self-duality* of the second order cone, and is fairly evident from Fig. 1.2.2. For a proof, we write

$$\begin{aligned} \inf_{\|(x_1,\ldots,x_{n-1})\| \leq x_n} y'x &= \inf_{x_n \geq 0} \left\{ y_n x_n + \inf_{\|(x_1,\ldots,x_{n-1})\| \leq x_n} \sum_{i=1}^{n-1} y_i x_i \right\} \\ &= \inf_{x_n \geq 0} \big\{ y_n x_n - \|(y_1, \ldots, y_{n-1})\|\, x_n \big\} \\ &= \begin{cases} 0 & \text{if } \|(y_1, \ldots, y_{n-1})\| \leq y_n, \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}$$

where the second equality follows because the minimum of the inner product of a vector $z \in \Re^{n-1}$ with vectors in the unit ball of $\Re^{n-1}$ is $-\|z\|$. Combining the preceding two relations, we have

$$y \in \hat{C} \quad \text{if and only if} \quad 0 \leq y_n - \|(y_1, \ldots, y_{n-1})\|,$$

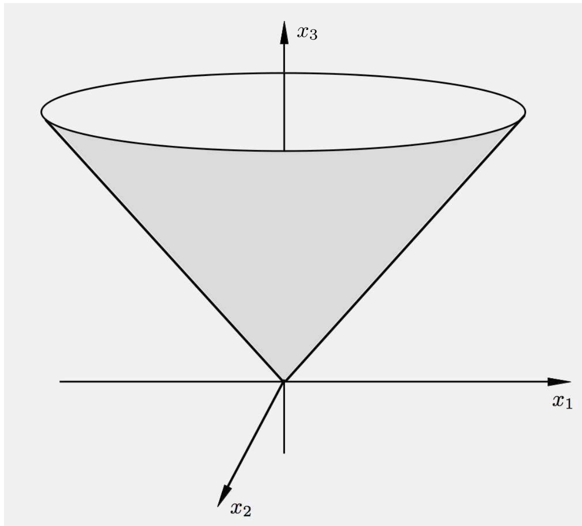

Figure 1.2.2. The second order cone

$$C = \left\{ (x_1, \ldots, x_n) \;\Big|\; x_n \geq \sqrt{x_1^2 + \cdots + x_{n-1}^2} \right\},$$

in $\Re^3$.

so $\hat{C} = C$.

The second order cone programming problem (SOCP for short) is

$$\begin{aligned} &\text{minimize} \quad c'x \\ &\text{subject to} \quad A_i x - b_i \in C_i, \ i = 1, \ldots, m, \end{aligned} \tag{1.24}$$

where $x \in \Re^n$, c is a vector in $\Re^n$, and for $i = 1, \ldots, m$, A_i is an $n_i \times n$ matrix, b_i is a vector in $\Re^{n_i}$, and C_i is the second order cone of $\Re^{n_i}$. It is seen to be a special case of the primal problem in the left-hand side of the duality relation (1.22), where

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_m \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}, \qquad C = C_1 \times \cdots \times C_m.$$

Note that linear inequality constraints of the form $a_i'x - b_i \geq 0$ can be written as

$$\begin{pmatrix} 0 \\ a_i' \end{pmatrix} x - \begin{pmatrix} 0 \\ b_i \end{pmatrix} \in C_i,$$

where C_i is the second order cone of $\Re^2$. As a result, linear-conic problems involving second order cones contain as special cases linear programming problems.

We now observe that from the right-hand side of the duality relation (1.22), and the self-duality relation $C = \hat{C}$, the corresponding dual linear-conic problem has the form

$$\begin{aligned} &\text{maximize} \quad \sum_{i=1}^{m} b_i'\lambda_i \\ &\text{subject to} \quad \sum_{i=1}^{m} A_i'\lambda_i = c, \quad \lambda_i \in C_i,\ i = 1, \ldots, m, \end{aligned} \tag{1.25}$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)$. By applying the Linear-Conic Duality Theorem (Prop. 1.2.3), we have the following.

Proposition 1.2.4: (Second Order Cone Duality Theorem) Consider the primal SOCP (1.24), and its dual problem (1.25).

(a) If the optimal value of the primal problem is finite and there exists a feasible solution $\overline{x}$ such that

$$A_i\overline{x} - b_i \in \text{int}(C_i), \qquad i = 1, \ldots, m,$$

then there is no duality gap, and the dual problem has an optimal solution.

(b) If the optimal value of the dual problem is finite and there exists a feasible solution $\overline{\lambda} = (\overline{\lambda}_1, \ldots, \overline{\lambda}_m)$ such that

$$\overline{\lambda}_i \in \text{int}(C_i), \qquad i = 1, \ldots, m,$$

then there is no duality gap, and the primal problem has an optimal solution.

Note that while the Linear-Conic Duality Theorem requires a relative interior point condition, the preceding proposition requires an interior point condition. The reason is that the second order cone has nonempty interior, so its relative interior coincides with its interior.

The SOCP arises in many application contexts, and significantly, it can be solved numerically with powerful specialized algorithms that belong to the class of interior point methods, which will be discussed in Section 6.8. We refer to the literature for a more detailed description and analysis (see e.g., the books [BeN01], [BoV04]).

Generally, SOCPs can be recognized from the presence of convex quadratic functions in the cost or the constraint functions. The following are illustrative examples. The first example relates to the field of robust optimization, which involves optimization under uncertainty described by set membership.

Example 1.2.3: (Robust Linear Programming)

Frequently, there is uncertainty about the data of an optimization problem, so one would like to have a solution that is adequate for a whole range of the uncertainty. A popular formulation of this type, is to assume that the constraints contain parameters that take values in a given set, and require that the constraints are satisfied for all values in that set. This approach is also known as a set membership description of the uncertainty and has been used in fields other than optimization, such as set membership estimation, and minimax control (see the textbook [Ber07], which also surveys earlier work).

As an example, consider the problem

$$\begin{aligned} &\text{minimize } \ c'x \\ &\text{subject to } \ a_j'x \le b_j, \quad \forall\, (a_j, b_j) \in T_j, \quad j = 1, \ldots, r, \end{aligned} \tag{1.26}$$

where $c \in \Re^n$ is a given vector, and T_j is a given subset of $\Re^{n+1}$ to which the constraint parameter vectors (a_j, b_j) must belong. The vector x must be chosen so that the constraint $a_j'x \le b_j$ is satisfied for all $(a_j, b_j) \in T_j$, $j = 1, \ldots, r$.

Generally, when T_j contains an infinite number of elements, this problem involves a correspondingly infinite number of constraints. To convert the problem to one involving a finite number of constraints, we note that

$$a_j'x \le b_j, \ \ \forall\, (a_j, b_j) \in T_j \qquad \text{if and only if} \qquad g_j(x) \le 0,$$

where

$$g_j(x) = \sup_{(a_j, b_j) \in T_j} \{a_j'x - b_j\}. \tag{1.27}$$

Thus, the robust linear programming problem (1.26) is equivalent to

$$\begin{aligned} &\text{minimize } \ c'x \\ &\text{subject to } \ g_j(x) \le 0, \qquad j = 1, \ldots, r. \end{aligned}$$

For special choices of the set T_j, the function g_j can be expressed in closed form, and in the case where T_j is an ellipsoid, it turns out that the constraint $g_j(x) \le 0$ can be expressed in terms of a second order cone. To see this, let

$$T_j = \Big\{ (\overline{a}_j + P_j u_j, \overline{b}_j + q_j' u_j) \mid \|u_j\| \le 1,\, u_j \in \Re^{n_j} \Big\}, \tag{1.28}$$

where P_j is a given $n \times n_j$ matrix, $\overline{a}_j \in \Re^n$ and $q_j \in \Re^{n_j}$ are given vectors, and $\overline{b}_j$ is a given scalar. Then, from Eqs. (1.27) and (1.28),

$$\begin{aligned} g_j(x) &= \sup_{\|u_j\| \le 1} \Big\{ (\overline{a}_j + P_j u_j)'x - (\overline{b}_j + q_j' u_j) \Big\} \\ &= \sup_{\|u_j\| \le 1} (P_j'x - q_j)'u_j + \overline{a}_j'x - \overline{b}_j, \end{aligned}$$

and finally

$$g_j(x) = \|P_j'x - q_j\| + \overline{a}_j'x - \overline{b}_j.$$

Thus,

$$g_j(x) \leq 0 \qquad \text{if and only if} \qquad (P_j'x - q_j, \overline{b}_j - \overline{a}_j'x) \in C_j,$$

where C_j is the second order cone of $\Re^{n_j+1}$; i.e., the "robust" constraint $g_j(x) \leq 0$ is equivalent to a second order cone constraint. It follows that in the case of ellipsoidal uncertainty, the robust linear programming problem (1.26) is a SOCP of the form (1.24).

Example 1.2.4: (Quadratically Constrained Quadratic Problems)

Consider the quadratically constrained quadratic problem

$$\begin{aligned} &\text{minimize} \quad x'Q_0x + 2q_0'x + p_0 \\ &\text{subject to} \quad x'Q_jx + 2q_j'x + p_j \leq 0, \quad j = 1, \ldots, r, \end{aligned}$$

where $Q_0, \ldots, Q_r$ are symmetric $n \times n$ positive definite matrices, $q_0, \ldots, q_r$ are vectors in $\Re^n$, and $p_0, \ldots, p_r$ are scalars. We show that the problem can be converted to the second order cone format. A similar conversion is also possible for the quadratic programming problem where Q_0 is positive definite and $Q_j = 0$, $j = 1, \ldots, r$.

Indeed, since each Q_j is symmetric and positive definite, we have

$$\begin{aligned} x'Q_jx + 2q_j'x + p_j &= \left(Q_j^{1/2}x\right)' Q_j^{1/2}x + 2\left(Q_j^{-1/2}q_j\right)' Q_j^{1/2}x + p_j \\ &= \|Q_j^{1/2}x + Q_j^{-1/2}q_j\|^2 + p_j - q_j'Q_j^{-1}q_j, \end{aligned}$$

for $j = 0, 1, \ldots, r$. Thus, the problem can be written as

$$\begin{aligned} &\text{minimize} \quad \|Q_0^{1/2}x + Q_0^{-1/2}q_0\|^2 + p_0 - q_0'Q_0^{-1}q_0 \\ &\text{subject to} \quad \|Q_j^{1/2}x + Q_j^{-1/2}q_j\|^2 + p_j - q_j'Q_j^{-1}q_j \leq 0, \quad j = 1, \ldots, r, \end{aligned}$$

or, by neglecting the constant $p_0 - q_0'Q_0^{-1}q_0$,

$$\begin{aligned} &\text{minimize} \quad \|Q_0^{1/2}x + Q_0^{-1/2}q_0\| \\ &\text{subject to} \quad \|Q_j^{1/2}x + Q_j^{-1/2}q_j\| \leq \left(q_j'Q_j^{-1}q_j - p_j\right)^{1/2}, \quad j = 1, \ldots, r. \end{aligned}$$

By introducing an auxiliary variable x_{n+1}, the problem can be written as

$$\begin{aligned} &\text{minimize} \quad x_{n+1} \\ &\text{subject to} \quad \|Q_0^{1/2}x + Q_0^{-1/2}q_0\| \leq x_{n+1} \\ &\qquad\qquad\quad \|Q_j^{1/2}x + Q_j^{-1/2}q_j\| \leq \left(q_j'Q_j^{-1}q_j - p_j\right)^{1/2}, \quad j = 1, \ldots, r. \end{aligned}$$