Chapter 5

Duality

5.1 The Lagrange dual function

5.1.1 The Lagrangian

We consider an optimization problem in the standard form (4.1):

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p, \end{array} \tag{5.1}$$

with variable $x \in \mathbf{R}^n$. We assume its domain $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \;\cap\; \bigcap_{i=1}^{p} \mathbf{dom}\, h_i$ is nonempty, and denote the optimal value of (5.1) by $p^\star$. We do not assume the problem (5.1) is convex.

The basic idea in Lagrangian duality is to take the constraints in (5.1) into account by augmenting the objective function with a weighted sum of the constraint functions. We define the *Lagrangian* $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ associated with the problem (5.1) as

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x),$$

with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$. We refer to λ_i as the *Lagrange multiplier* associated with the ith inequality constraint $f_i(x) \leq 0$; similarly we refer to ν_i as the Lagrange multiplier associated with the ith equality constraint $h_i(x) = 0$. The vectors λ and ν are called the *dual variables* or *Lagrange multiplier vectors* associated with the problem (5.1).

5.1.2 The Lagrange dual function

We define the *Lagrange dual function* (or just *dual function*) $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ as the minimum value of the Lagrangian over x: for $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right).$$

When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of affine functions of (λ, ν), it is concave, even when the problem (5.1) is not convex.

5.1.3 Lower bounds on optimal value

The dual function yields lower bounds on the optimal value $p^\star$ of the problem (5.1): For any $\lambda \succeq 0$ and any ν we have

$$g(\lambda, \nu) \leq p^\star. \tag{5.2}$$

This important property is easily verified. Suppose $\tilde{x}$ is a feasible point for the problem (5.1), *i.e.*, $f_i(\tilde{x}) \leq 0$ and $h_i(\tilde{x}) = 0$, and $\lambda \succeq 0$. Then we have

$$\sum_{i=1}^m \lambda_i f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) \leq 0,$$

since each term in the first sum is nonpositive, and each term in the second sum is zero, and therefore

$$L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^m \lambda_i f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) \leq f_0(\tilde{x}).$$

Hence

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

Since $g(\lambda, \nu) \leq f_0(\tilde{x})$ holds for every feasible point $\tilde{x}$, the inequality (5.2) follows. The lower bound (5.2) is illustrated in figure 5.1, for a simple problem with $x \in \mathbf{R}$ and one inequality constraint.

The inequality (5.2) holds, but is vacuous, when $g(\lambda, \nu) = -\infty$. The dual function gives a nontrivial lower bound on $p^\star$ only when $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}\, g$, *i.e.*, $g(\lambda, \nu) > -\infty$. We refer to a pair (λ, ν) with $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}\, g$ as *dual feasible*, for reasons that will become clear later.

5.1.4 Linear approximation interpretation

The Lagrangian and lower bound property can be given a simple interpretation, based on a linear approximation of the indicator functions of the sets $\{0\}$ and $-\mathbf{R}_+$.

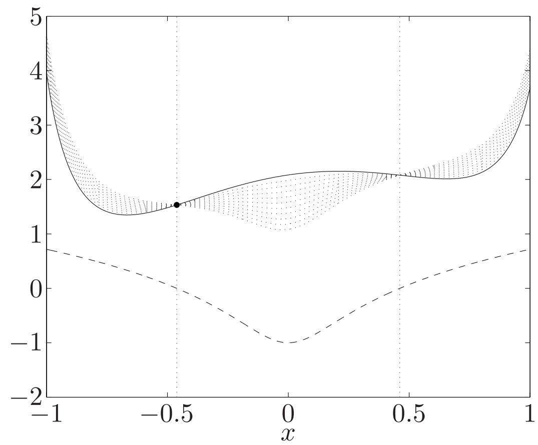

Figure 5.1 *Lower bound from a dual feasible point.* The solid curve shows the objective function f_0, and the dashed curve shows the constraint function f_1. The feasible set is the interval $[-0.46, 0.46]$, which is indicated by the two dotted vertical lines. The optimal point and value are $x^\star = -0.46$, $p^\star = 1.54$ (shown as a circle). The dotted curves show $L(x,\lambda)$ for $\lambda = 0.1,\ 0.2, \ldots, 1.0$. Each of these has a minimum value smaller than $p^\star$, since on the feasible set (and for $\lambda \geq 0$) we have $L(x,\lambda) \leq f_0(x)$.

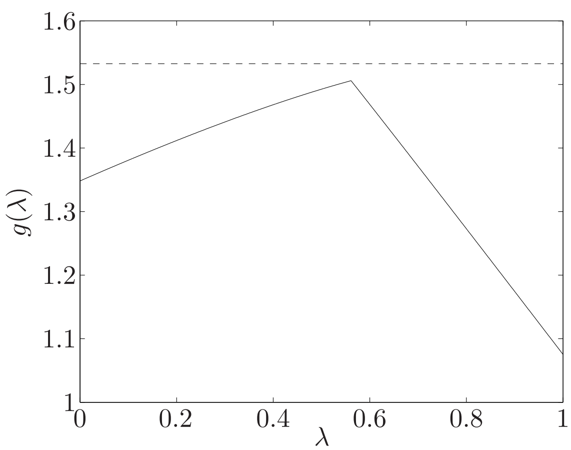

Figure 5.2 The dual function g for the problem in figure 5.1. Neither f_0 nor f_1 is convex, but the dual function is concave. The horizontal dashed line shows $p^\star$, the optimal value of the problem.

We first rewrite the original problem (5.1) as an unconstrained problem,

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^m I_-(f_i(x)) + \sum_{i=1}^p I_0(h_i(x)), \tag{5.3}$$

where $I_- : \mathbf{R} \to \mathbf{R}$ is the indicator function for the nonpositive reals,

$$I_-(u) = \begin{cases} 0 & u \le 0 \\ \infty & u > 0, \end{cases}$$

and similarly, I_0 is the indicator function of $\{0\}$. In the formulation (5.3), the function $I_-(u)$ can be interpreted as expressing our irritation or displeasure associated with a constraint function value $u = f_i(x)$: It is zero if $f_i(x) \le 0$, and infinite if $f_i(x) > 0$. In a similar way, $I_0(u)$ gives our displeasure for an equality constraint value $u = h_i(x)$. We can think of I_- as a "brick wall" or "infinitely hard" displeasure function; our displeasure rises from zero to infinite as $f_i(x)$ transitions from nonpositive to positive.

Now suppose in the formulation (5.3) we replace the function $I_-(u)$ with the linear function $\lambda_i u$, where $\lambda_i \ge 0$, and the function $I_0(u)$ with $\nu_i u$. The objective becomes the Lagrangian function $L(x, \lambda, \nu)$, and the dual function value $g(\lambda, \nu)$ is the optimal value of the problem

$$\text{minimize} \quad L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x). \tag{5.4}$$

In this formulation, we use a linear or "soft" displeasure function in place of I_- and I_0. For an inequality constraint, our displeasure is zero when $f_i(x) = 0$, and is positive when $f_i(x) > 0$ (assuming $\lambda_i > 0$); our displeasure grows as the constraint becomes "more violated". Unlike the original formulation, in which any nonpositive value of $f_i(x)$ is acceptable, in the soft formulation we actually derive pleasure from constraints that have margin, *i.e.*, from $f_i(x) < 0$.

Clearly the approximation of the indicator function $I_-(u)$ with a linear function $\lambda_i u$ is rather poor. But the linear function is at least an *underestimator* of the indicator function. Since $\lambda_i u \le I_-(u)$ and $\nu_i u \le I_0(u)$ for all u, we see immediately that the dual function yields a lower bound on the optimal value of the original problem.

The idea of replacing the "hard" constraints with "soft" versions will come up again when we consider interior-point methods (§11.2.1).

5.1.5 Examples

In this section we give some examples for which we can derive an analytical expression for the Lagrange dual function.

Least-squares solution of linear equations

We consider the problem

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b, \end{array} \tag{5.5}$$

where $A \in \mathbf{R}^{p \times n}$. This problem has no inequality constraints and p (linear) equality constraints. The Lagrangian is $L(x, \nu) = x^T x + \nu^T (Ax - b)$, with domain $\mathbf{R}^n \times$

$\mathbf{R}^p$. The dual function is given by $g(\nu) = \inf_x L(x, \nu)$. Since $L(x, \nu)$ is a convex quadratic function of x, we can find the minimizing x from the optimality condition

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0,$$

which yields $x = -(1/2)A^T \nu$. Therefore the dual function is

$$g(\nu) = L(-(1/2)A^T \nu, \nu) = -(1/4)\nu^T AA^T \nu - b^T \nu,$$

which is a concave quadratic function, with domain $\mathbf{R}^p$. The lower bound property (5.2) states that for any $\nu \in \mathbf{R}^p$, we have

$$-(1/4)\nu^T AA^T \nu - b^T \nu \leq \inf\{x^T x \mid Ax = b\}.$$

Standard form LP

Consider an LP in standard form,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0, \end{array} \tag{5.6}$$

which has inequality constraint functions $f_i(x) = -x_i$, $i = 1, \ldots, n$. To form the Lagrangian we introduce multipliers λ_i for the n inequality constraints and multipliers ν_i for the equality constraints, and obtain

$$L(x, \lambda, \nu) = c^T x - \sum_{i=1}^n \lambda_i x_i + \nu^T (Ax - b) = -b^T \nu + (c + A^T \nu - \lambda)^T x.$$

The dual function is

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = -b^T \nu + \inf_x (c + A^T \nu - \lambda)^T x,$$

which is easily determined analytically, since a linear function is bounded below only when it is identically zero. Thus, $g(\lambda, \nu) = -\infty$ except when $c + A^T \nu - \lambda = 0$, in which case it is $-b^T \nu$:

$$g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Note that the dual function g is finite only on a proper affine subset of $\mathbf{R}^m \times \mathbf{R}^p$. We will see that this is a common occurrence.

The lower bound property (5.2) is nontrivial only when λ and ν satisfy $\lambda \succeq 0$ and $A^T \nu - \lambda + c = 0$. When this occurs, $-b^T \nu$ is a lower bound on the optimal value of the LP (5.6).

Two-way partitioning problem

We consider the (nonconvex) problem

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n, \end{array} \tag{5.7}$$

where $W \in \mathbf{S}^n$. The constraints restrict the values of x_i to 1 or -1, so the problem is equivalent to finding the vector with components ± 1 that minimizes $x^T W x$. The feasible set here is finite (it contains 2^n points) so this problem can in principle be solved by simply checking the objective value of each feasible point. Since the number of feasible points grows exponentially, however, this is possible only for small problems (say, with $n \leq 30$). In general (and for n larger than, say, 50) the problem (5.7) is very difficult to solve.

We can interpret the problem (5.7) as a two-way partitioning problem on a set of n elements, say, $\{1, \dots, n\}$: A feasible x corresponds to the partition

$$\{1, \dots, n\} = \{i \mid x_i = -1\} \cup \{i \mid x_i = 1\}.$$

The matrix coefficient W_{ij} can be interpreted as the cost of having the elements i and j in the same partition, and $-W_{ij}$ is the cost of having i and j in different partitions. The objective in (5.7) is the total cost, over all pairs of elements, and the problem (5.7) is to find the partition with least total cost.

We now derive the dual function for this problem. The Lagrangian is

$$\begin{aligned} L(x, \nu) &= x^T W x + \sum_{i=1}^n \nu_i (x_i^2 - 1) \\ &= x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu. \end{aligned}$$

We obtain the Lagrange dual function by minimizing over x:

$$\begin{aligned} g(\nu) &= \inf_x x^T (W + \mathbf{diag}(\nu)) x - \mathbf{1}^T \nu \\ &= \begin{cases} -\mathbf{1}^T \nu & W + \mathbf{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}$$

where we use the fact that the infimum of a quadratic form is either zero (if the form is positive semidefinite) or $-\infty$ (if the form is not positive semidefinite).

This dual function provides lower bounds on the optimal value of the difficult problem (5.7). For example, we can take the specific value of the dual variable

$$\nu = -\lambda_{\min}(W) \mathbf{1},$$

which is dual feasible, since

$$W + \mathbf{diag}(\nu) = W - \lambda_{\min}(W) I \succeq 0.$$

This yields the bound on the optimal value $p^\star$

$$p^\star \geq -\mathbf{1}^T \nu = n \lambda_{\min}(W). \tag{5.8}$$

Remark 5.1 This lower bound on $p^\star$ can also be obtained without using the Lagrange dual function. First, we replace the constraints $x_1^2 = 1, \dots, x_n^2 = 1$ with $\sum_{i=1}^n x_i^2 = n$, to obtain the modified problem

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & \sum_{i=1}^n x_i^2 = n. \end{array} \tag{5.9}$$

The constraints of the original problem (5.7) imply the constraint here, so the optimal value of the problem (5.9) is a lower bound on $p^\star$, the optimal value of (5.7). But the modified problem (5.9) is easily solved as an eigenvalue problem, with optimal value $n\lambda_{\min}(W)$.

5.1.6 The Lagrange dual function and conjugate functions

Recall from §3.3 that the conjugate f^* of a function $f : \mathbf{R}^n \to \mathbf{R}$ is given by

$$f^*(y) = \sup_{x \in \mathbf{dom}\, f} \left(y^T x - f(x) \right).$$

The conjugate function and Lagrange dual function are closely related. To see one simple connection, consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x = 0 \end{array}$$

(which is not very interesting, and solvable by inspection). This problem has Lagrangian $L(x,\nu) = f(x) + \nu^T x$, and dual function

$$g(\nu) = \inf_x \left(f(x) + \nu^T x \right) = -\sup_x \left((-\nu)^T x - f(x) \right) = -f^*(-\nu).$$

More generally (and more usefully), consider an optimization problem with linear inequality and equality constraints,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \preceq b \\ & Cx = d. \end{array} \tag{5.10}$$

Using the conjugate of f_0 we can write the dual function for the problem (5.10) as

$$\begin{aligned} g(\lambda,\nu) &= \inf_x \left(f_0(x) + \lambda^T (Ax - b) + \nu^T (Cx - d) \right) \\ &= -b^T\lambda - d^T\nu + \inf_x \left(f_0(x) + (A^T\lambda + C^T\nu)^T x \right) \\ &= -b^T\lambda - d^T\nu - f_0^*(-A^T\lambda - C^T\nu). \end{aligned} \tag{5.11}$$

The domain of g follows from the domain of f_0^*:

$$\mathbf{dom}\, g = \{(\lambda,\nu) \mid -A^T\lambda - C^T\nu \in \mathbf{dom}\, f_0^*\}.$$

Let us illustrate this with a few examples.

Equality constrained norm minimization

Consider the problem

$$\begin{array}{ll} \text{minimize} & \|x\| \\ \text{subject to} & Ax = b, \end{array} \tag{5.12}$$

where $\|\cdot\|$ is any norm. Recall (from example 3.26 on page 93) that the conjugate of $f_0 = \|\cdot\|$ is given by

$$f_0^*(y) = \begin{cases} 0 & \|y\|_* \leq 1 \\ \infty & \text{otherwise,} \end{cases}$$

the indicator function of the dual norm unit ball.

Using the result (5.11) above, the dual function for the problem (5.12) is given by

$$g(\nu) = -b^T\nu - f_0^*(-A^T\nu) = \begin{cases} -b^T\nu & \|A^T\nu\|_* \leq 1 \\ -\infty & \text{otherwise.} \end{cases}$$

Entropy maximization

Consider the entropy maximization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1 \end{array} \tag{5.13}$$

where $\mathbf{dom}\, f_0 = \mathbf{R}_{++}^n$. The conjugate of the negative entropy function $u \log u$, with scalar variable u, is e^{v-1} (see example 3.21 on page 91). Since f_0 is a sum of negative entropy functions of different variables, we conclude that its conjugate is

$$f_0^*(y) = \sum_{i=1}^n e^{y_i - 1},$$

with $\mathbf{dom}\, f_0^* = \mathbf{R}^n$. Using the result (5.11) above, the dual function of (5.13) is given by

$$g(\lambda, \nu) = -b^T\lambda - \nu - \sum_{i=1}^n e^{-a_i^T\lambda - \nu - 1} = -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T\lambda}$$

where a_i is the ith column of A.

Minimum volume covering ellipsoid

Consider the problem with variable $X \in \mathbf{S}^n$,

$$\begin{array}{ll} \text{minimize} & f_0(X) = \log\det X^{-1} \\ \text{subject to} & a_i^T X a_i \leq 1, \quad i = 1, \ldots, m, \end{array} \tag{5.14}$$

where $\mathbf{dom}\, f_0 = \mathbf{S}_{++}^n$. The problem (5.14) has a simple geometric interpretation. With each $X \in \mathbf{S}_{++}^n$ we associate the ellipsoid, centered at the origin,

$$\mathcal{E}_X = \{z \mid z^T X z \leq 1\}.$$

The volume of this ellipsoid is proportional to $\left(\det X^{-1}\right)^{1/2}$, so the objective of (5.14) is, except for a constant and a factor of two, the logarithm of the volume of $\mathcal{E}_X$. The constraints of the problem (5.14) are that $a_i \in \mathcal{E}_X$. Thus the problem (5.14) is to determine the minimum volume ellipsoid, centered at the origin, that includes the points $a_1, \ldots, a_m$.

The inequality constraints in problem (5.14) are affine; they can be expressed as

$$\mathbf{tr}\left((a_i a_i^T)X\right) \leq 1.$$

In example 3.23 (page 92) we found that the conjugate of f_0 is

$$f_0^*(Y) = \log\det(-Y)^{-1} - n,$$

with $\mathbf{dom}\, f_0^* = -\mathbf{S}^n_{++}$. Applying the result (5.11) above, the dual function for the problem (5.14) is given by

$$g(\lambda) = \begin{cases} \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - \mathbf{1}^T\lambda + n & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ -\infty & \text{otherwise.} \end{cases} \tag{5.15}$$

Thus, for any $\lambda \succeq 0$ with $\sum_{i=1}^m \lambda_i a_i a_i^T \succ 0$, the number

$$\log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - \mathbf{1}^T\lambda + n$$

is a lower bound on the optimal value of the problem (5.14).

5.2 The Lagrange dual problem

For each pair (λ, ν) with $\lambda \succeq 0$, the Lagrange dual function gives us a lower bound on the optimal value $p^\star$ of the optimization problem (5.1). Thus we have a lower bound that depends on some parameters λ, ν. A natural question is: What is the *best* lower bound that can be obtained from the Lagrange dual function?

This leads to the optimization problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0. \end{array} \tag{5.16}$$

This problem is called the *Lagrange dual problem* associated with the problem (5.1). In this context the original problem (5.1) is sometimes called the *primal problem*. The term *dual feasible*, to describe a pair (λ, ν) with $\lambda \succeq 0$ and $g(\lambda, \nu) > -\infty$, now makes sense. It means, as the name implies, that (λ, ν) is feasible for the dual problem (5.16). We refer to $(\lambda^\star, \nu^\star)$ as *dual optimal* or *optimal Lagrange multipliers* if they are optimal for the problem (5.16).

The Lagrange dual problem (5.16) is a convex optimization problem, since the objective to be maximized is concave and the constraint is convex. This is the case whether or not the primal problem (5.1) is convex.

5.2.1 Making dual constraints explicit

The examples above show that it is not uncommon for the domain of the dual function,

$$\mathbf{dom}\, g = \{(\lambda, \nu) \mid g(\lambda, \nu) > -\infty\},$$

to have dimension smaller than $m + p$. In many cases we can identify the affine hull of $\mathbf{dom}\, g$, and describe it as a set of linear equality constraints. Roughly speaking, this means we can identify the equality constraints that are 'hidden' or 'implicit' in the objective g of the dual problem (5.16). In this case we can form an equivalent problem, in which these equality constraints are given explicitly as constraints. The following examples demonstrate this idea.

Lagrange dual of standard form LP

On page 219 we found that the Lagrange dual function for the standard form LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0 \end{array} \tag{5.17}$$

is given by

$$g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Strictly speaking, the Lagrange dual problem of the standard form LP is to maximize this dual function g subject to $\lambda \succeq 0$, *i.e.*,

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases} \\ \text{subject to} & \lambda \succeq 0. \end{array} \tag{5.18}$$

Here g is finite only when $A^T \nu - \lambda + c = 0$. We can form an equivalent problem by making these equality constraints explicit:

$$\begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu - \lambda + c = 0 \\ & \lambda \succeq 0. \end{array} \tag{5.19}$$

This problem, in turn, can be expressed as

$$\begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c \succeq 0, \end{array} \tag{5.20}$$

which is an LP in inequality form.

Note the subtle distinctions between these three problems. The Lagrange dual of the standard form LP (5.17) is the problem (5.18), which is equivalent to (but not the same as) the problems (5.19) and (5.20). With some abuse of terminology, we refer to the problem (5.19) or the problem (5.20) as the Lagrange dual of the standard form LP (5.17).

Lagrange dual of inequality form LP

In a similar way we can find the Lagrange dual problem of a linear program in inequality form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b. \end{array} \tag{5.21}$$

The Lagrangian is

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

so the dual function is

$$g(\lambda) = \inf_x L(x, \lambda) = -b^T \lambda + \inf_x (A^T \lambda + c)^T x.$$

The infimum of a linear function is $-\infty$, except in the special case when it is identically zero, so the dual function is

$$g(\lambda) = \begin{cases} -b^T \lambda & A^T \lambda + c = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The dual variable λ is dual feasible if $\lambda \succeq 0$ and $A^T \lambda + c = 0$.

The Lagrange dual of the LP (5.21) is to maximize g over all $\lambda \succeq 0$. Again we can reformulate this by explicitly including the dual feasibility conditions as constraints, as in

$$\begin{array}{ll} \text{maximize} & -b^T \lambda \\ \text{subject to} & A^T \lambda + c = 0 \\ & \lambda \succeq 0, \end{array} \tag{5.22}$$

which is an LP in standard form.

Note the interesting symmetry between the standard and inequality form LPs and their duals: The dual of a standard form LP is an LP with only inequality constraints, and vice versa. One can also verify that the Lagrange dual of (5.22) is (equivalent to) the primal problem (5.21).

5.2.2 Weak duality

The optimal value of the Lagrange dual problem, which we denote $d^\star$, is, by definition, the best lower bound on $p^\star$ that can be obtained from the Lagrange dual function. In particular, we have the simple but important inequality

$$d^\star \leq p^\star, \tag{5.23}$$

which holds even if the original problem is not convex. This property is called *weak duality*.

The weak duality inequality (5.23) holds when $d^\star$ and $p^\star$ are infinite. For example, if the primal problem is unbounded below, so that $p^\star = -\infty$, we must have $d^\star = -\infty$, *i.e.*, the Lagrange dual problem is infeasible. Conversely, if the dual problem is unbounded above, so that $d^\star = \infty$, we must have $p^\star = \infty$, *i.e.*, the primal problem is infeasible.

We refer to the difference $p^\star - d^\star$ as the *optimal duality gap* of the original problem, since it gives the gap between the optimal value of the primal problem and the best (*i.e.*, greatest) lower bound on it that can be obtained from the Lagrange dual function. The optimal duality gap is always nonnegative.

The bound (5.23) can sometimes be used to find a lower bound on the optimal value of a problem that is difficult to solve, since the dual problem is always convex, and in many cases can be solved efficiently, to find $d^\star$. As an example, consider the two-way partitioning problem (5.7) described on page 219. The dual problem is an SDP,

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T\nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0, \end{array}$$

with variable $\nu \in \mathbf{R}^n$. This problem can be solved efficiently, even for relatively large values of n, such as $n = 1000$. Its optimal value is a lower bound on the optimal value of the two-way partitioning problem, and is always at least as good as the lower bound (5.8) based on $\lambda_{\min}(W)$.

5.2.3 Strong duality and Slater's constraint qualification

If the equality

$$d^\star = p^\star \tag{5.24}$$

holds, *i.e.*, the optimal duality gap is zero, then we say that *strong duality* holds. This means that the best bound that can be obtained from the Lagrange dual function is tight.

Strong duality does not, in general, hold. But if the primal problem (5.1) is convex, *i.e.*, of the form

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \ldots, m, \\ & Ax = b, \end{array} \tag{5.25}$$

with $f_0, \ldots, f_m$ convex, we usually (but not always) have strong duality. There are many results that establish conditions on the problem, beyond convexity, under which strong duality holds. These conditions are called *constraint qualifications*.

One simple constraint qualification is *Slater's condition*: There exists an $x \in \mathbf{relint}\, \mathcal{D}$ such that

$$f_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b. \tag{5.26}$$

Such a point is sometimes called *strictly feasible*, since the inequality constraints hold with strict inequalities. Slater's theorem states that strong duality holds, if Slater's condition holds (and the problem is convex).

Slater's condition can be refined when some of the inequality constraint functions f_i are affine. If the first k constraint functions $f_1, \ldots, f_k$ are affine, then strong duality holds provided the following weaker condition holds: There exists an $x \in \mathbf{relint}\, \mathcal{D}$ with

$$f_i(x) \le 0, \quad i = 1, \ldots, k, \qquad f_i(x) < 0, \quad i = k+1, \ldots, m, \qquad Ax = b. \tag{5.27}$$

In other words, the affine inequalities do not need to hold with strict inequality. Note that the refined Slater condition (5.27) reduces to feasibility when the constraints are all linear equalities and inequalities, and $\mathbf{dom}\, f_0$ is open.

Slater's condition (and the refinement (5.27)) not only implies strong duality for convex problems. It also implies that the dual optimal value is attained when $d^\star > -\infty$, *i.e.*, there exists a dual feasible $(\lambda^\star, \nu^\star)$ with $g(\lambda^\star, \nu^\star) = d^\star = p^\star$. We will prove that strong duality obtains, when the primal problem is convex and Slater's condition holds, in §5.3.2.

5.2.4 Examples

Least-squares solution of linear equations

Recall the problem (5.5):

$$\begin{array}{ll} \text{minimize} & x^T x \\ \text{subject to} & Ax = b. \end{array}$$

The associated dual problem is

$$\text{maximize} \quad -(1/4)\nu^T A A^T \nu - b^T \nu,$$

which is an unconstrained concave quadratic maximization problem.

Slater's condition is simply that the primal problem is feasible, so $p^\star = d^\star$ provided $b \in \mathcal{R}(A)$, *i.e.*, $p^\star < \infty$. In fact for this problem we always have strong duality, even when $p^\star = \infty$. This is the case when $b \notin \mathcal{R}(A)$, so there is a z with $A^T z = 0$, $b^T z \neq 0$. It follows that the dual function is unbounded above along the line $\{tz \mid t \in \mathbf{R}\}$, so $d^\star = \infty$ as well.

Lagrange dual of LP

By the weaker form of Slater's condition, we find that strong duality holds for any LP (in standard or inequality form) provided the primal problem is feasible. Applying this result to the duals, we conclude that strong duality holds for LPs if the dual is feasible. This leaves only one possible situation in which strong duality for LPs can fail: both the primal and dual problems are infeasible. This pathological case can, in fact, occur; see exercise 5.23.

Lagrange dual of QCQP

We consider the QCQP

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \ldots, m, \end{array} \tag{5.28}$$

with $P_0 \in \mathbf{S}^n_{++}$, and $P_i \in \mathbf{S}^n_+$, $i = 1, \ldots, m$. The Lagrangian is

$$L(x, \lambda) = (1/2)x^T P(\lambda) x + q(\lambda)^T x + r(\lambda),$$

where

$$P(\lambda) = P_0 + \sum_{i=1}^m \lambda_i P_i, \qquad q(\lambda) = q_0 + \sum_{i=1}^m \lambda_i q_i, \qquad r(\lambda) = r_0 + \sum_{i=1}^m \lambda_i r_i.$$

It is possible to derive an expression for $g(\lambda)$ for general λ, but it is quite complicated. If $\lambda \succeq 0$, however, we have $P(\lambda) \succ 0$ and

$$g(\lambda) = \inf_x L(x, \lambda) = -(1/2)q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda).$$

We can therefore express the dual problem as

$$\begin{array}{ll} \text{maximize} & -(1/2)q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda) \\ \text{subject to} & \lambda \succeq 0. \end{array} \tag{5.29}$$

The Slater condition says that strong duality between (5.29) and (5.28) holds if the quadratic inequality constraints are strictly feasible, *i.e.*, there exists an x with

$$(1/2)x^T P_i x + q_i^T x + r_i < 0, \quad i = 1, \dots, m.$$

Entropy maximization

Our next example is the entropy maximization problem (5.13):

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1, \end{array}$$

with domain $\mathcal{D} = \mathbf{R}_+^n$. The Lagrange dual function was derived on page 222; the dual problem is

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T \lambda} \\ \text{subject to} & \lambda \succeq 0, \end{array} \tag{5.30}$$

with variables $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}$. The (weaker) Slater condition for (5.13) tells us that the optimal duality gap is zero if there exists an $x \succ 0$ with $Ax \preceq b$ and $\mathbf{1}^T x = 1$.

We can simplify the dual problem (5.30) by maximizing over the dual variable ν analytically. For fixed λ, the objective function is maximized when the derivative with respect to ν is zero, *i.e.*,

$$\nu = \log \sum_{i=1}^n e^{-a_i^T \lambda} - 1.$$

Substituting this optimal value of ν into the dual problem gives

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \log\left(\sum_{i=1}^n e^{-a_i^T\lambda}\right) \\ \text{subject to} & \lambda \succeq 0, \end{array}$$

which is a geometric program (in convex form) with nonnegativity constraints.

Minimum volume covering ellipsoid

We consider the problem (5.14):

$$\begin{array}{ll} \text{minimize} & \log\det X^{-1} \\ \text{subject to} & a_i^T X a_i \le 1, \quad i = 1, \dots, m, \end{array}$$

with domain $\mathcal{D} = \mathbf{S}^n_{++}$. The Lagrange dual function is given by (5.15), so the dual problem can be expressed as

$$\begin{array}{ll} \text{maximize} & \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - \mathbf{1}^T\lambda + n \\ \text{subject to} & \lambda \succeq 0 \end{array} \qquad (5.31)$$

where we take $\log\det X = -\infty$ if $X \not\succ 0$.

The (weaker) Slater condition for the problem (5.14) is that there exists an $X \in \mathbf{S}^n_{++}$ with $a_i^T X a_i \leq 1$, for $i = 1, \ldots, m$. This is always satisfied, so strong duality always obtains between (5.14) and the dual problem (5.31).

A nonconvex quadratic problem with strong duality

On rare occasions strong duality obtains for a *nonconvex* problem. As an important example, we consider the problem of minimizing a nonconvex quadratic function over the unit ball,

$$\begin{array}{ll} \text{minimize} & x^T A x + 2b^T x \\ \text{subject to} & x^T x \leq 1, \end{array} \qquad (5.32)$$

where $A \in \mathbf{S}^n$, $A \not\succeq 0$, and $b \in \mathbf{R}^n$. Since $A \not\succeq 0$, this is not a convex problem. This problem is sometimes called the *trust region problem*, and arises in minimizing a second-order approximation of a function over the unit ball, which is the region in which the approximation is assumed to be approximately valid.

The Lagrangian is

$$L(x,\lambda) = x^T A x + 2b^T x + \lambda(x^T x - 1) = x^T(A + \lambda I)x + 2b^T x - \lambda,$$

so the dual function is given by

$$g(\lambda) = \left\{ \begin{array}{ll} -b^T(A+\lambda I)^\dagger b - \lambda & A + \lambda I \succeq 0, \quad b \in \mathcal{R}(A + \lambda I) \\ -\infty & \text{otherwise}, \end{array} \right.$$

where $(A + \lambda I)^\dagger$ is the pseudo-inverse of $A + \lambda I$. The Lagrange dual problem is thus

$$\begin{array}{ll} \text{maximize} & -b^T(A+\lambda I)^\dagger b - \lambda \\ \text{subject to} & A + \lambda I \succeq 0, \quad b \in \mathcal{R}(A + \lambda I), \end{array} \qquad (5.33)$$

with variable $\lambda \in \mathbf{R}$. Although it is not obvious from this expression, this is a convex optimization problem. In fact, it is readily solved since it can be expressed as

$$\begin{array}{ll} \text{maximize} & -\sum_{i=1}^n (q_i^T b)^2/(\lambda_i + \lambda) - \lambda \\ \text{subject to} & \lambda \geq -\lambda_{\min}(A), \end{array}$$

where λ_i and q_i are the eigenvalues and corresponding (orthonormal) eigenvectors of A, and we interpret $(q_i^T b)^2/0$ as 0 if $q_i^T b = 0$ and as ∞ otherwise.

Despite the fact that the original problem (5.32) is not convex, we always have zero optimal duality gap for this problem: The optimal values of (5.32) and (5.33) are always the same. In fact, a more general result holds: strong duality holds for any optimization problem with quadratic objective and one quadratic inequality constraint, provided Slater's condition holds; see §B.1.

5.2.5 Mixed strategies for matrix games

In this section we use strong duality to derive a basic result for zero-sum matrix games. We consider a game with two players. Player 1 makes a choice (or *move*) $k \in \{1, \ldots, n\}$, and player 2 makes a choice $l \in \{1, \ldots, m\}$. Player 1 then makes a payment of P_{kl} to player 2, where $P \in \mathbf{R}^{n \times m}$ is the *payoff matrix* for the game. The goal of player 1 is to make the payment as small as possible, while the goal of player 2 is to maximize it.

The players use randomized or *mixed strategies*, which means that each player makes his or her choice randomly and independently of the other player's choice, according to a probability distribution:

$$\mathbf{prob}(k = i) = u_i, \quad i = 1, \ldots, n, \qquad \mathbf{prob}(l = i) = v_i, \quad i = 1, \ldots, m.$$

Here u and v give the probability distributions of the choices of the two players, *i.e.*, their associated strategies. The expected payoff from player 1 to player 2 is then

$$\sum_{k=1}^{n} \sum_{l=1}^{m} u_k v_l P_{kl} = u^T P v.$$

Player 1 wishes to choose u to minimize $u^T P v$, while player 2 wishes to choose v to maximize $u^T P v$.

Let us first analyze the game from the point of view of player 1, assuming her strategy u is known to player 2 (which clearly gives an advantage to player 2). Player 2 will choose v to maximize $u^T P v$, which results in the expected payoff

$$\sup\{u^T P v \mid v \succeq 0, \ \mathbf{1}^T v = 1\} = \max_{i=1,\ldots,m} (P^T u)_i.$$

The best thing player 1 can do is to choose u to minimize this worst-case payoff to player 2, *i.e.*, to choose a strategy u that solves the problem

$$\begin{array}{ll} \text{minimize} & \max_{i=1,\ldots,m} (P^T u)_i \\ \text{subject to} & u \succeq 0, \quad \mathbf{1}^T u = 1, \end{array} \tag{5.34}$$

which is a piecewise-linear convex optimization problem. We will denote the optimal value of this problem as $p_1^\star$. This is the smallest expected payoff player 1 can arrange to have, assuming that player 2 knows the strategy of player 1, and plays to his own maximum advantage.

In a similar way we can consider the situation in which v, the strategy of player 2, is known to player 1 (which gives an advantage to player 1). In this case player 1 chooses u to minimize $u^T P v$, which results in an expected payoff of

$$\inf\{u^T P v \mid u \succeq 0, \ \mathbf{1}^T u = 1\} = \min_{i=1,\ldots,n} (P v)_i.$$

Player 2 chooses v to maximize this, *i.e.*, chooses a strategy v that solves the problem

$$\begin{array}{ll} \text{maximize} & \min_{i=1,\ldots,n} (P v)_i \\ \text{subject to} & v \succeq 0, \quad \mathbf{1}^T v = 1, \end{array} \tag{5.35}$$

which is another convex optimization problem, with piecewise-linear (concave) objective. We will denote the optimal value of this problem as $p_2^\star$. This is the largest expected payoff player 2 can guarantee getting, assuming that player 1 knows the strategy of player 2.

It is intuitively obvious that knowing your opponent's strategy gives an advantage (or at least, cannot hurt), and indeed, it is easily shown that we always have $p_1^\star \geq p_2^\star$. We can interpret the difference, $p_1^\star - p_2^\star$, which is nonnegative, as the advantage conferred on a player by knowing the opponent's strategy.

Using duality, we can establish a result that is at first surprising: $p_1^\star = p_2^\star$. In other words, in a matrix game with mixed strategies, there is *no* advantage to knowing your opponent's strategy. We will establish this result by showing that the two problems (5.34) and (5.35) are Lagrange dual problems, for which strong duality obtains.

We start by formulating (5.34) as an LP,

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & u \succeq 0, \quad \mathbf{1}^T u = 1 \\ & P^T u \preceq t\mathbf{1}, \end{array}$$

with extra variable $t \in \mathbf{R}$. Introducing the multiplier λ for $P^T u \preceq t\mathbf{1}$, μ for $u \succeq 0$, and ν for $\mathbf{1}^T u = 1$, the Lagrangian is

$$t + \lambda^T(P^T u - t\mathbf{1}) - \mu^T u + \nu(1 - \mathbf{1}^T u) = \nu + (1 - \mathbf{1}^T\lambda)t + (P\lambda - \nu\mathbf{1} - \mu)^T u,$$

so the dual function is

$$g(\lambda, \mu, \nu) = \begin{cases} \nu & \mathbf{1}^T\lambda = 1, \quad P\lambda - \nu\mathbf{1} = \mu \\ -\infty & \text{otherwise.} \end{cases}$$

The dual problem is then

$$\begin{array}{ll} \text{maximize} & \nu \\ \text{subject to} & \lambda \succeq 0, \quad \mathbf{1}^T\lambda = 1, \quad \mu \succeq 0 \\ & P\lambda - \nu\mathbf{1} = \mu. \end{array}$$

Eliminating μ we obtain the following Lagrange dual of (5.34):

$$\begin{array}{ll} \text{maximize} & \nu \\ \text{subject to} & \lambda \succeq 0, \quad \mathbf{1}^T\lambda = 1 \\ & P\lambda \succeq \nu\mathbf{1}, \end{array}$$

with variables λ, ν. But this is clearly equivalent to (5.35). Since the LPs are feasible, we have strong duality; the optimal values of (5.34) and (5.35) are equal.

5.3 Geometric interpretation

5.3.1 Weak and strong duality via set of values

We can give a simple geometric interpretation of the dual function in terms of the set

$$\mathcal{G} = \{(f_1(x), \dots, f_m(x), h_1(x), \dots, h_p(x), f_0(x)) \in \mathbf{R}^m \times \mathbf{R}^p \times \mathbf{R} \mid x \in \mathcal{D}\}, \quad (5.36)$$

which is the set of values taken on by the constraint and objective functions. The optimal value $p^\star$ of (5.1) is easily expressed in terms of $\mathcal{G}$ as

$$p^\star = \inf\{t \mid (u, v, t) \in \mathcal{G},\ u \preceq 0,\ v = 0\}.$$

To evaluate the dual function at (λ, ν), we minimize the affine function

$$(\lambda, \nu, 1)^T(u, v, t) = \sum_{i=1}^m \lambda_i u_i + \sum_{i=1}^p \nu_i v_i + t$$

over $(u, v, t) \in \mathcal{G}$, *i.e.*, we have

$$g(\lambda, \nu) = \inf\{(\lambda, \nu, 1)^T(u, v, t) \mid (u, v, t) \in \mathcal{G}\}.$$

In particular, we see that if the infimum is finite, then the inequality

$$(\lambda, \nu, 1)^T(u, v, t) \geq g(\lambda, \nu)$$

defines a supporting hyperplane to $\mathcal{G}$. This is sometimes referred to as a *nonvertical* supporting hyperplane, because the last component of the normal vector is nonzero.

Now suppose $\lambda \succeq 0$. Then, obviously, $t \geq (\lambda, \nu, 1)^T(u, v, t)$ if $u \preceq 0$ and $v = 0$. Therefore

$$\begin{array}{rcl} p^\star & = & \inf\{t \mid (u, v, t) \in \mathcal{G},\ u \preceq 0,\ v = 0\} \\ & \geq & \inf\{(\lambda, \nu, 1)^T(u, v, t) \mid (u, v, t) \in \mathcal{G},\ u \preceq 0,\ v = 0\} \\ & \geq & \inf\{(\lambda, \nu, 1)^T(u, v, t) \mid (u, v, t) \in \mathcal{G}\} \\ & = & g(\lambda, \nu), \end{array}$$

i.e., we have weak duality. This interpretation is illustrated in figures 5.3 and 5.4, for a simple problem with one inequality constraint.

Epigraph variation

In this section we describe a variation on the geometric interpretation of duality in terms of $\mathcal{G}$, which explains why strong duality obtains for (most) convex problems. We define the set $\mathcal{A} \subseteq \mathbf{R}^m \times \mathbf{R}^p \times \mathbf{R}$ as

$$\mathcal{A} = \mathcal{G} + \left(\mathbf{R}_+^m \times \{0\} \times \mathbf{R}_+\right), \quad (5.37)$$

or, more explicitly,

$$\begin{array}{l} \mathcal{A} = \{(u, v, t) \mid \exists x \in \mathcal{D},\ f_i(x) \leq u_i,\ i = 1, \dots, m, \\ \qquad h_i(x) = v_i,\ i = 1, \dots, p,\ f_0(x) \leq t\}, \end{array}$$

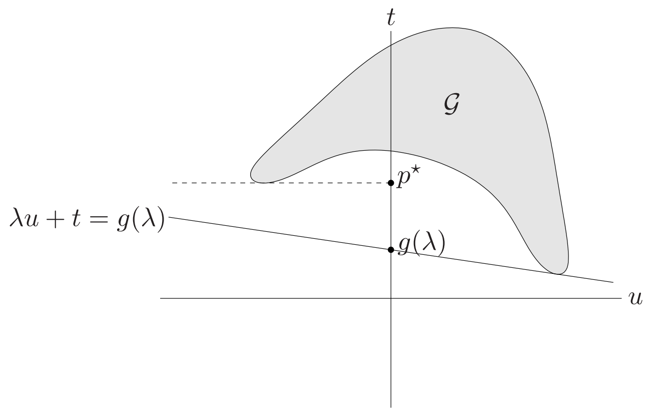

Figure 5.3 Geometric interpretation of dual function and lower bound $g(\lambda) \leq p^\star$, for a problem with one (inequality) constraint. Given λ, we minimize $(\lambda, 1)^T(u, t)$ over $\mathcal{G} = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$. This yields a supporting hyperplane with slope $-\lambda$. The intersection of this hyperplane with the $u = 0$ axis gives $g(\lambda)$.

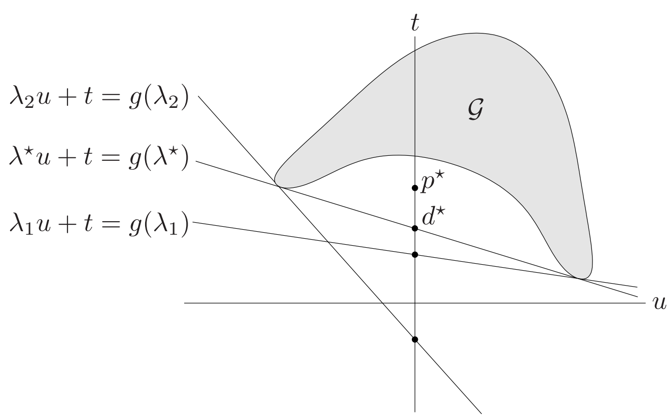

Figure 5.4 Supporting hyperplanes corresponding to three dual feasible values of λ, including the optimum $\lambda^\star$. Strong duality does not hold; the optimal duality gap $p^\star - d^\star$ is positive.

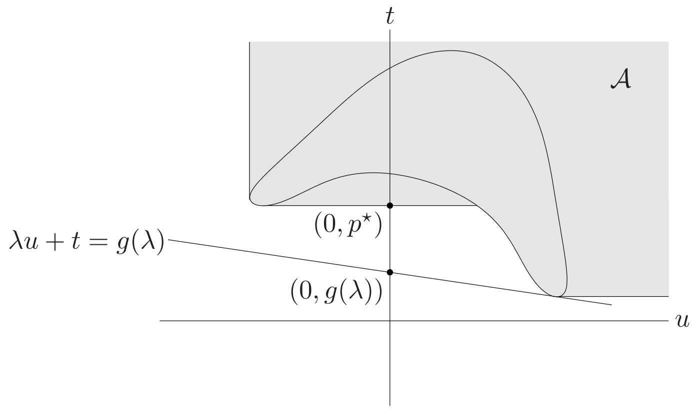

Figure 5.5 Geometric interpretation of dual function and lower bound $g(\lambda) \leq p^\star$, for a problem with one (inequality) constraint. Given λ, we minimize $(\lambda, 1)^T(u, t)$ over $\mathcal{A} = \{(u, t) \mid \exists x \in \mathcal{D},\ f_0(x) \leq t,\ f_1(x) \leq u\}$. This yields a supporting hyperplane with slope $-\lambda$. The intersection of this hyperplane with the $u = 0$ axis gives $g(\lambda)$.

We can think of $\mathcal{A}$ as a sort of epigraph form of $\mathcal{G}$, since $\mathcal{A}$ includes all the points in $\mathcal{G}$, as well as points that are 'worse', *i.e.*, those with larger objective or inequality constraint function values.

We can express the optimal value in terms of $\mathcal{A}$ as

$$p^\star = \inf\{t \mid (0, 0, t) \in \mathcal{A}\}.$$

To evaluate the dual function at a point (λ, ν) with $\lambda \succeq 0$, we can minimize the affine function $(\lambda, \nu, 1)^T(u, v, t)$ over $\mathcal{A}$: If $\lambda \succeq 0$, then

$$g(\lambda, \nu) = \inf\{(\lambda, \nu, 1)^T(u, v, t) \mid (u, v, t) \in \mathcal{A}\}.$$

If the infimum is finite, then

$$(\lambda, \nu, 1)^T(u, v, t) \geq g(\lambda, \nu)$$

defines a nonvertical supporting hyperplane to $\mathcal{A}$.

In particular, since $(0, 0, p^\star) \in \mathbf{bd}\,\mathcal{A}$, we have

$$p^\star = (\lambda, \nu, 1)^T(0, 0, p^\star) \geq g(\lambda, \nu), \tag{5.38}$$

the weak duality lower bound. Strong duality holds if and only if we have equality in (5.38) for some dual feasible (λ, ν), *i.e.*, there exists a nonvertical supporting hyperplane to $\mathcal{A}$ at its boundary point $(0, 0, p^\star)$.

This second interpretation is illustrated in figure 5.5.

5.3.2 Proof of strong duality under constraint qualification

In this section we prove that Slater's constraint qualification guarantees strong duality (and that the dual optimum is attained) for a convex problem. We consider the primal problem (5.25), with $f_0, \ldots, f_m$ convex, and assume Slater's condition holds: There exists $\tilde{x} \in \mathbf{relint}\,\mathcal{D}$ with $f_i(\tilde{x}) < 0$, $i = 1, \ldots, m$, and $A\tilde{x} = b$. In order to simplify the proof, we make two additional assumptions: first that $\mathcal{D}$ has nonempty interior (hence, $\mathbf{relint}\,\mathcal{D} = \mathbf{int}\,\mathcal{D}$) and second, that $\mathbf{rank}\,A = p$. We assume that $p^\star$ is finite. (Since there is a feasible point, we can only have $p^\star = -\infty$ or $p^\star$ finite; if $p^\star = -\infty$, then $d^\star = -\infty$ by weak duality.)

The set $\mathcal{A}$ defined in (5.37) is readily shown to be convex if the underlying problem is convex. We define a second convex set $\mathcal{B}$ as

$$\mathcal{B} = \{(0, 0, s) \in \mathbf{R}^m \times \mathbf{R}^p \times \mathbf{R} \mid s < p^\star\}.$$

The sets $\mathcal{A}$ and $\mathcal{B}$ do not intersect. To see this, suppose $(u, v, t) \in \mathcal{A} \cap \mathcal{B}$. Since $(u, v, t) \in \mathcal{B}$ we have $u = 0$, $v = 0$, and $t < p^\star$. Since $(u, v, t) \in \mathcal{A}$, there exists an x with $f_i(x) \le 0$, $i = 1, \ldots, m$, $Ax - b = 0$, and $f_0(x) \le t < p^\star$, which is impossible since $p^\star$ is the optimal value of the primal problem.

By the separating hyperplane theorem of §2.5.1 there exists $(\tilde{\lambda}, \tilde{\nu}, \mu) \ne 0$ and α such that

$$(u, v, t) \in \mathcal{A} \implies \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \ge \alpha, \tag{5.39}$$

and

$$(u, v, t) \in \mathcal{B} \implies \tilde{\lambda}^T u + \tilde{\nu}^T v + \mu t \le \alpha. \tag{5.40}$$

From (5.39) we conclude that $\tilde{\lambda} \succeq 0$ and $\mu \ge 0$. (Otherwise $\tilde{\lambda}^T u + \mu t$ is unbounded below over $\mathcal{A}$, contradicting (5.39).) The condition (5.40) simply means that $\mu t \le \alpha$ for all $t < p^\star$, and hence, $\mu p^\star \le \alpha$. Together with (5.39) we conclude that for any $x \in \mathcal{D}$,

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) + \mu f_0(x) \ge \alpha \ge \mu p^\star. \tag{5.41}$$

Assume that $\mu > 0$. In that case we can divide (5.41) by μ to obtain

$$L(x, \tilde{\lambda}/\mu, \tilde{\nu}/\mu) \ge p^\star$$

for all $x \in \mathcal{D}$, from which it follows, by minimizing over x, that $g(\lambda, \nu) \ge p^\star$, where we define

$$\lambda = \tilde{\lambda}/\mu, \qquad \nu = \tilde{\nu}/\mu.$$

By weak duality we have $g(\lambda, \nu) \le p^\star$, so in fact $g(\lambda, \nu) = p^\star$. This shows that strong duality holds, and that the dual optimum is attained, at least in the case when $\mu > 0$.

Now consider the case $\mu = 0$. From (5.41), we conclude that for all $x \in \mathcal{D}$,

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(x) + \tilde{\nu}^T (Ax - b) \ge 0. \tag{5.42}$$

Applying this to the point $\tilde{x}$ that satisfies the Slater condition, we have

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{x}) \ge 0.$$

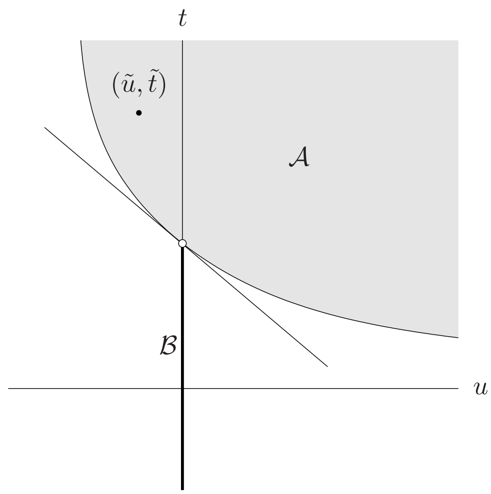

Figure 5.6 Illustration of strong duality proof, for a convex problem that satisfies Slater's constraint qualification. The set $\mathcal{A}$ is shown shaded, and the set $\mathcal{B}$ is the thick vertical line segment, not including the point $(0, p^\star)$, shown as a small open circle. The two sets are convex and do not intersect, so they can be separated by a hyperplane. Slater's constraint qualification guarantees that any separating hyperplane must be nonvertical, since it must pass to the left of the point $(\tilde{u}, \tilde{t}) = (f_1(\tilde{x}), f_0(\tilde{x}))$, where $\tilde{x}$ is strictly feasible.

Since $f_i(\tilde{x}) < 0$ and $\tilde{\lambda}_i \geq 0$, we conclude that $\tilde{\lambda} = 0$. From $(\tilde{\lambda}, \tilde{\nu}, \mu) \neq 0$ and $\tilde{\lambda} = 0$, $\mu = 0$, we conclude that $\tilde{\nu} \neq 0$. Then (5.42) implies that for all $x \in \mathcal{D}$, $\tilde{\nu}^T(Ax - b) \geq 0$. But $\tilde{x}$ satisfies $\tilde{\nu}^T(A\tilde{x} - b) = 0$, and since $\tilde{x} \in \mathbf{int}\, \mathcal{D}$, there are points in $\mathcal{D}$ with $\tilde{\nu}^T(Ax - b) < 0$ unless $A^T\tilde{\nu} = 0$. This, of course, contradicts our assumption that $\mathbf{rank}\, A = p$.

The geometric idea behind the proof is illustrated in figure 5.6, for a simple problem with one inequality constraint. The hyperplane separating $\mathcal{A}$ and $\mathcal{B}$ defines a supporting hyperplane to $\mathcal{A}$ at $(0, p^\star)$. Slater's constraint qualification is used to establish that the hyperplane must be nonvertical (*i.e.*, has a normal vector of the form $(\lambda^\star, 1)$). (For a simple example of a convex problem with one inequality constraint for which strong duality fails, see exercise 5.21.)

5.3.3 Multicriterion interpretation

There is a natural connection between Lagrange duality for a problem without equality constraints,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m, \end{array} \tag{5.43}$$

and the scalarization method for the (unconstrained) multicriterion problem

$$\text{minimize (w.r.t. } \mathbf{R}_{+}^{m+1}) \quad F(x) = (f_1(x), \ldots, f_m(x), f_0(x)) \tag{5.44}$$

(see §4.7.4). In scalarization, we choose a positive vector $\tilde{\lambda}$, and minimize the scalar function $\tilde{\lambda}^T F(x)$; any minimizer is guaranteed to be Pareto optimal. Since we can scale $\tilde{\lambda}$ by a positive constant, without affecting the minimizers, we can, without loss of generality, take $\tilde{\lambda} = (\lambda, 1)$. Thus, in scalarization we minimize the function

$$\tilde{\lambda}^T F(x) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x),$$

which is exactly the Lagrangian for the problem (5.43).

To establish that every Pareto optimal point of a convex multicriterion problem minimizes the function $\tilde{\lambda}^T F(x)$ for some nonnegative weight vector $\tilde{\lambda}$, we considered the set $\mathcal{A}$, defined in (4.62),

$$\mathcal{A} = \{t \in \mathbf{R}^{m+1} \mid \exists x \in \mathcal{D},\ f_i(x) \le t_i,\ i = 0, \ldots, m\},$$

which is exactly the same as the set $\mathcal{A}$ defined in (5.37), that arises in Lagrange duality. Here too we constructed the required weight vector as a supporting hyperplane to the set, at an arbitrary Pareto optimal point. In multicriterion optimization, we interpret the components of the weight vector as giving the relative weights between the objective functions. When we fix the last component of the weight vector (associated with f_0) to be one, the other weights have the interpretation of the cost relative to f_0, *i.e.*, the cost relative to the objective.

5.4 Saddle-point interpretation

In this section we give several interpretations of Lagrange duality. The material of this section will not be used in the sequel.

5.4.1 Max-min characterization of weak and strong duality

It is possible to express the primal and the dual optimization problems in a form that is more symmetric. To simplify the discussion we assume there are no equality constraints; the results are easily extended to cover them.

First note that

$$\begin{aligned} \sup_{\lambda \succeq 0} L(x, \lambda) &= \sup_{\lambda \succeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right) \\ &= \begin{cases} f_0(x) & f_i(x) \le 0, \quad i = 1, \ldots, m \\ \infty & \text{otherwise.} \end{cases} \end{aligned}$$

Indeed, suppose x is not feasible, and $f_i(x) > 0$ for some i. Then $\sup_{\lambda\succeq 0} L(x,\lambda) = \infty$, as can be seen by choosing $\lambda_j = 0$, $j \neq i$, and $\lambda_i \to \infty$. On the other hand, if $f_i(x) \leq 0$, $i = 1,\ldots,m$, then the optimal choice of λ is $\lambda = 0$ and $\sup_{\lambda\succeq 0} L(x,\lambda) = f_0(x)$. This means that we can express the optimal value of the primal problem as

$$p^\star = \inf_x \sup_{\lambda\succeq 0} L(x,\lambda).$$

By the definition of the dual function, we also have

$$d^\star = \sup_{\lambda\succeq 0} \inf_x L(x,\lambda).$$

Thus, weak duality can be expressed as the inequality

$$\sup_{\lambda\succeq 0} \inf_x L(x,\lambda) \leq \inf_x \sup_{\lambda\succeq 0} L(x,\lambda), \tag{5.45}$$

and strong duality as the equality

$$\sup_{\lambda\succeq 0} \inf_x L(x,\lambda) = \inf_x \sup_{\lambda\succeq 0} L(x,\lambda).$$

Strong duality means that the order of the minimization over x and the maximization over $\lambda \succeq 0$ can be switched without affecting the result.

In fact, the inequality (5.45) does not depend on any properties of L: We have

$$\sup_{z\in Z} \inf_{w\in W} f(w,z) \leq \inf_{w\in W} \sup_{z\in Z} f(w,z) \tag{5.46}$$

for any $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ (and any $W \subseteq \mathbf{R}^n$ and $Z \subseteq \mathbf{R}^m$). This general inequality is called the *max-min inequality*. When equality holds, *i.e.*,

$$\sup_{z\in Z} \inf_{w\in W} f(w,z) = \inf_{w\in W} \sup_{z\in Z} f(w,z) \tag{5.47}$$

we say that f (and W and Z) satisfy the *strong max-min property* or the *saddle-point* property. Of course the strong max-min property holds only in special cases, for example, when $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$ is the Lagrangian of a problem for which strong duality obtains, $W = \mathbf{R}^n$, and $Z = \mathbf{R}^m_+$.

5.4.2 Saddle-point interpretation

We refer to a pair $\tilde{w} \in W$, $\tilde{z} \in Z$ as a *saddle-point* for f (and W and Z) if

$$f(\tilde{w},z) \leq f(\tilde{w},\tilde{z}) \leq f(w,\tilde{z})$$

for all $w \in W$ and $z \in Z$. In other words, $\tilde{w}$ minimizes $f(w,\tilde{z})$ (over $w \in W$) and $\tilde{z}$ maximizes $f(\tilde{w},z)$ (over $z \in Z$):

$$f(\tilde{w},\tilde{z}) = \inf_{w\in W} f(w,\tilde{z}), \qquad f(\tilde{w},\tilde{z}) = \sup_{z\in Z} f(\tilde{w},z).$$

This implies that the strong max-min property (5.47) holds, and that the common value is $f(\tilde{w}, \tilde{z})$.

Returning to our discussion of Lagrange duality, we see that if $x^\star$ and $\lambda^\star$ are primal and dual optimal points for a problem in which strong duality obtains, they form a saddle-point for the Lagrangian. The converse is also true: If (x, λ) is a saddle-point of the Lagrangian, then x is primal optimal, λ is dual optimal, and the optimal duality gap is zero.

5.4.3 Game interpretation

We can interpret the max-min inequality (5.46), the max-min equality (5.47), and the saddle-point property, in terms of a continuous *zero-sum game*. If the first player chooses $w \in W$, and the second player selects $z \in Z$, then player 1 pays an amount $f(w, z)$ to player 2. Player 1 therefore wants to minimize f, while player 2 wants to maximize f. (The game is called continuous since the choices are vectors, and not discrete.)

Suppose that player 1 makes his choice first, and then player 2, after learning the choice of player 1, makes her selection. Player 2 wants to maximize the payoff $f(w, z)$, and so will choose $z \in Z$ to maximize $f(w, z)$. The resulting payoff will be $\sup_{z \in Z} f(w, z)$, which depends on w, the choice of the first player. (We assume here that the supremum is achieved; if not the optimal payoff can be arbitrarily close to $\sup_{z \in Z} f(w, z)$.) Player 1 knows (or assumes) that player 2 will follow this strategy, and so will choose $w \in W$ to make this worst-case payoff to player 2 as small as possible. Thus player 1 chooses

$$\operatorname*{argmin}_{w \in W} \sup_{z \in Z} f(w, z),$$

which results in the payoff

$$\inf_{w \in W} \sup_{z \in Z} f(w, z)$$

from player 1 to player 2.

Now suppose the order of play is reversed: Player 2 must choose $z \in Z$ first, and then player 1 chooses $w \in W$ (with knowledge of z). Following a similar argument, if the players follow the optimal strategy, player 2 should choose $z \in Z$ to maximize $\inf_{w \in W} f(w, z)$, which results in the payoff of

$$\sup_{z \in Z} \inf_{w \in W} f(w, z)$$

from player 1 to player 2.

The max-min inequality (5.46) states the (intuitively obvious) fact that it is better for a player to go second, or more precisely, for a player to know his or her opponent's choice before choosing. In other words, the payoff to player 2 will be larger if player 1 must choose first. When the saddle-point property (5.47) holds, there is no advantage to playing second.

If $(\tilde{w}, \tilde{z})$ is a saddle-point for f (and W and Z), then it is called a *solution* of the game; $\tilde{w}$ is called the optimal choice or strategy for player 1, and $\tilde{z}$ is called the optimal choice or strategy for player 2. In this case there is no advantage to playing second.

Now consider the special case where the payoff function is the Lagrangian, $W = \mathbf{R}^n$ and $Z = \mathbf{R}^m_+$. Here player 1 chooses the primal variable x, while player 2 chooses the dual variable $\lambda \succeq 0$. By the argument above, the optimal choice for player 2, if she must choose first, is any $\lambda^\star$ which is dual optimal, which results in a payoff to player 2 of $d^\star$. Conversely, if player 1 must choose first, his optimal choice is any primal optimal $x^\star$, which results in a payoff of $p^\star$.

The optimal duality gap for the problem is exactly equal to the advantage afforded the player who goes second, *i.e.*, the player who has the advantage of knowing his or her opponent's choice before choosing. If strong duality holds, then there is no advantage to the players of knowing their opponent's choice.

5.4.4 Price or tax interpretation

Lagrange duality has an interesting economic interpretation. Suppose the variable x denotes how an enterprise operates and $f_0(x)$ denotes the cost of operating at x, *i.e.*, $-f_0(x)$ is the profit (say, in dollars) made at the operating condition x. Each constraint $f_i(x) \leq 0$ represents some limit, such as a limit on resources (*e.g.*, warehouse space, labor) or a regulatory limit (*e.g.*, environmental). The operating condition that maximizes profit while respecting the limits can be found by solving the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m. \end{array}$$

The resulting optimal profit is $-p^\star$.

Now imagine a second scenario in which the limits can be violated, by paying an additional cost which is linear in the amount of violation, measured by f_i. Thus the payment made by the enterprise for the ith limit or constraint is $\lambda_i f_i(x)$. Payments are also made *to* the firm for constraints that are not tight; if $f_i(x) < 0$, then $\lambda_i f_i(x)$ represents a payment to the firm. The coefficient λ_i has the interpretation of the price for violating $f_i(x) \leq 0$; its units are dollars per unit violation (as measured by f_i). For the same price the enterprise can sell any 'unused' portion of the ith constraint. We assume $\lambda_i \geq 0$, *i.e.*, the firm must pay for violations (and receives income if a constraint is not tight).

As an example, suppose the first constraint in the original problem, $f_1(x) \leq 0$, represents a limit on warehouse space (say, in square meters). In this new arrangement, we open the possibility that the firm can rent extra warehouse space at a cost of λ_1 dollars per square meter and also rent out unused space, at the same rate.

The total cost to the firm, for operating condition x, and constraint prices λ_i, is $L(x, \lambda) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x)$. The firm will obviously operate so as to minimize its total cost $L(x, \lambda)$, which yields a cost $g(\lambda)$. The dual function therefore represents the optimal cost to the firm, as a function of the constraint price vector λ. The optimal dual value, $d^\star$, is the optimal cost to the enterprise under the least favorable set of prices.

Using this interpretation we can paraphrase weak duality as follows: The optimal cost to the firm in the second scenario (in which constraint violations can be bought and sold) is less than or equal to the cost in the original situation (which has constraints that cannot be violated), even with the most unfavorable prices. This is obvious: If $x^\star$ is optimal in the first scenario, then the operating cost of $x^\star$ in the second scenario will be lower than $f_0(x^\star)$, since some income can be derived from the constraints that are not tight. The optimal duality gap is then the minimum possible advantage to the enterprise of being allowed to pay for constraint violations (and receive payments for nontight constraints).

Now suppose strong duality holds, and the dual optimum is attained. We can interpret a dual optimal $\lambda^\star$ as a set of prices for which there is no advantage to the firm in being allowed to pay for constraint violations (or receive payments for nontight constraints). For this reason a dual optimal $\lambda^\star$ is sometimes called a set of *shadow prices* for the original problem.

5.5 Optimality conditions

We remind the reader that we do not assume the problem (5.1) is convex, unless explicitly stated.

5.5.1 Certificate of suboptimality and stopping criteria

If we can find a dual feasible (λ, ν), we establish a lower bound on the optimal value of the primal problem: $p^\star \geq g(\lambda, \nu)$. Thus a dual feasible point (λ, ν) provides a *proof* or *certificate* that $p^\star \geq g(\lambda, \nu)$. Strong duality means there exist arbitrarily good certificates.

Dual feasible points allow us to bound how suboptimal a given feasible point is, without knowing the exact value of $p^\star$. Indeed, if x is primal feasible and (λ, ν) is dual feasible, then

$$f_0(x) - p^\star \leq f_0(x) - g(\lambda, \nu).$$

In particular, this establishes that x is ϵ-suboptimal, with $\epsilon = f_0(x) - g(\lambda, \nu)$. (It also establishes that (λ, ν) is ϵ-suboptimal for the dual problem.)

We refer to the gap between primal and dual objectives,

$$f_0(x) - g(\lambda, \nu),$$

as the *duality gap* associated with the primal feasible point x and dual feasible point (λ, ν). A primal dual feasible pair x, (λ, ν) localizes the optimal value of the primal (and dual) problems to an interval:

$$p^\star \in [g(\lambda, \nu), f_0(x)], \qquad d^\star \in [g(\lambda, \nu), f_0(x)],$$

the width of which is the duality gap.

If the duality gap of the primal dual feasible pair x, (λ, ν) is zero, *i.e.*, $f_0(x) = g(\lambda, \nu)$, then x is primal optimal and (λ, ν) is dual optimal. We can think of (λ, ν)

as a certificate that proves x is optimal (and, similarly, we can think of x as a certificate that proves (λ, ν) is dual optimal).

These observations can be used in optimization algorithms to provide nonheuristic stopping criteria. Suppose an algorithm produces a sequence of primal feasible $x^{(k)}$ and dual feasible $(\lambda^{(k)}, \nu^{(k)})$, for $k = 1, 2, \ldots$, and $\epsilon_{\rm abs} > 0$ is a given required absolute accuracy. Then the stopping criterion (*i.e.*, the condition for terminating the algorithm)

$$f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \leq \epsilon_{\rm abs}$$

guarantees that when the algorithm terminates, $x^{(k)}$ is $\epsilon_{\rm abs}$-suboptimal. Indeed, $(\lambda^{(k)}, \nu^{(k)})$ is a certificate that proves it. (Of course strong duality must hold if this method is to work for arbitrarily small tolerances $\epsilon_{\rm abs}$.)

A similar condition can be used to guarantee a given relative accuracy $\epsilon_{\rm rel} > 0$. If

$$g(\lambda^{(k)}, \nu^{(k)}) > 0, \qquad \frac{f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)})}{g(\lambda^{(k)}, \nu^{(k)})} \leq \epsilon_{\rm rel}$$

holds, or

$$f_0(x^{(k)}) < 0, \qquad \frac{f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)})}{-f_0(x^{(k)})} \leq \epsilon_{\rm rel}$$

holds, then $p^\star \neq 0$ and the relative error

$$\frac{f_0(x^{(k)}) - p^\star}{|p^\star|}$$

is guaranteed to be less than or equal to $\epsilon_{\rm rel}$.

5.5.2 Complementary slackness

Suppose that the primal and dual optimal values are attained and equal (so, in particular, strong duality holds). Let $x^\star$ be a primal optimal and $(\lambda^\star, \nu^\star)$ be a dual optimal point. This means that

$$\begin{array}{rcl} f_0(x^\star) & = & g(\lambda^\star, \nu^\star) \\ & = & \inf_x \left(f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \right) \\ & \leq & f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ & \leq & f_0(x^\star). \end{array}$$

The first line states that the optimal duality gap is zero, and the second line is the definition of the dual function. The third line follows since the infimum of the Lagrangian over x is less than or equal to its value at $x = x^\star$. The last inequality follows from $\lambda_i^\star \geq 0$, $f_i(x^\star) \leq 0$, $i = 1, \ldots, m$, and $h_i(x^\star) = 0$, $i = 1, \ldots, p$. We conclude that the two inequalities in this chain hold with equality.

We can draw several interesting conclusions from this. For example, since the inequality in the third line is an equality, we conclude that $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$ over x. (The Lagrangian $L(x, \lambda^\star, \nu^\star)$ can have other minimizers; $x^\star$ is simply *a* minimizer.)

Another important conclusion is that

$$\sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) = 0.$$

Since each term in this sum is nonpositive, we conclude that

$$\lambda_i^\star f_i(x^\star) = 0, \quad i = 1, \ldots, m. \tag{5.48}$$

This condition is known as *complementary slackness*; it holds for any primal optimal $x^\star$ and any dual optimal $(\lambda^\star, \nu^\star)$ (when strong duality holds). We can express the complementary slackness condition as

$$\lambda_i^\star > 0 \implies f_i(x^\star) = 0,$$

or, equivalently,

$$f_i(x^\star) < 0 \implies \lambda_i^\star = 0.$$

Roughly speaking, this means the ith optimal Lagrange multiplier is zero unless the ith constraint is active at the optimum.

5.5.3 KKT optimality conditions

We now assume that the functions $f_0, \ldots, f_m, h_1, \ldots, h_p$ are differentiable (and therefore have open domains), but we make no assumptions yet about convexity.

KKT conditions for nonconvex problems

As above, let $x^\star$ and $(\lambda^\star, \nu^\star)$ be any primal and dual optimal points with zero duality gap. Since $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$ over x, it follows that its gradient must vanish at $x^\star$, *i.e.*,

$$\nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star) = 0.$$

Thus we have

$$\begin{array}{rcll} f_i(x^\star) & \leq & 0, & i = 1, \ldots, m \\ h_i(x^\star) & = & 0, & i = 1, \ldots, p \\ \lambda_i^\star & \geq & 0, & i = 1, \ldots, m \\ \lambda_i^\star f_i(x^\star) & = & 0, & i = 1, \ldots, m \\ \nabla f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star \nabla h_i(x^\star) & = & 0, & \end{array} \tag{5.49}$$

which are called the *Karush-Kuhn-Tucker* (KKT) conditions.

To summarize, for *any* optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions (5.49).

KKT conditions for convex problems

When the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal. In other words, if f_i are convex and h_i are affine, and $\tilde{x}$, $\tilde{\lambda}$, $\tilde{\nu}$ are any points that satisfy the KKT conditions

$$\begin{array}{rcll} f_i(\tilde{x}) & \leq & 0, & i=1,\ldots,m \\ h_i(\tilde{x}) & = & 0, & i=1,\ldots,p \\ \tilde{\lambda}_i & \geq & 0, & i=1,\ldots,m \\ \tilde{\lambda}_i f_i(\tilde{x}) & = & 0, & i=1,\ldots,m \\ \nabla f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{x}) & = & 0, & \end{array}$$

then $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal and dual optimal, with zero duality gap.

To see this, note that the first two conditions state that $\tilde{x}$ is primal feasible. Since $\tilde{\lambda}_i \geq 0$, $L(x, \tilde{\lambda}, \tilde{\nu})$ is convex in x; the last KKT condition states that its gradient with respect to x vanishes at $x = \tilde{x}$, so it follows that $\tilde{x}$ minimizes $L(x, \tilde{\lambda}, \tilde{\nu})$ over x. From this we conclude that

$$\begin{array}{rcl} g(\tilde{\lambda}, \tilde{\nu}) & = & L(\tilde{x}, \tilde{\lambda}, \tilde{\nu}) \\ & = & f_0(\tilde{x}) + \displaystyle\sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i h_i(\tilde{x}) \\ & = & f_0(\tilde{x}), \end{array}$$

where in the last line we use $h_i(\tilde{x}) = 0$ and $\tilde{\lambda}_i f_i(\tilde{x}) = 0$. This shows that $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ have zero duality gap, and therefore are primal and dual optimal. In summary, for any *convex* optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal, and have zero duality gap.

If a convex optimization problem with differentiable objective and constraint functions satisfies Slater's condition, then the KKT conditions provide necessary and sufficient conditions for optimality: Slater's condition implies that the optimal duality gap is zero and the dual optimum is attained, so x is optimal if and only if there are (λ, ν) that, together with x, satisfy the KKT conditions.

The KKT conditions play an important role in optimization. In a few special cases it is possible to solve the KKT conditions (and therefore, the optimization problem) analytically. More generally, many algorithms for convex optimization are conceived as, or can be interpreted as, methods for solving the KKT conditions.

Example 5.1 *Equality constrained convex quadratic minimization.* We consider the problem

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b, \end{array} \tag{5.50}$$

where $P \in \mathbf{S}_+^n$. The KKT conditions for this problem are

$$Ax^\star = b, \qquad Px^\star + q + A^T \nu^\star = 0,$$

which we can write as

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

Solving this set of $m+n$ equations in the $m+n$ variables $x^\star$, $\nu^\star$ gives the optimal primal and dual variables for (5.50).

Example 5.2 *Water-filling.* We consider the convex optimization problem

$$\begin{array}{ll} \text{minimize} & -\sum_{i=1}^n \log(\alpha_i + x_i) \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1, \end{array}$$

where $\alpha_i > 0$. This problem arises in information theory, in allocating power to a set of n communication channels. The variable x_i represents the transmitter power allocated to the ith channel, and $\log(\alpha_i + x_i)$ gives the capacity or communication rate of the channel, so the problem is to allocate a total power of one to the channels, in order to maximize the total communication rate.

Introducing Lagrange multipliers $\lambda^\star \in \mathbf{R}^n$ for the inequality constraints $x^\star \succeq 0$, and a multiplier $\nu^\star \in \mathbf{R}$ for the equality constraint $\mathbf{1}^T x = 1$, we obtain the KKT conditions

$$x^\star \succeq 0, \qquad \mathbf{1}^T x^\star = 1, \qquad \lambda^\star \succeq 0, \qquad \lambda_i^\star x_i^\star = 0, \quad i = 1, \ldots, n,$$

$$-1/(\alpha_i + x_i^\star) - \lambda_i^\star + \nu^\star = 0, \quad i = 1, \ldots, n.$$

We can directly solve these equations to find $x^\star$, $\lambda^\star$, and $\nu^\star$. We start by noting that $\lambda^\star$ acts as a slack variable in the last equation, so it can be eliminated, leaving

$$x^\star \succeq 0, \qquad \mathbf{1}^T x^\star = 1, \qquad x_i^\star \left(\nu^\star - 1/(\alpha_i + x_i^\star)\right) = 0, \quad i = 1, \ldots, n,$$

$$\nu^\star \geq 1/(\alpha_i + x_i^\star), \quad i = 1, \ldots, n.$$

If $\nu^\star < 1/\alpha_i$, this last condition can only hold if $x_i^\star > 0$, which by the third condition implies that $\nu^\star = 1/(\alpha_i + x_i^\star)$. Solving for $x_i^\star$, we conclude that $x_i^\star = 1/\nu^\star - \alpha_i$ if $\nu^\star < 1/\alpha_i$. If $\nu^\star \geq 1/\alpha_i$, then $x_i^\star > 0$ is impossible, because it would imply $\nu^\star \geq 1/\alpha_i > 1/(\alpha_i + x_i^\star)$, which violates the complementary slackness condition. Therefore, $x_i^\star = 0$ if $\nu^\star \geq 1/\alpha_i$. Thus we have

$$x_i^\star = \begin{cases} 1/\nu^\star - \alpha_i & \nu^\star < 1/\alpha_i \\ 0 & \nu^\star \geq 1/\alpha_i, \end{cases}$$

or, put more simply, $x_i^\star = \max\{0, 1/\nu^\star - \alpha_i\}$. Substituting this expression for $x_i^\star$ into the condition $\mathbf{1}^T x^\star = 1$ we obtain

$$\sum_{i=1}^n \max\{0, 1/\nu^\star - \alpha_i\} = 1.$$

The lefthand side is a piecewise-linear increasing function of $1/\nu^\star$, with breakpoints at α_i, so the equation has a unique solution which is readily determined.

This solution method is called *water-filling* for the following reason. We think of α_i as the ground level above patch i, and then flood the region with water to a depth $1/\nu$, as illustrated in figure 5.7. The total amount of water used is then $\sum_{i=1}^n \max\{0, 1/\nu^\star - \alpha_i\}$. We then increase the flood level until we have used a total amount of water equal to one. The depth of water above patch i is then the optimal value $x_i^\star$.

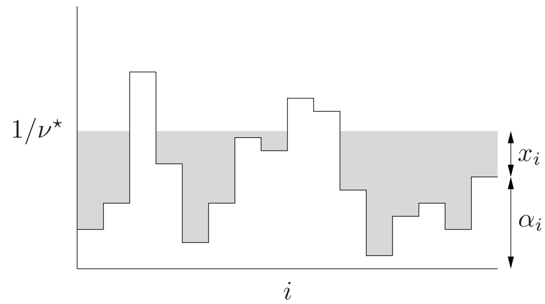

Figure 5.7 Illustration of water-filling algorithm. The height of each patch is given by α_i. The region is flooded to a level $1/\nu^\star$ which uses a total quantity of water equal to one. The height of the water (shown shaded) above each patch is the optimal value of $x_i^\star$.

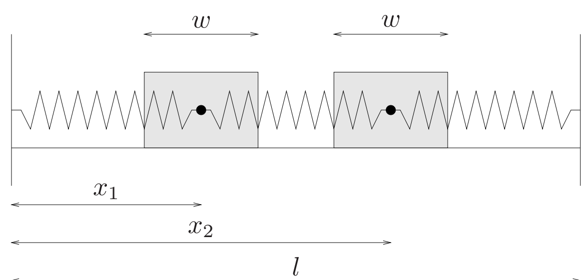

Figure 5.8 Two blocks connected by springs to each other, and the left and right walls. The blocks have width $w > 0$, and cannot penetrate each other or the walls.

5.5.4 Mechanics interpretation of KKT conditions

The KKT conditions can be given a nice interpretation in mechanics (which indeed, was one of Lagrange's primary motivations). We illustrate the idea with a simple example. The system shown in figure 5.8 consists of two blocks attached to each other, and to walls at the left and right, by three springs. The position of the blocks are given by $x \in \mathbf{R}^2$, where x_1 is the displacement of the (middle of the) left block, and x_2 is the displacement of the right block. The left wall is at position 0, and the right wall is at position l.

The potential energy in the springs, as a function of the block positions, is given by

$$f_0(x_1, x_2) = \frac{1}{2}k_1 x_1^2 + \frac{1}{2}k_2(x_2 - x_1)^2 + \frac{1}{2}k_3(l - x_2)^2,$$

where $k_i > 0$ are the stiffness constants of the three springs. The equilibrium position $x^\star$ is the position that minimizes the potential energy subject to the inequalities

$$w/2 - x_1 \leq 0, \qquad w + x_1 - x_2 \leq 0, \qquad w/2 - l + x_2 \leq 0. \tag{5.51}$$

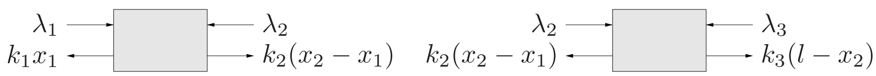

Figure 5.9 Force analysis of the block-spring system. The total force on each block, due to the springs and also to contact forces, must be zero. The Lagrange multipliers, shown on top, are the contact forces between the walls and blocks. The spring forces are shown at bottom.

These constraints are called *kinematic constraints*, and express the fact that the blocks have width $w > 0$, and cannot penetrate each other or the walls. The equilibrium position is therefore given by the solution of the optimization problem

$$\begin{array}{ll} \text{minimize} & (1/2)\left(k_1x_1^2 + k_2(x_2-x_1)^2 + k_3(l-x_2)^2\right) \\ \text{subject to} & w/2 - x_1 \le 0 \\ & w + x_1 - x_2 \le 0 \\ & w/2 - l + x_2 \le 0, \end{array} \qquad (5.52)$$

which is a QP.

With λ_1, λ_2, λ_3 as Lagrange multipliers, the KKT conditions for this problem consist of the kinematic constraints (5.51), the nonnegativity constraints $\lambda_i \ge 0$, the complementary slackness conditions

$$\lambda_1(w/2 - x_1) = 0, \qquad \lambda_2(w - x_2 + x_1) = 0, \qquad \lambda_3(w/2 - l + x_2) = 0, \qquad (5.53)$$

and the zero gradient condition

$$\left[\begin{array}{c} k_1x_1 - k_2(x_2-x_1) \\ k_2(x_2-x_1) - k_3(l-x_2) \end{array}\right] + \lambda_1 \left[\begin{array}{c} -1 \\ 0 \end{array}\right] + \lambda_2 \left[\begin{array}{c} 1 \\ -1 \end{array}\right] + \lambda_3 \left[\begin{array}{c} 0 \\ 1 \end{array}\right] = 0. \qquad (5.54)$$

The equation (5.54) can be interpreted as the force balance equations for the two blocks, provided we interpret the Lagrange multipliers as *contact forces* that act between the walls and blocks, as illustrated in figure 5.9. The first equation states that the sum of the forces on the first block is zero: The term $-k_1x_1$ is the force exerted on the left block by the left spring, the term $k_2(x_2-x_1)$ is the force exerted by the middle spring, λ_1 is the force exerted by the left wall, and $-\lambda_2$ is the force exerted by the right block. The contact forces must point away from the contact surface (as expressed by the constraints $\lambda_1 \ge 0$ and $-\lambda_2 \le 0$), and are nonzero only when there is contact (as expressed by the first two complementary slackness conditions (5.53)). In a similar way, the second equation in (5.54) is the force balance for the second block, and the last condition in (5.53) states that λ_3 is zero unless the right block touches the wall.

In this example, the potential energy and kinematic constraint functions are convex, and (the refined form of) Slater's constraint qualification holds provided $2w \le l$, *i.e.*, there is enough room between the walls to fit the two blocks, so we can conclude that the energy formulation of the equilibrium given by (5.52), gives the same result as the force balance formulation, given by the KKT conditions.

5.5.5 Solving the primal problem via the dual

We mentioned at the beginning of §5.5.3 that if strong duality holds and a dual optimal solution $(\lambda^\star, \nu^\star)$ exists, then any primal optimal point is also a minimizer of $L(x, \lambda^\star, \nu^\star)$. This fact sometimes allows us to compute a primal optimal solution from a dual optimal solution.

More precisely, suppose we have strong duality and an optimal $(\lambda^\star, \nu^\star)$ is known. Suppose that the minimizer of $L(x, \lambda^\star, \nu^\star)$, *i.e.*, the solution of

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x), \tag{5.55}$$

is unique. (For a convex problem this occurs, for example, if $L(x, \lambda^\star, \nu^\star)$ is a strictly convex function of x.) Then if the solution of (5.55) is primal feasible, it must be primal optimal; if it is not primal feasible, then no primal optimal point can exist, *i.e.*, we can conclude that the primal optimum is not attained. This observation is interesting when the dual problem is easier to solve than the primal problem, for example, because it can be solved analytically, or has some special structure that can be exploited.

Example 5.3 *Entropy maximization.* We consider the entropy maximization problem

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1 \end{array}$$

with domain $\mathbf{R}_{++}^n$, and its dual problem

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T\lambda} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

where a_i are the columns of A (see pages 222 and 228). We assume that the weak form of Slater's condition holds, *i.e.*, there exists an $x \succ 0$ with $Ax \preceq b$ and $\mathbf{1}^T x = 1$, so strong duality holds and an optimal solution $(\lambda^\star, \nu^\star)$ exists.

Suppose we have solved the dual problem. The Lagrangian at $(\lambda^\star, \nu^\star)$ is

$$L(x, \lambda^\star, \nu^\star) = \sum_{i=1}^n x_i \log x_i + \lambda^{\star T}(Ax - b) + \nu^\star(\mathbf{1}^T x - 1)$$

which is strictly convex on $\mathcal{D}$ and bounded below, so it has a unique solution $x^\star$, given by

$$x_i^\star = 1/\exp(a_i^T\lambda^\star + \nu^\star + 1), \quad i = 1, \ldots, n.$$

If $x^\star$ is primal feasible, it must be the optimal solution of the primal problem (5.13). If $x^\star$ is not primal feasible, then we can conclude that the primal optimum is not attained.

Example 5.4 *Minimizing a separable function subject to an equality constraint.* We consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n f_i(x_i) \\ \text{subject to} & a^T x = b, \end{array}$$

where $a \in \mathbf{R}^n$, $b \in \mathbf{R}$, and $f_i : \mathbf{R} \to \mathbf{R}$ are differentiable and strictly convex. The objective function is called *separable* since it is a sum of functions of the individual variables $x_1, \ldots, x_n$. We assume that the domain of f_0 intersects the constraint set, *i.e.*, there exists a point $x_0 \in \mathbf{dom}\, f_0$ with $a^T x_0 = b$. This implies the problem has a unique optimal point $x^\star$.

The Lagrangian is

$$L(x, \nu) = \sum_{i=1}^n f_i(x_i) + \nu(a^T x - b) = -b\nu + \sum_{i=1}^n (f_i(x_i) + \nu a_i x_i),$$

which is also separable, so the dual function is

$$\begin{aligned} g(\nu) &= -b\nu + \inf_x \left(\sum_{i=1}^n (f_i(x_i) + \nu a_i x_i) \right) \\ &= -b\nu + \sum_{i=1}^n \inf_{x_i} (f_i(x_i) + \nu a_i x_i) \\ &= -b\nu - \sum_{i=1}^n f_i^*(-\nu a_i). \end{aligned}$$

The dual problem is thus

$$\text{maximize} \quad -b\nu - \textstyle\sum_{i=1}^n f_i^*(-\nu a_i),$$

with (scalar) variable $\nu \in \mathbf{R}$.

Now suppose we have found an optimal dual variable $\nu^\star$. (There are several simple methods for solving a convex problem with one scalar variable, such as the bisection method.) Since each f_i is strictly convex, the function $L(x, \nu^\star)$ is strictly convex in x, and so has a unique minimizer $\tilde{x}$. But we also know that $x^\star$ minimizes $L(x, \nu^\star)$, so we must have $\tilde{x} = x^\star$. We can recover $x^\star$ from $\nabla_x L(x, \nu^\star) = 0$, *i.e.*, by solving the equations $f_i'(x_i^\star) = -\nu^\star a_i$.

5.6 Perturbation and sensitivity analysis

When strong duality obtains, the optimal dual variables give very useful information about the sensitivity of the optimal value with respect to perturbations of the constraints.

5.6.1 The perturbed problem

We consider the following perturbed version of the original optimization problem (5.1):

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p, \end{array} \tag{5.56}$$

with variable $x \in \mathbf{R}^n$. This problem coincides with the original problem (5.1) when $u = 0$, $v = 0$. When u_i is positive it means that we have relaxed the ith inequality constraint; when u_i is negative, it means that we have tightened the constraint. Thus the perturbed problem (5.56) results from the original problem (5.1) by tightening or relaxing each inequality constraint by u_i, and changing the righthand side of the equality constraints by v_i.

We define $p^\star(u, v)$ as the optimal value of the perturbed problem (5.56):

$$\begin{aligned} p^\star(u,v) = \inf\{f_0(x) \mid \exists x \in \mathcal{D},\ & f_i(x) \le u_i,\ i = 1,\dots,m, \\ & h_i(x) = v_i,\ i = 1,\dots,p\}. \end{aligned}$$

We can have $p^\star(u, v) = \infty$, which corresponds to perturbations of the constraints that result in infeasibility. Note that $p^\star(0, 0) = p^\star$, the optimal value of the unperturbed problem (5.1). (We hope this slight abuse of notation will cause no confusion.) Roughly speaking, the function $p^\star : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ gives the optimal value of the problem as a function of perturbations to the righthand sides of the constraints.

When the original problem is convex, the function $p^\star$ is a convex function of u and v; indeed, its epigraph is precisely the closure of the set $\mathcal{A}$ defined in (5.37) (see exercise 5.32).

5.6.2 A global inequality

Now we assume that strong duality holds, and that the dual optimum is attained. (This is the case if the original problem is convex, and Slater's condition is satisfied). Let $(\lambda^\star, \nu^\star)$ be optimal for the dual (5.16) of the unperturbed problem. Then for all u and v we have

$$p^\star(u,v) \ge p^\star(0,0) - {\lambda^\star}^T u - {\nu^\star}^T v. \tag{5.57}$$

To establish this inequality, suppose that x is any feasible point for the perturbed problem, *i.e.*, $f_i(x) \le u_i$ for $i = 1,\dots,m$, and $h_i(x) = v_i$ for $i = 1,\dots,p$. Then we have, by strong duality,

$$\begin{aligned} p^\star(0,0) = g(\lambda^\star,\nu^\star) &\le f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \\ &\le f_0(x) + {\lambda^\star}^T u + {\nu^\star}^T v. \end{aligned}$$

(The first inequality follows from the definition of $g(\lambda^\star, \nu^\star)$; the second follows since $\lambda^\star \succeq 0$.) We conclude that for any x feasible for the perturbed problem, we have

$$f_0(x) \ge p^\star(0,0) - {\lambda^\star}^T u - {\nu^\star}^T v,$$

from which (5.57) follows.

Sensitivity interpretations

When strong duality holds, various sensitivity interpretations of the optimal Lagrange variables follow directly from the inequality (5.57). Some of the conclusions are:

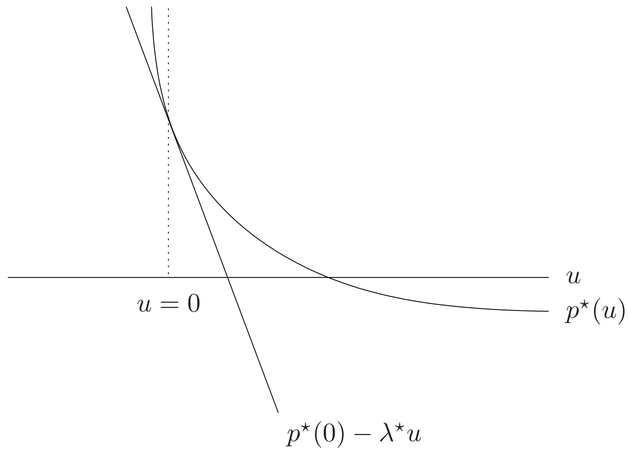

Figure 5.10 Optimal value $p^\star(u)$ of a convex problem with one constraint $f_1(x) \leq u$, as a function of u. For $u = 0$, we have the original unperturbed problem; for $u < 0$ the constraint is tightened, and for $u > 0$ the constraint is loosened. The affine function $p^\star(0) - \lambda^\star u$ is a lower bound on $p^\star$.

- If $\lambda_i^\star$ is large and we tighten the ith constraint (*i.e.*, choose $u_i < 0$), then the optimal value $p^\star(u, v)$ is guaranteed to increase greatly.
- If $\nu_i^\star$ is large and positive and we take $v_i < 0$, or if $\nu_i^\star$ is large and negative and we take $v_i > 0$, then the optimal value $p^\star(u, v)$ is guaranteed to increase greatly.
- If $\lambda_i^\star$ is small, and we loosen the ith constraint ($u_i > 0$), then the optimal value $p^\star(u, v)$ will not decrease too much.
- If $\nu_i^\star$ is small and positive, and $v_i > 0$, or if $\nu_i^\star$ is small and negative and $v_i < 0$, then the optimal value $p^\star(u, v)$ will not decrease too much.

The inequality (5.57), and the conclusions listed above, give a *lower bound* on the perturbed optimal value, but no upper bound. For this reason the results are *not* symmetric with respect to loosening or tightening a constraint. For example, suppose that $\lambda_i^\star$ is large, and we loosen the ith constraint a bit (*i.e.*, take u_i small and positive). In this case the inequality (5.57) is not useful; it does not, for example, imply that the optimal value will decrease considerably.

The inequality (5.57) is illustrated in figure 5.10 for a convex problem with one inequality constraint. The inequality states that the affine function $p^\star(0) - \lambda^\star u$ is a lower bound on the convex function $p^\star$.

5.6.3 Local sensitivity analysis

Suppose now that $p^\star(u, v)$ is differentiable at $u = 0$, $v = 0$. Then, provided strong duality holds, the optimal dual variables $\lambda^\star$, $\nu^\star$ are related to the gradient of $p^\star$ at

$u = 0$, $v = 0$:

$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \qquad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}. \tag{5.58}$$

This property can be seen in the example shown in figure 5.10, where $-\lambda^\star$ is the slope of $p^\star$ near $u = 0$.

Thus, when $p^\star(u, v)$ is differentiable at $u = 0$, $v = 0$, and strong duality holds, the optimal Lagrange multipliers are exactly the local sensitivities of the optimal value with respect to constraint perturbations. In contrast to the nondifferentiable case, this interpretation *is* symmetric: Tightening the ith inequality constraint a small amount (*i.e.*, taking u_i small and negative) yields an increase in $p^\star$ of approximately $-\lambda_i^\star u_i$; loosening the ith constraint a small amount (*i.e.*, taking u_i small and positive) yields a decrease in $p^\star$ of approximately $\lambda_i^\star u_i$.

To show (5.58), suppose $p^\star(u, v)$ is differentiable and strong duality holds. For the perturbation $u = te_i$, $v = 0$, where e_i is the ith unit vector, we have

$$\lim_{t \to 0} \frac{p^\star(te_i, 0) - p^\star}{t} = \frac{\partial p^\star(0,0)}{\partial u_i}.$$

The inequality (5.57) states that for $t > 0$,

$$\frac{p^\star(te_i, 0) - p^\star}{t} \geq -\lambda_i^\star,$$

while for $t < 0$ we have the opposite inequality. Taking the limit $t \to 0$, with $t > 0$, yields

$$\frac{\partial p^\star(0,0)}{\partial u_i} \geq -\lambda_i^\star,$$

while taking the limit with $t < 0$ yields the opposite inequality, so we conclude that

$$\frac{\partial p^\star(0,0)}{\partial u_i} = -\lambda_i^\star.$$

The same method can be used to establish

$$\frac{\partial p^\star(0,0)}{\partial v_i} = -\nu_i^\star.$$

The local sensitivity result (5.58) gives us a quantitative measure of how active a constraint is at the optimum $x^\star$. If $f_i(x^\star) < 0$, then the constraint is inactive, and it follows that the constraint can be tightened or loosened a small amount without affecting the optimal value. By complementary slackness, the associated optimal Lagrange multiplier must be zero. But now suppose that $f_i(x^\star) = 0$, *i.e.*, the ith constraint is active at the optimum. The ith optimal Lagrange multiplier tells us how active the constraint is: If $\lambda_i^\star$ is small, it means that the constraint can be loosened or tightened a bit without much effect on the optimal value; if $\lambda_i^\star$ is large, it means that if the constraint is loosened or tightened a bit, the effect on the optimal value will be great.

Shadow price interpretation

We can also give a simple geometric interpretation of the result (5.58) in terms of economics. We consider (for simplicity) a convex problem with no equality constraints, which satisfies Slater's condition. The variable $x \in \mathbf{R}^m$ determines how a firm operates, and the objective f_0 is the cost, *i.e.*, $-f_0$ is the profit. Each constraint $f_i(x) \leq 0$ represents a limit on some resource such as labor, steel, or warehouse space. The (negative) perturbed optimal cost function $-p^\star(u)$ tells us how much more or less profit could be made if more, or less, of each resource were made available to the firm. If it is differentiable near $u = 0$, then we have

$$\lambda_i^\star = -\frac{\partial p^\star(0)}{\partial u_i}.$$

In other words, $\lambda_i^\star$ tells us approximately how much more profit the firm could make, for a small increase in availability of resource i.

It follows that $\lambda_i^\star$ would be the natural or equilibrium *price* for resource i, if it were possible for the firm to buy or sell it. Suppose, for example, that the firm can buy or sell resource i, at a price that is less than $\lambda_i^\star$. In this case it would certainly buy some of the resource, which would allow it to operate in a way that increases its profit more than the cost of buying the resource. Conversely, if the price exceeds $\lambda_i^\star$, the firm would sell some of its allocation of resource i, and obtain a net gain since its income from selling some of the resource would be larger than its drop in profit due to the reduction in availability of the resource.

5.7 Examples

In this section we show by example that simple equivalent reformulations of a problem can lead to very different dual problems. We consider the following types of reformulations:

- Introducing new variables and associated equality constraints.
- Replacing the objective with an increasing function of the original objective.
- Making explicit constraints implicit, *i.e.*, incorporating them into the domain of the objective.

5.7.1 Introducing new variables and equality constraints

Consider an unconstrained problem of the form

$$\text{minimize} \quad f_0(Ax + b). \tag{5.59}$$

Its Lagrange dual function is the constant $p^\star$. So while we do have strong duality, *i.e.*, $p^\star = d^\star$, the Lagrangian dual is neither useful nor interesting.

Now let us reformulate the problem (5.59) as

$$\begin{array}{ll} \text{minimize} & f_0(y) \\ \text{subject to} & Ax + b = y. \end{array} \tag{5.60}$$

Here we have introduced new variables y, as well as new equality constraints $Ax + b = y$. The problems (5.59) and (5.60) are clearly equivalent.

The Lagrangian of the reformulated problem is

$$L(x, y, \nu) = f_0(y) + \nu^T(Ax + b - y).$$

To find the dual function we minimize L over x and y. Minimizing over x we find that $g(\nu) = -\infty$ unless $A^T\nu = 0$, in which case we are left with

$$g(\nu) = b^T\nu + \inf_y (f_0(y) - \nu^T y) = b^T\nu - f_0^*(\nu),$$

where f_0^* is the conjugate of f_0. The dual problem of (5.60) can therefore be expressed as

$$\begin{array}{ll} \text{maximize} & b^T\nu - f_0^*(\nu) \\ \text{subject to} & A^T\nu = 0. \end{array} \tag{5.61}$$

Thus, the dual of the reformulated problem (5.60) is considerably more useful than the dual of the original problem (5.59).

Example 5.5 *Unconstrained geometric program.* Consider the unconstrained geometric program

$$\begin{array}{ll} \text{minimize} & \log\left(\sum_{i=1}^m \exp(a_i^T x + b_i)\right). \end{array}$$

We first reformulate it by introducing new variables and equality constraints:

$$\begin{array}{ll} \text{minimize} & f_0(y) = \log\left(\sum_{i=1}^m \exp y_i\right) \\ \text{subject to} & Ax + b = y, \end{array}$$

where a_i^T are the rows of A. The conjugate of the log-sum-exp function is

$$f_0^*(\nu) = \begin{cases} \sum_{i=1}^m \nu_i \log \nu_i & \nu \succeq 0,\ \mathbf{1}^T\nu = 1 \\ \infty & \text{otherwise} \end{cases}$$

(example 3.25, page 93), so the dual of the reformulated problem can be expressed as

$$\begin{array}{ll} \text{maximize} & b^T\nu - \sum_{i=1}^m \nu_i \log \nu_i \\ \text{subject to} & \mathbf{1}^T\nu = 1 \\ & A^T\nu = 0 \\ & \nu \succeq 0, \end{array} \tag{5.62}$$

which is an entropy maximization problem.

Example 5.6 *Norm approximation problem.* We consider the unconstrained norm approximation problem

$$\text{minimize} \quad \|Ax - b\|, \tag{5.63}$$

where $\|\cdot\|$ is any norm. Here too the Lagrange dual function is constant, equal to the optimal value of (5.63), and therefore not useful.

Once again we reformulate the problem as

$$\begin{array}{ll} \text{minimize} & \|y\| \\ \text{subject to} & Ax - b = y. \end{array}$$

The Lagrange dual problem is, following (5.61),

$$\begin{array}{ll} \text{maximize} & b^T \nu \\ \text{subject to} & \|\nu\|_* \leq 1 \\ & A^T \nu = 0, \end{array} \tag{5.64}$$

where we use the fact that the conjugate of a norm is the indicator function of the dual norm unit ball (example 3.26, page 93).

The idea of introducing new equality constraints can be applied to the constraint functions as well. Consider, for example, the problem

$$\begin{array}{ll} \text{minimize} & f_0(A_0 x + b_0) \\ \text{subject to} & f_i(A_i x + b_i) \leq 0, \quad i = 1, \ldots, m, \end{array} \tag{5.65}$$

where $A_i \in \mathbf{R}^{k_i \times n}$ and $f_i : \mathbf{R}^{k_i} \to \mathbf{R}$ are convex. (For simplicity we do not include equality constraints here.) We introduce a new variable $y_i \in \mathbf{R}^{k_i}$, for $i = 0, \ldots, m$, and reformulate the problem as

$$\begin{array}{ll} \text{minimize} & f_0(y_0) \\ \text{subject to} & f_i(y_i) \leq 0, \quad i = 1, \ldots, m \\ & A_i x + b_i = y_i, \quad i = 0, \ldots, m. \end{array} \tag{5.66}$$

The Lagrangian for this problem is

$$L(x, y_0, \ldots, y_m, \lambda, \nu_0, \ldots, \nu_m) = f_0(y_0) + \sum_{i=1}^m \lambda_i f_i(y_i) + \sum_{i=0}^m \nu_i^T (A_i x + b_i - y_i).$$

To find the dual function we minimize over x and y_i. The minimum over x is $-\infty$ unless

$$\sum_{i=0}^m A_i^T \nu_i = 0,$$

in which case we have, for $\lambda \succ 0$,

$$\begin{aligned} & g(\lambda, \nu_0, \ldots, \nu_m) \\ & \quad = \sum_{i=0}^m \nu_i^T b_i + \inf_{y_0, \ldots, y_m} \left(f_0(y_0) + \sum_{i=1}^m \lambda_i f_i(y_i) - \sum_{i=0}^m \nu_i^T y_i \right) \\ & \quad = \sum_{i=0}^m \nu_i^T b_i + \inf_{y_0} \left(f_0(y_0) - \nu_0^T y_0 \right) + \sum_{i=1}^m \lambda_i \inf_{y_i} \left(f_i(y_i) - (\nu_i / \lambda_i)^T y_i \right) \\ & \quad = \sum_{i=0}^m \nu_i^T b_i - f_0^*(\nu_0) - \sum_{i=1}^m \lambda_i f_i^*(\nu_i / \lambda_i). \end{aligned}$$

The last expression involves the perspective of the conjugate function, and is therefore concave in the dual variables. Finally, we address the question of what happens when $\lambda \succeq 0$, but some λ_i are zero. If $\lambda_i = 0$ and $\nu_i \neq 0$, then the dual function is $-\infty$. If $\lambda_i = 0$ and $\nu_i = 0$, however, the terms involving y_i, ν_i, and λ_i are all zero. Thus, the expression above for g is valid for all $\lambda \succeq 0$, if we take $\lambda_i f_i^*(\nu_i/\lambda_i) = 0$ when $\lambda_i = 0$ and $\nu_i = 0$, and $\lambda_i f_i^*(\nu_i/\lambda_i) = \infty$ when $\lambda_i = 0$ and $\nu_i \neq 0$.

Therefore we can express the dual of the problem (5.66) as

$$
\begin{array}{ll}
\text{maximize} & \sum_{i=0}^m \nu_i^T b_i - f_0^*(\nu_0) - \sum_{i=1}^m \lambda_i f_i^*(\nu_i/\lambda_i) \\
\text{subject to} & \lambda \succeq 0 \\
& \sum_{i=0}^m A_i^T \nu_i = 0.
\end{array}
\tag{5.67}
$$

Example 5.7 *Inequality constrained geometric program.* The inequality constrained geometric program

$$
\begin{array}{ll}
\text{minimize} & \log\left(\sum_{k=1}^{K_0} e^{a_{0k}^T x + b_{0k}}\right) \\
\text{subject to} & \log\left(\sum_{k=1}^{K_i} e^{a_{ik}^T x + b_{ik}}\right) \leq 0, \quad i = 1, \ldots, m
\end{array}
$$

is of the form (5.65) with $f_i : \mathbf{R}^{K_i} \to \mathbf{R}$ given by $f_i(y) = \log\left(\sum_{k=1}^{K_i} e^{y_k}\right)$. The conjugate of this function is

$$
f_i^*(\nu) = \begin{cases} \sum_{k=1}^{K_i} \nu_k \log \nu_k & \nu \succeq 0, \quad \mathbf{1}^T \nu = 1 \\ \infty & \text{otherwise.} \end{cases}
$$

Using (5.67) we can immediately write down the dual problem as

$$
\begin{array}{ll}
\text{maximize} & b_0^T \nu_0 - \sum_{k=1}^{K_0} \nu_{0k} \log \nu_{0k} + \sum_{i=1}^m \left(b_i^T \nu_i - \sum_{k=1}^{K_i} \nu_{ik} \log(\nu_{ik}/\lambda_i)\right) \\
\text{subject to} & \nu_0 \succeq 0, \quad \mathbf{1}^T \nu_0 = 1 \\
& \nu_i \succeq 0, \quad \mathbf{1}^T \nu_i = \lambda_i, \quad i = 1, \ldots, m \\
& \lambda_i \geq 0, \quad i = 1, \ldots, m \\
& \sum_{i=0}^m A_i^T \nu_i = 0,
\end{array}
$$

which further simplifies to

$$
\begin{array}{ll}
\text{maximize} & b_0^T \nu_0 - \sum_{k=1}^{K_0} \nu_{0k} \log \nu_{0k} + \sum_{i=1}^m \left(b_i^T \nu_i - \sum_{k=1}^{K_i} \nu_{ik} \log(\nu_{ik}/\mathbf{1}^T \nu_i)\right) \\
\text{subject to} & \nu_i \succeq 0, \quad i = 0, \ldots, m \\
& \mathbf{1}^T \nu_0 = 1 \\
& \sum_{i=0}^m A_i^T \nu_i = 0.
\end{array}
$$

5.7.2 Transforming the objective

If we replace the objective f_0 by an increasing function of f_0, the resulting problem is clearly equivalent (see §4.1.3). The dual of this equivalent problem, however, can be very different from the dual of the original problem.

Example 5.8 We consider again the minimum norm problem

$$
\text{minimize} \quad \|Ax - b\|,
$$

where $\|\cdot\|$ is some norm. We reformulate this problem as

$$\begin{array}{ll} \text{minimize} & (1/2)\|y\|^2 \\ \text{subject to} & Ax - b = y. \end{array}$$

Here we have introduced new variables, and replaced the objective by half its square. Evidently it is equivalent to the original problem.

The dual of the reformulated problem is

$$\begin{array}{ll} \text{maximize} & -(1/2)\|\nu\|_*^2 + b^T\nu \\ \text{subject to} & A^T\nu = 0, \end{array}$$

where we use the fact that the conjugate of $(1/2)\|\cdot\|^2$ is $(1/2)\|\cdot\|_*^2$ (see example 3.27, page 93).

Note that this dual problem is not the same as the dual problem (5.64) derived earlier.

5.7.3 Implicit constraints

The next simple reformulation we study is to include some of the constraints in the objective function, by modifying the objective function to be infinite when the constraint is violated.

Example 5.9 *Linear program with box constraints.* We consider the linear program

$$\begin{array}{ll} \text{minimize} & c^Tx \\ \text{subject to} & Ax = b \\ & l \preceq x \preceq u \end{array} \tag{5.68}$$

where $A \in \mathbf{R}^{p\times n}$ and $l \prec u$. The constraints $l \preceq x \preceq u$ are sometimes called *box constraints* or *variable bounds*.

We can, of course, derive the dual of this linear program. The dual will have a Lagrange multiplier ν associated with the equality constraint, λ_1 associated with the inequality constraint $x \preceq u$, and λ_2 associated with the inequality constraint $l \preceq x$. The dual is

$$\begin{array}{ll} \text{maximize} & -b^T\nu - \lambda_1^Tu + \lambda_2^Tl \\ \text{subject to} & A^T\nu + \lambda_1 - \lambda_2 + c = 0 \\ & \lambda_1 \succeq 0, \quad \lambda_2 \succeq 0. \end{array} \tag{5.69}$$

Instead, let us first reformulate the problem (5.68) as

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax = b, \end{array} \tag{5.70}$$

where we define

$$f_0(x) = \begin{cases} c^Tx & l \preceq x \preceq u \\ \infty & \text{otherwise.} \end{cases}$$

The problem (5.70) is clearly equivalent to (5.68); we have merely made the explicit box constraints implicit.

The dual function for the problem (5.70) is

$$\begin{aligned} g(\nu) &= \inf_{l \preceq x \preceq u} \left(c^T x + \nu^T (Ax - b) \right) \\ &= -b^T \nu - u^T (A^T \nu + c)^- + l^T (A^T \nu + c)^+ \end{aligned}$$

where $y_i^+ = \max\{y_i, 0\}$, $y_i^- = \max\{-y_i, 0\}$. So here we are able to derive an analytical formula for g, which is a concave piecewise-linear function.

The dual problem is the unconstrained problem

$$\text{maximize} \quad -b^T \nu - u^T (A^T \nu + c)^- + l^T (A^T \nu + c)^+, \tag{5.71}$$

which has a quite different form from the dual of the original problem.

(The problems (5.69) and (5.71) are closely related, in fact, equivalent; see exercise 5.8.)

5.8 Theorems of alternatives

5.8.1 Weak alternatives via the dual function

In this section we apply Lagrange duality theory to the problem of determining feasibility of a system of inequalities and equalities

$$f_i(x) \leq 0, \quad i = 1, \ldots, m, \qquad h_i(x) = 0, \quad i = 1, \ldots, p. \tag{5.72}$$

We assume the domain of the inequality system (5.72), $\mathcal{D} = \bigcap_{i=1}^m \mathbf{dom}\, f_i \cap \bigcap_{i=1}^p \mathbf{dom}\, h_i$, is nonempty. We can think of (5.72) as the standard problem (5.1), with objective $f_0 = 0$, *i.e.*,

$$\begin{array}{ll} \text{minimize} & 0 \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p. \end{array} \tag{5.73}$$

This problem has optimal value

$$p^\star = \begin{cases} 0 & \text{(5.72) is feasible} \\ \infty & \text{(5.72) is infeasible,} \end{cases} \tag{5.74}$$

so solving the optimization problem (5.73) is the same as solving the inequality system (5.72).

The dual function

We associate with the inequality system (5.72) the dual function

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \left(\sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right),$$

which is the same as the dual function for the optimization problem (5.73). Since $f_0 = 0$, the dual function is positive homogeneous in (λ, ν): For $\alpha > 0$, $g(\alpha\lambda, \alpha\nu) = \alpha g(\lambda, \nu)$. The dual problem associated with (5.73) is to maximize $g(\lambda, \nu)$ subject to $\lambda \succeq 0$. Since g is homogeneous, the optimal value of this dual problem is given by

$$d^\star = \begin{cases} \infty & \lambda \succeq 0,\ g(\lambda, \nu) > 0 \text{ is feasible} \\ 0 & \lambda \succeq 0,\ g(\lambda, \nu) > 0 \text{ is infeasible.} \end{cases} \tag{5.75}$$

Weak duality tells us that $d^\star \leq p^\star$. Combining this fact with (5.74) and (5.75) yields the following: If the inequality system

$$\lambda \succeq 0, \qquad g(\lambda, \nu) > 0 \tag{5.76}$$

is feasible (which means $d^\star = \infty$), then the inequality system (5.72) is infeasible (since we then have $p^\star = \infty$). Indeed, we can interpret any solution (λ, ν) of the inequalities (5.76) as a *proof* or *certificate* of infeasibility of the system (5.72).

We can restate this implication in terms of feasibility of the original system: If the original inequality system (5.72) is feasible, then the inequality system (5.76) must be infeasible. We can interpret an x which satisfies (5.72) as a certificate establishing infeasibility of the inequality system (5.76).

Two systems of inequalities (and equalities) are called *weak alternatives* if at most one of the two is feasible. Thus, the systems (5.72) and (5.76) are weak alternatives. This is true whether or not the inequalities (5.72) are convex (*i.e.*, f_i convex, h_i affine); moreover, the alternative inequality system (5.76) is always convex (*i.e.*, g is concave and the constraints $\lambda_i \geq 0$ are convex).

Strict inequalities

We can also study feasibility of the *strict* inequality system

$$f_i(x) < 0, \quad i = 1, \ldots, m, \qquad h_i(x) = 0, \quad i = 1, \ldots, p. \tag{5.77}$$

With g defined as for the nonstrict inequality system, we have the alternative inequality system

$$\lambda \succeq 0, \qquad \lambda \neq 0, \qquad g(\lambda, \nu) \geq 0. \tag{5.78}$$

We can show directly that (5.77) and (5.78) are weak alternatives. Suppose there exists an $\tilde{x}$ with $f_i(\tilde{x}) < 0$, $h_i(\tilde{x}) = 0$. Then for any $\lambda \succeq 0$, $\lambda \neq 0$, and ν,

$$\lambda_1 f_1(\tilde{x}) + \cdots + \lambda_m f_m(\tilde{x}) + \nu_1 h_1(\tilde{x}) + \cdots + \nu_p h_p(\tilde{x}) < 0.$$

It follows that

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} \left(\sum_{i=1}^m \lambda_i f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right) \\ &\leq \sum_{i=1}^m \lambda_i f_i(\tilde{x}) + \sum_{i=1}^p \nu_i h_i(\tilde{x}) \\ &< 0. \end{aligned}$$

Therefore, feasibility of (5.77) implies that there does not exist (λ, ν) satisfying (5.78).

Thus, we can prove infeasibility of (5.77) by producing a solution of the system (5.78); we can prove infeasibility of (5.78) by producing a solution of the system (5.77).

5.8.2 Strong alternatives

When the original inequality system is convex, *i.e.*, f_i are convex and h_i are affine, and some type of constraint qualification holds, then the pairs of weak alternatives described above are *strong alternatives*, which means that *exactly one* of the two alternatives holds. In other words, each of the inequality systems is feasible if and only if the other is infeasible.

In this section we assume that f_i are convex and h_i are affine, so the inequality system (5.72) can be expressed as

$$f_i(x) \leq 0, \quad i = 1, \ldots, m, \qquad Ax = b,$$

where $A \in \mathbf{R}^{p \times n}$.

Strict inequalities

We first study the strict inequality system

$$f_i(x) < 0, \quad i = 1, \ldots, m, \qquad Ax = b, \tag{5.79}$$

and its alternative

$$\lambda \succeq 0, \qquad \lambda \neq 0, \qquad g(\lambda, \nu) \geq 0. \tag{5.80}$$

We need one technical condition: There exists an $x \in \mathbf{relint}\, \mathcal{D}$ with $Ax = b$. In other words we not only assume that the linear equality constraints are consistent, but also that they have a solution in $\mathbf{relint}\, \mathcal{D}$. (Very often $\mathcal{D} = \mathbf{R}^n$, so the condition is satisfied if the equality constraints are consistent.) Under this condition, exactly one of the inequality systems (5.79) and (5.80) is feasible. In other words, the inequality systems (5.79) and (5.80) are strong alternatives.

We will establish this result by considering the related optimization problem

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & f_i(x) - s \leq 0, \quad i = 1, \ldots, m \\ & Ax = b \end{array} \tag{5.81}$$

with variables x, s, and domain $\mathcal{D} \times \mathbf{R}$. The optimal value $p^\star$ of this problem is negative if and only if there exists a solution to the strict inequality system (5.79).

The Lagrange dual function for the problem (5.81) is

$$\inf_{x \in \mathcal{D},\, s} \left(s + \sum_{i=1}^{m} \lambda_i (f_i(x) - s) + \nu^T (Ax - b) \right) = \left\{ \begin{array}{ll} g(\lambda, \nu) & \mathbf{1}^T \lambda = 1 \\ -\infty & \text{otherwise.} \end{array} \right.$$

Therefore we can express the dual problem of (5.81) as

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0, \quad \mathbf{1}^T \lambda = 1. \end{array}$$

Now we observe that Slater's condition holds for the problem (5.81). By the hypothesis there exists an $\tilde{x} \in \textbf{relint}\, \mathcal{D}$ with $A\tilde{x} = b$. Choosing any $\tilde{s} > \max_i f_i(\tilde{x})$ yields a point $(\tilde{x}, \tilde{s})$ which is strictly feasible for (5.81). Therefore we have $d^\star = p^\star$, and the dual optimum $d^\star$ is attained. In other words, there exist $(\lambda^\star, \nu^\star)$ such that

$$g(\lambda^\star, \nu^\star) = p^\star, \qquad \lambda^\star \succeq 0, \qquad \mathbf{1}^T \lambda^\star = 1. \tag{5.82}$$

Now suppose that the strict inequality system (5.79) is infeasible, which means that $p^\star \geq 0$. Then $(\lambda^\star, \nu^\star)$ from (5.82) satisfy the alternate inequality system (5.80). Similarly, if the alternate inequality system (5.80) is feasible, then $d^\star = p^\star \geq 0$, which shows that the strict inequality system (5.79) is infeasible. Thus, the inequality systems (5.79) and (5.80) are strong alternatives; each is feasible if and only if the other is not.

Nonstrict inequalities

We now consider the nonstrict inequality system

$$f_i(x) \leq 0, \quad i = 1, \ldots, m, \qquad Ax = b, \tag{5.83}$$

and its alternative

$$\lambda \succeq 0, \qquad g(\lambda, \nu) > 0. \tag{5.84}$$

We will show these are strong alternatives, provided the following conditions hold: There exists an $x \in \textbf{relint}\, \mathcal{D}$ with $Ax = b$, and the optimal value $p^\star$ of (5.81) is attained. This holds, for example, if $\mathcal{D} = \mathbf{R}^n$ and $\max_i f_i(x) \to \infty$ as $x \to \infty$. With these assumptions we have, as in the strict case, that $p^\star = d^\star$, and that both the primal and dual optimal values are attained. Now suppose that the nonstrict inequality system (5.83) is infeasible, which means that $p^\star > 0$. (Here we use the assumption that the primal optimal value is attained.) Then $(\lambda^\star, \nu^\star)$ from (5.82) satisfy the alternate inequality system (5.84). Thus, the inequality systems (5.83) and (5.84) are strong alternatives; each is feasible if and only if the other is not.

5.8.3 Examples

Linear inequalities

Consider the system of linear inequalities $Ax \preceq b$. The dual function is

$$g(\lambda) = \inf_x \lambda^T (Ax - b) = \begin{cases} -b^T \lambda & A^T \lambda = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The alternative inequality system is therefore

$$\lambda \succeq 0, \qquad A^T \lambda = 0, \qquad b^T \lambda < 0.$$

These are, in fact, strong alternatives. This follows since the optimum in the related problem (5.81) is achieved, unless it is unbounded below.

We now consider the system of strict linear inequalities $Ax \prec b$, which has the strong alternative system

$$\lambda \succeq 0, \qquad \lambda \neq 0, \qquad A^T\lambda = 0, \qquad b^T\lambda \leq 0.$$

In fact we have encountered (and proved) this result before, in §2.5.1; see (2.17) and (2.18) (on page 50).

Intersection of ellipsoids

We consider m ellipsoids, described as

$$\mathcal{E}_i = \{x \mid f_i(x) \leq 0\},$$

with $f_i(x) = x^TA_ix + 2b_i^Tx + c_i$, $i = 1, \ldots, m$, where $A_i \in \mathbf{S}^n_{++}$. We ask when the intersection of these ellipsoids has nonempty interior. This is equivalent to feasibility of the set of strict quadratic inequalities

$$f_i(x) = x^TA_ix + 2b_i^Tx + c_i < 0, \quad i = 1, \ldots, m. \tag{5.85}$$

The dual function g is

$$\begin{aligned} g(\lambda) &= \inf_x \left(x^TA(\lambda)x + 2b(\lambda)^Tx + c(\lambda)\right) \\ &= \begin{cases} -b(\lambda)^TA(\lambda)^\dagger b(\lambda) + c(\lambda) & A(\lambda) \succeq 0, \quad b(\lambda) \in \mathcal{R}(A(\lambda)) \\ -\infty & \text{otherwise,} \end{cases} \end{aligned}$$

where

$$A(\lambda) = \sum_{i=1}^m \lambda_i A_i, \qquad b(\lambda) = \sum_{i=1}^m \lambda_i b_i, \qquad c(\lambda) = \sum_{i=1}^m \lambda_i c_i.$$

Note that for $\lambda \succeq 0$, $\lambda \neq 0$, we have $A(\lambda) \succ 0$, so we can simplify the expression for the dual function as

$$g(\lambda) = -b(\lambda)^TA(\lambda)^{-1}b(\lambda) + c(\lambda).$$

The strong alternative of the system (5.85) is therefore

$$\lambda \succeq 0, \qquad \lambda \neq 0, \qquad -b(\lambda)^TA(\lambda)^{-1}b(\lambda) + c(\lambda) \geq 0. \tag{5.86}$$

We can give a simple geometric interpretation of this pair of strong alternatives. For any nonzero $\lambda \succeq 0$, the (possibly empty) ellipsoid

$$\mathcal{E}_\lambda = \{x \mid x^TA(\lambda)x + 2b(\lambda)^Tx + c(\lambda) \leq 0\}$$

contains $\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_m$, since $f_i(x) \leq 0$ implies $\sum_{i=1}^m \lambda_i f_i(x) \leq 0$. Now, $\mathcal{E}_\lambda$ has empty interior if and only if

$$\inf_x \left(x^TA(\lambda)x + 2b(\lambda)^Tx + c(\lambda)\right) = -b(\lambda)^TA(\lambda)^{-1}b(\lambda) + c(\lambda) \geq 0.$$

Therefore the alternative system (5.86) means that $\mathcal{E}_\lambda$ has empty interior.

Weak duality is obvious: If (5.86) holds, then $\mathcal{E}_\lambda$ contains the intersection $\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_m$, and has empty interior, so naturally the intersection has empty interior. The fact that these are strong alternatives states the (not obvious) fact that if the intersection $\mathcal{E}_1 \cap \cdots \cap \mathcal{E}_m$ has empty interior, then we can construct an ellipsoid $\mathcal{E}_\lambda$ that contains the intersection and has empty interior.

Farkas' lemma

In this section we describe a pair of strong alternatives for a mixture of strict and nonstrict linear inequalities, known as *Farkas' lemma*: The system of inequalities

$$Ax \preceq 0, \qquad c^T x < 0, \tag{5.87}$$

where $A \in \mathbf{R}^{m\times n}$ and $c \in \mathbf{R}^n$, and the system of equalities and inequalities

$$A^T y + c = 0, \qquad y \succeq 0, \tag{5.88}$$

are strong alternatives.

We can prove Farkas' lemma directly, using LP duality. Consider the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq 0, \end{array} \tag{5.89}$$

and its dual

$$\begin{array}{ll} \text{maximize} & 0 \\ \text{subject to} & A^T y + c = 0 \\ & y \succeq 0. \end{array} \tag{5.90}$$

The primal LP (5.89) is homogeneous, and so has optimal value 0, if (5.87) is not feasible, and optimal value $-\infty$, if (5.87) is feasible. The dual LP (5.90) has optimal value 0, if (5.88) is feasible, and optimal value $-\infty$, if (5.88) is infeasible.

Since $x = 0$ is feasible in (5.89), we can rule out the one case in which strong duality can fail for LPs, so we must have $p^\star = d^\star$. Combined with the remarks above, this shows that (5.87) and (5.88) are strong alternatives.

Example 5.10 *Arbitrage-free bounds on price.* We consider a set of n assets, with prices at the beginning of an investment period $p_1, \ldots, p_n$, respectively. At the end of the investment period, the value of the assets is $v_1, \ldots, v_n$. If $x_1, \ldots, x_n$ represents the initial investment in each asset (with $x_j < 0$ meaning a short position in asset j), the cost of the initial investment is $p^T x$, and the final value of the investment is $v^T x$.

The value of the assets at the end of the investment period, v, is uncertain. We will assume that only m possible scenarios, or outcomes, are possible. If outcome i occurs, the final value of the assets is $v^{(i)}$, and therefore, the overall value of the investments is $v^{(i)T} x$.

If there is an investment vector x with $p^T x < 0$, and in all possible scenarios, the final value is nonnegative, *i.e.*, $v^{(i)T} x \geq 0$ for $i = 1, \ldots, m$, then an *arbitrage* is said to exist. The condition $p^T x < 0$ means you are *paid* to accept the investment mix, and the condition $v^{(i)T} x \geq 0$ for $i = 1, \ldots, m$ means that no matter what outcome occurs, the final value is nonnegative, so an arbitrage corresponds to a guaranteed money-making investment strategy. It is generally assumed that the prices and values are such that no arbitrage exists. This means that the inequality system

$$Vx \succeq 0, \qquad p^T x < 0$$

is infeasible, where $V_{ij} = v_j^{(i)}$.

Using Farkas' lemma, we have no arbitrage if and only if there exists y such that

$$-V^T y + p = 0, \qquad y \succeq 0.$$

We can use this characterization of arbitrage-free prices and values to solve several interesting problems.

Suppose, for example, that the values V are known, and all prices except the last one, p_n, are known. The set of prices p_n that are consistent with the no-arbitrage assumption is an interval, which can be found by solving a pair of LPs. The optimal value of the LP

$$\begin{array}{ll} \text{minimize} & p_n \\ \text{subject to} & V^T y = p, \quad y \succeq 0, \end{array}$$

with variables p_n and y, gives the smallest possible arbitrage-free price for asset n. Solving the same LP with maximization instead of minimization yields the largest possible price for asset n. If the two values are equal, *i.e.*, the no-arbitrage assumption leads us to a unique price for asset n, we say the market is *complete*. For an example, see exercise 5.38.

This method can be used to find bounds on the price of a derivative or option that is based on the final value of other underlying assets, *i.e.*, when the value or payoff of asset n is a function of the values of the other assets.

5.9 Generalized inequalities

In this section we examine how Lagrange duality extends to a problem with generalized inequality constraints

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p, \end{array} \tag{5.91}$$

where $K_i \subseteq \mathbf{R}^{k_i}$ are proper cones. For now, we do not assume convexity of the problem (5.91). We assume the domain of (5.91), $\mathcal{D} = \bigcap_{i=0}^m \mathbf{dom}\, f_i \,\cap\, \bigcap_{i=1}^p \mathbf{dom}\, h_i$, is nonempty.

5.9.1 The Lagrange dual

With each generalized inequality $f_i(x) \preceq_{K_i} 0$ in (5.91) we associate a Lagrange multiplier *vector* $\lambda_i \in \mathbf{R}^{k_i}$ and define the associated Lagrangian as

$$L(x, \lambda, \nu) = f_0(x) + \lambda_1^T f_1(x) + \cdots + \lambda_m^T f_m(x) + \nu_1 h_1(x) + \cdots + \nu_p h_p(x),$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)$ and $\nu = (\nu_1, \ldots, \nu_p)$. The dual function is defined exactly as in a problem with scalar inequalities:

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i^T f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right).$$

Since the Lagrangian is affine in the dual variables (λ, ν), and the dual function is a pointwise infimum of the Lagrangian, the dual function is concave.

As in a problem with scalar inequalities, the dual function gives lower bounds on $p^\star$, the optimal value of the primal problem (5.91). For a problem with scalar inequalities, we require $\lambda_i \geq 0$. Here the nonnegativity requirement on the dual variables is replaced by the condition

$$\lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m,$$

where K_i^* denotes the dual cone of K_i. In other words, the Lagrange multipliers associated with inequalities must be *dual* nonnegative.

Weak duality follows immediately from the definition of dual cone. If $\lambda_i \succeq_{K_i^*} 0$ and $f_i(\tilde{x}) \preceq_{K_i} 0$, then $\lambda_i^T f_i(\tilde{x}) \leq 0$. Therefore for any primal feasible point $\tilde{x}$ and any $\lambda_i \succeq_{K_i^*} 0$, we have

$$f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq f_0(\tilde{x}).$$

Taking the infimum over $\tilde{x}$ yields $g(\lambda, \nu) \leq p^\star$.

The Lagrange dual optimization problem is

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m. \end{array} \tag{5.92}$$

We always have *weak duality*, *i.e.*, $d^\star \leq p^\star$, where $d^\star$ denotes the optimal value of the dual problem (5.92), whether or not the primal problem (5.91) is convex.

Slater's condition and strong duality

As might be expected, *strong* duality ($d^\star = p^\star$) holds when the primal problem is convex and satisfies an appropriate constraint qualification. For example, a generalized version of Slater's condition for the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\ & Ax = b, \end{array}$$

where f_0 is convex and f_i is K_i-convex, is that there exists an $x \in \mathbf{relint}\, \mathcal{D}$ with $Ax = b$ and $f_i(x) \prec_{K_i} 0$, $i = 1, \ldots, m$. This condition implies strong duality (and also, that the dual optimum is attained).

Example 5.11 *Lagrange dual of semidefinite program.* We consider a semidefinite program in inequality form,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + \cdots + x_n F_n + G \preceq 0 \end{array} \tag{5.93}$$

where $F_1, \ldots, F_n, G \in \mathbf{S}^k$. (Here f_1 is affine, and K_1 is $\mathbf{S}_+^k$, the positive semidefinite cone.)

We associate with the constraint a dual variable or multiplier $Z \in \mathbf{S}^k$, so the Lagrangian is

$$\begin{array}{rcl} L(x, Z) & = & c^T x + \mathbf{tr}\left((x_1 F_1 + \cdots + x_n F_n + G)\, Z\right) \\ & = & x_1(c_1 + \mathbf{tr}(F_1 Z)) + \cdots + x_n(c_n + \mathbf{tr}(F_n Z)) + \mathbf{tr}(GZ), \end{array}$$

which is affine in x. The dual function is given by

$$g(Z) = \inf_x L(x, Z) = \begin{cases} \mathbf{tr}(GZ) & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \dots, n \\ -\infty & \text{otherwise.} \end{cases}$$

The dual problem can therefore be expressed as

$$\begin{array}{ll} \text{maximize} & \mathbf{tr}(GZ) \\ \text{subject to} & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \dots, n \\ & Z \succeq 0. \end{array}$$

(We use the fact that $\mathbf{S}^k_+$ is self-dual, *i.e.*, $(\mathbf{S}^k_+)^* = \mathbf{S}^k_+$; see §2.6.)

Strong duality obtains if the semidefinite program (5.93) is strictly feasible, *i.e.*, there exists an x with

$$x_1 F_1 + \cdots + x_n F_n + G \prec 0.$$

Example 5.12 *Lagrange dual of cone program in standard form.* We consider the cone program

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq_K 0, \end{array}$$

where $A \in \mathbf{R}^{m \times n}$, $b \in \mathbf{R}^m$, and $K \subseteq \mathbf{R}^n$ is a proper cone. We associate with the equality constraint a multiplier $\nu \in \mathbf{R}^m$, and with the nonnegativity constraint a multiplier $\lambda \in \mathbf{R}^n$. The Lagrangian is

$$L(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (Ax - b),$$

so the dual function is

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

The dual problem can be expressed as

$$\begin{array}{ll} \text{maximize} & -b^T \nu \\ \text{subject to} & A^T \nu + c = \lambda \\ & \lambda \succeq_{K^*} 0. \end{array}$$

By eliminating λ and defining $y = -\nu$, this problem can be simplified to

$$\begin{array}{ll} \text{maximize} & b^T y \\ \text{subject to} & A^T y \preceq_{K^*} c, \end{array}$$

which is a cone program in inequality form, involving the dual generalized inequality.

Strong duality obtains if the Slater condition holds, *i.e.*, there is an $x \succ_K 0$ with $Ax = b$.

5.9.2 Optimality conditions

The optimality conditions of §5.5 are readily extended to problems with generalized inequalities. We first derive the complementary slackness conditions.

Complementary slackness

Assume that the primal and dual optimal values are equal, and attained at the optimal points $x^\star$, $\lambda^\star$, $\nu^\star$. As in §5.5.2, the complementary slackness conditions follow directly from the equality $f_0(x^\star) = g(\lambda^\star, \nu^\star)$, along with the definition of g. We have

$$\begin{aligned} f_0(x^\star) &= g(\lambda^\star, \nu^\star) \\ &\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^{\star T} f_i(x^\star) + \sum_{i=1}^p \nu_i^\star h_i(x^\star) \\ &\leq f_0(x^\star), \end{aligned}$$

and therefore we conclude that $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$, and also that the two sums in the second line are zero. Since the second sum is zero (since $x^\star$ satisfies the equality constraints), we have $\sum_{i=1}^m \lambda_i^{\star T} f_i(x^\star) = 0$. Since each term in this sum is nonpositive, we conclude that

$$\lambda_i^{\star T} f_i(x^\star) = 0, \quad i = 1, \ldots, m, \tag{5.94}$$

which generalizes the complementary slackness condition (5.48). From (5.94) we can conclude that

$$\lambda_i^\star \succ_{K_i^*} 0 \implies f_i(x^\star) = 0, \qquad f_i(x^\star) \prec_{K_i} 0, \implies \lambda_i^\star = 0.$$

However, in contrast to problems with scalar inequalities, it is possible to satisfy (5.94) with $\lambda_i^\star \neq 0$ and $f_i(x^\star) \neq 0$.

KKT conditions

Now we add the assumption that the functions f_i, h_i are differentiable, and generalize the KKT conditions of §5.5.3 to problems with generalized inequalities. Since $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$, its gradient with respect to x vanishes at $x^\star$:

$$\nabla f_0(x^\star) + \sum_{i=1}^m Df_i(x^\star)^T \lambda_i^\star + \sum_{i=1}^p \nu_i^\star \nabla h_i(x^\star) = 0,$$

where $Df_i(x^\star) \in \mathbf{R}^{k_i \times n}$ is the derivative of f_i evaluated at $x^\star$ (see §A.4.1). Thus, if strong duality holds, any primal optimal $x^\star$ and any dual optimal $(\lambda^\star, \nu^\star)$ must satisfy the optimality conditions (or KKT conditions)

$$\begin{array}{rcll} f_i(x^\star) & \preceq_{K_i} & 0, & i = 1, \ldots, m \\ h_i(x^\star) & = & 0, & i = 1, \ldots, p \\ \lambda_i^\star & \succeq_{K_i^*} & 0, & i = 1, \ldots, m \\ \lambda_i^{\star T} f_i(x^\star) & = & 0, & i = 1, \ldots, m \\ \nabla f_0(x^\star) + \sum_{i=1}^m Df_i(x^\star)^T \lambda_i^\star + \sum_{i=1}^p \nu_i^\star \nabla h_i(x^\star) & = & 0. & \end{array} \tag{5.95}$$

If the primal problem is convex, the converse also holds, *i.e.*, the conditions (5.95) are sufficient conditions for optimality of $x^\star$, $(\lambda^\star, \nu^\star)$.

5.9.3 Perturbation and sensitivity analysis

The results of §5.6 can be extended to problems involving generalized inequalities. We consider the associated perturbed version of the problem,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \preceq_{K_i} u_i, \quad i = 1, \ldots, m \\ & h_i(x) = v_i, \quad i = 1, \ldots, p, \end{array}$$

where $u_i \in \mathbf{R}^{k_i}$, and $v \in \mathbf{R}^p$. We define $p^\star(u, v)$ as the optimal value of the perturbed problem. As in the case with scalar inequalities, $p^\star$ is a convex function when the original problem is convex.

Now let $(\lambda^\star, \nu^\star)$ be optimal for the dual of the original (unperturbed) problem, which we assume has zero duality gap. Then for all u and v we have

$$p^\star(u, v) \geq p^\star - \sum_{i=1}^{m} \lambda_i^{\star T} u_i - \nu^{\star T} v,$$

the analog of the global sensitivity inequality (5.57). The local sensitivity result holds as well: If $p^\star(u, v)$ is differentiable at $u = 0$, $v = 0$, then the optimal dual variables $\lambda_i^\star$ satisfies

$$\lambda_i^\star = -\nabla_{u_i} p^\star(0, 0),$$

the analog of (5.58).

Example 5.13 *Semidefinite program in inequality form.* We consider a semidefinite program in inequality form, as in example 5.11. The primal problem is

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F(x) = x_1 F_1 + \cdots + x_n F_n + G \preceq 0, \end{array}$$

with variable $x \in \mathbf{R}^n$ (and $F_1, \ldots, F_n, G \in \mathbf{S}^k$), and the dual problem is

$$\begin{array}{ll} \text{maximize} & \mathbf{tr}(GZ) \\ \text{subject to} & \mathbf{tr}(F_i Z) + c_i = 0, \quad i = 1, \ldots, n \\ & Z \succeq 0, \end{array}$$

with variable $Z \in \mathbf{S}^k$.

Suppose that $x^\star$ and $Z^\star$ are primal and dual optimal, respectively, with zero duality gap. The complementary slackness condition is $\mathbf{tr}(F(x^\star)Z^\star) = 0$. Since $F(x^\star) \preceq 0$ and $Z^\star \succeq 0$, we can conclude that $F(x^\star)Z^\star = 0$. Thus, the complementary slackness condition can be expressed as

$$\mathcal{R}(F(x^\star)) \perp \mathcal{R}(Z^\star),$$

i.e., the ranges of the primal and dual matrices are orthogonal.

Let $p^\star(U)$ denote the optimal value of the perturbed SDP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & F(x) = x_1 F_1 + \cdots + x_n F_n + G \preceq U. \end{array}$$

Then we have, for all U, $p^\star(U) \geq p^\star - \mathbf{tr}(Z^\star U)$. If $p^\star(U)$ is differentiable at $U = 0$, then we have

$$\nabla p^\star(0) = -Z^\star.$$

This means that for U small, the optimal value of the perturbed SDP is very close to (the lower bound) $p^\star - \mathbf{tr}(Z^\star U)$.

5.9.4 Theorems of alternatives

We can derive theorems of alternatives for systems of generalized inequalities and equalities

$$f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m, \qquad h_i(x) = 0, \quad i = 1, \ldots, p, \tag{5.96}$$

where $K_i \subseteq \mathbf{R}^{k_i}$ are proper cones. We will also consider systems with strict inequalities,

$$f_i(x) \prec_{K_i} 0, \quad i = 1, \ldots, m, \qquad h_i(x) = 0, \quad i = 1, \ldots, p. \tag{5.97}$$

We assume that $\mathcal{D} = \bigcap_{i=0}^m \mathbf{dom}\, f_i \,\cap\, \bigcap_{i=1}^p \mathbf{dom}\, h_i$ is nonempty.

Weak alternatives

We associate with the systems (5.96) and (5.97) the dual function

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \left(\sum_{i=1}^m \lambda_i^T f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$$

where $\lambda = (\lambda_1, \ldots, \lambda_m)$ with $\lambda_i \in \mathbf{R}^{k_i}$ and $\nu \in \mathbf{R}^p$. In analogy with (5.76), we claim that

$$\lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m, \qquad g(\lambda, \nu) > 0 \tag{5.98}$$

is a weak alternative to the system (5.96). To verify this, suppose there exists an x satisfying (5.96) and (λ, ν) satisfying (5.98). Then we have a contradiction:

$$0 < g(\lambda, \nu) \leq \lambda_1^T f_1(x) + \cdots + \lambda_m^T f_m(x) + \nu_1 h_1(x) + \cdots + \nu_p h_p(x) \leq 0.$$

Therefore at least one of the two systems (5.96) and (5.98) must be infeasible, *i.e.*, the two systems are weak alternatives.

In a similar way, we can prove that (5.97) and the system

$$\lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m, \qquad \lambda \neq 0, \qquad g(\lambda, \nu) \geq 0.$$

form a pair of weak alternatives.

Strong alternatives

We now assume that the functions f_i are K_i-convex, and the functions h_i are affine. We first consider a system with strict inequalities

$$f_i(x) \prec_{K_i} 0, \quad i = 1, \ldots, m, \qquad Ax = b, \tag{5.99}$$

and its alternative

$$\lambda_i \succeq_{K_i^*} 0, \quad i=1,\ldots,m, \qquad \lambda \neq 0, \qquad g(\lambda,\nu) \geq 0. \tag{5.100}$$

We have already seen that (5.99) and (5.100) are weak alternatives. They are also strong alternatives provided the following constraint qualification holds: There exists an $\tilde{x} \in \textbf{relint}\,\mathcal{D}$ with $A\tilde{x} = b$. To prove this, we select a set of vectors $e_i \succ_{K_i} 0$, and consider the problem

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & f_i(x) \preceq_{K_i} se_i, \quad i=1,\ldots,m \\ & Ax = b \end{array} \tag{5.101}$$

with variables x and $s \in \mathbf{R}$. Slater's condition holds since $(\tilde{x},\tilde{s})$ satisfies the strict inequalities $f_i(\tilde{x}) \prec_{K_i} \tilde{s}e_i$ provided $\tilde{s}$ is large enough.

The dual of (5.101) is

$$\begin{array}{ll} \text{maximize} & g(\lambda,\nu) \\ \text{subject to} & \lambda_i \succeq_{K_i^*} 0, \quad i=1,\ldots,m \\ & \sum_{i=1}^m e_i^T\lambda_i = 1 \end{array} \tag{5.102}$$

with variables $\lambda = (\lambda_1,\ldots,\lambda_m)$ and ν.

Now suppose the system (5.99) is infeasible. Then the optimal value of (5.101) is nonnegative. Since Slater's condition is satisfied, we have strong duality and the dual optimum is attained. Therefore there exist $(\tilde{\lambda},\tilde{\nu})$ that satisfy the constraints of (5.102) and $g(\tilde{\lambda},\tilde{\nu}) \geq 0$, *i.e.*, the system (5.100) has a solution.

As we noted in the case of scalar inequalities, existence of an $x \in \textbf{relint}\,\mathcal{D}$ with $Ax = b$ is not sufficient for the system of nonstrict inequalities

$$f_i(x) \preceq_{K_i} 0, \quad i=1,\ldots,m, \qquad Ax = b$$

and its alternative

$$\lambda_i \succeq_{K_i^*} 0, \quad i=1,\ldots,m, \qquad g(\lambda,\nu) > 0$$

to be strong alternatives. An additional condition is required, *e.g.*, that the optimal value of (5.101) is attained.

Example 5.14 *Feasibility of a linear matrix inequality.* The following systems are strong alternatives:

$$F(x) = x_1F_1 + \cdots + x_nF_n + G \prec 0,$$

where $F_i, G \in \mathbf{S}^k$, and

$$Z \succeq 0, \qquad Z \neq 0, \qquad \mathbf{tr}(GZ) \geq 0, \qquad \mathbf{tr}(F_iZ) = 0, \quad i=1,\ldots,n,$$

where $Z \in \mathbf{S}^k$. This follows from the general result, if we take for K the positive semidefinite cone $\mathbf{S}_+^k$, and

$$g(Z) = \inf_x \left(\mathbf{tr}(F(x)Z)\right) = \begin{cases} \mathbf{tr}(GZ) & \mathbf{tr}(F_iZ) = 0, \quad i=1,\ldots,n \\ -\infty & \text{otherwise.} \end{cases}$$

The nonstrict inequality case is slightly more involved, and we need an extra assumption on the matrices F_i to have strong alternatives. One such condition is

$$\sum_{i=1}^{n} v_i F_i \succeq 0 \Longrightarrow \sum_{i=1}^{n} v_i F_i = 0.$$

If this condition holds, the following systems are strong alternatives:

$$F(x) = x_1 F_1 + \cdots + x_n F_n + G \preceq 0$$

and

$$Z \succeq 0, \qquad \mathbf{tr}(GZ) > 0, \qquad \mathbf{tr}(F_i Z) = 0, \quad i = 1, \ldots, n$$

(see exercise 5.44).

Bibliography

Lagrange duality is covered in detail by Luenberger [Lue69, chapter 8], Rockafellar [Roc70, part VI], Whittle [Whi71], Hiriart-Urruty and Lemaréchal [HUL93], and Bertsekas, Nedić, and Ozdaglar [Ber03]. The name is derived from Lagrange's method of multipliers for optimization problems with equality constraints; see Courant and Hilbert [CH53, chapter IV].

The max-min result for matrix games in §5.2.5 predates linear programming duality. It is proved via a theorem of alternatives by von Neuman and Morgenstern [vNM53, page 153]. The strong duality result for linear programming on page 227 is due to von Neumann [vN63] and Gale, Kuhn, and Tucker [GKT51]. Strong duality for the nonconvex quadratic problem (5.32) is a fundamental result in the literature on trust region methods for nonlinear optimization (Nocedal and Wright [NW99, page 78]). It is also related to the S-procedure in control theory, discussed in appendix §B.1. For an extension of the proof of strong duality of §5.3.2 to the refined Slater condition (5.27), see Rockafellar [Roc70, page 277].

Conditions that guarantee the saddle-point property (5.47) can be found in Rockafellar [Roc70, part VII] and Bertsekas, Nedić, and Ozdaglar [Ber03, chapter 2]; see also exercise 5.25.

The KKT conditions are named after Karush (whose unpublished 1939 Master's thesis is summarized in Kuhn [Kuh76]), Kuhn, and Tucker [KT51]. Related optimality conditions were also derived by John [Joh85]. The water-filling algorithm in example 5.2 has applications in information theory and communications (Cover and Thomas [CT91, page 252]).

Farkas' lemma was published by Farkas [Far02]. It is the best known theorem of alternatives for systems of linear inequalities and equalities, but many variants exist; see Mangasarian [Man94, §2.4]. The application of Farkas' lemma to asset pricing (example 5.10) is discussed by Bertsimas and Tsitsiklis [BT97, page 167] and Ross [Ros99].

The extension of Lagrange duality to problems with generalized inequalities appears in Isii [Isi64], Luenberger [Lue69, chapter 8], Berman [Ber73], and Rockafellar [Roc89, page 47]. It is discussed in the context of cone programming in Nesterov and Nemirovski [NN94, §4.2] and Ben-Tal and Nemirovski [BTN01, lecture 2]. Theorems of alternatives for generalized inequalities were studied by Ben-Israel [BI69], Berman and Ben-Israel [BBI71], and Craven and Kohila [CK77]. Bellman and Fan [BF63], Wolkowicz [Wol81], and Lasserre [Las95] give extensions of Farkas' lemma to linear matrix inequalities.

Exercises

Basic definitions

5.1 *A simple example.* Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & x^2+1 \\ \text{subject to} & (x-2)(x-4) \le 0, \end{array}$$

with variable $x \in \mathbf{R}$.

(a) *Analysis of primal problem.* Give the feasible set, the optimal value, and the optimal solution.

(b) *Lagrangian and dual function.* Plot the objective x^2+1 versus x. On the same plot, show the feasible set, optimal point and value, and plot the Lagrangian $L(x,\lambda)$ versus x for a few positive values of λ. Verify the lower bound property ($p^\star \ge \inf_x L(x,\lambda)$ for $\lambda \ge 0$). Derive and sketch the Lagrange dual function g.

(c) *Lagrange dual problem.* State the dual problem, and verify that it is a concave maximization problem. Find the dual optimal value and dual optimal solution $\lambda^\star$. Does strong duality hold?

(d) *Sensitivity analysis.* Let $p^\star(u)$ denote the optimal value of the problem

$$\begin{array}{ll} \text{minimize} & x^2+1 \\ \text{subject to} & (x-2)(x-4) \le u, \end{array}$$

as a function of the parameter u. Plot $p^\star(u)$. Verify that $dp^\star(0)/du = -\lambda^\star$.

5.2 *Weak duality for unbounded and infeasible problems.* The weak duality inequality, $d^\star \le p^\star$, clearly holds when $d^\star = -\infty$ or $p^\star = \infty$. Show that it holds in the other two cases as well: If $p^\star = -\infty$, then we must have $d^\star = -\infty$, and also, if $d^\star = \infty$, then we must have $p^\star = \infty$.

5.3 *Problems with one inequality constraint.* Express the dual problem of

$$\begin{array}{ll} \text{minimize} & c^Tx \\ \text{subject to} & f(x) \le 0, \end{array}$$

with $c \ne 0$, in terms of the conjugate f^*. Explain why the problem you give is convex. We do not assume f is convex.

Examples and applications

5.4 *Interpretation of LP dual via relaxed problems.* Consider the inequality form LP

$$\begin{array}{ll} \text{minimize} & c^Tx \\ \text{subject to} & Ax \preceq b, \end{array}$$

with $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$. In this exercise we develop a simple geometric interpretation of the dual LP (5.22).

Let $w \in \mathbf{R}^m_+$. If x is feasible for the LP, *i.e.*, satisfies $Ax \preceq b$, then it also satisfies the inequality

$$w^TAx \le w^Tb.$$

Geometrically, for any $w \succeq 0$, the halfspace $H_w = \{x \mid w^TAx \le w^Tb\}$ contains the feasible set for the LP. Therefore if we minimize the objective c^Tx over the halfspace H_w we get a lower bound on $p^\star$.

(a) Derive an expression for the minimum value of $c^T x$ over the halfspace H_w (which will depend on the choice of $w \succeq 0$).

(b) Formulate the problem of finding the best such bound, by maximizing the lower bound over $w \succeq 0$.

(c) Relate the results of (a) and (b) to the Lagrange dual of the LP, given by (5.22).

5.5 *Dual of general LP.* Find the dual function of the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Gx \preceq h \\ & Ax = b. \end{array}$$

Give the dual problem, and make the implicit equality constraints explicit.

5.6 *Lower bounds in Chebyshev approximation from least-squares.* Consider the Chebyshev or ℓ_∞-norm approximation problem

$$\text{minimize} \quad \|Ax - b\|_\infty, \tag{5.103}$$

where $A \in \mathbf{R}^{m \times n}$ and $\mathbf{rank}\, A = n$. Let x_{ch} denote an optimal solution (there may be multiple optimal solutions; x_{ch} denotes one of them).

The Chebyshev problem has no closed-form solution, but the corresponding least-squares problem does. Define

$$x_{\text{ls}} = \operatorname{argmin} \|Ax - b\|_2 = (A^T A)^{-1} A^T b.$$

We address the following question. Suppose that for a particular A and b we have computed the least-squares solution x_{ls} (but not x_{ch}). How suboptimal is x_{ls} for the Chebyshev problem? In other words, how much larger is $\|Ax_{\text{ls}} - b\|_\infty$ than $\|Ax_{\text{ch}} - b\|_\infty$?

(a) Prove the lower bound

$$\|Ax_{\text{ls}} - b\|_\infty \le \sqrt{m}\, \|Ax_{\text{ch}} - b\|_\infty,$$

using the fact that for all $z \in \mathbf{R}^m$,

$$\frac{1}{\sqrt{m}} \|z\|_2 \le \|z\|_\infty \le \|z\|_2.$$

(b) In example 5.6 (page 254) we derived a dual for the general norm approximation problem. Applying the results to the ℓ_∞-norm (and its dual norm, the ℓ_1-norm), we can state the following dual for the Chebyshev approximation problem:

$$\begin{array}{ll} \text{maximize} & b^T \nu \\ \text{subject to} & \|\nu\|_1 \le 1 \\ & A^T \nu = 0. \end{array} \tag{5.104}$$

Any feasible ν corresponds to a lower bound $b^T \nu$ on $\|Ax_{\text{ch}} - b\|_\infty$.

Denote the least-squares residual as $r_{\text{ls}} = b - Ax_{\text{ls}}$. Assuming $r_{\text{ls}} \neq 0$, show that

$$\hat{\nu} = -r_{\text{ls}} / \|r_{\text{ls}}\|_1, \qquad \tilde{\nu} = r_{\text{ls}} / \|r_{\text{ls}}\|_1,$$

are both feasible in (5.104). By duality $b^T \hat{\nu}$ and $b^T \tilde{\nu}$ are lower bounds on $\|Ax_{\text{ch}} - b\|_\infty$. Which is the better bound? How do these bounds compare with the bound derived in part (a)?

5.7 *Piecewise-linear minimization.* We consider the convex piecewise-linear minimization problem

$$\text{minimize} \quad \max_{i=1,\ldots,m} (a_i^T x + b_i) \tag{5.105}$$

with variable $x \in \mathbf{R}^n$.

(a) Derive a dual problem, based on the Lagrange dual of the equivalent problem

$$\begin{array}{ll} \text{minimize} & \max_{i=1,\ldots,m} y_i \\ \text{subject to} & a_i^T x + b_i = y_i, \quad i = 1, \ldots, m, \end{array}$$

with variables $x \in \mathbf{R}^n$, $y \in \mathbf{R}^m$.

(b) Formulate the piecewise-linear minimization problem (5.105) as an LP, and form the dual of the LP. Relate the LP dual to the dual obtained in part (a).

(c) Suppose we approximate the objective function in (5.105) by the smooth function

$$f_0(x) = \log\left(\sum_{i=1}^m \exp(a_i^T x + b_i)\right),$$

and solve the unconstrained geometric program

$$\text{minimize} \quad \log\left(\sum_{i=1}^m \exp(a_i^T x + b_i)\right). \tag{5.106}$$

A dual of this problem is given by (5.62). Let $p^\star_{\text{pwl}}$ and $p^\star_{\text{gp}}$ be the optimal values of (5.105) and (5.106), respectively. Show that

$$0 \le p^\star_{\text{gp}} - p^\star_{\text{pwl}} \le \log m.$$

(d) Derive similar bounds for the difference between $p^\star_{\text{pwl}}$ and the optimal value of

$$\text{minimize} \quad (1/\gamma)\log\left(\sum_{i=1}^m \exp(\gamma(a_i^T x + b_i))\right),$$

where $\gamma > 0$ is a parameter. What happens as we increase γ?

5.8 Relate the two dual problems derived in example 5.9 on page 257.

5.9 *Suboptimality of a simple covering ellipsoid.* Recall the problem of determining the minimum volume ellipsoid, centered at the origin, that contains the points $a_1, \ldots, a_m \in \mathbf{R}^n$ (problem (5.14), page 222):

$$\begin{array}{ll} \text{minimize} & f_0(X) = \log\det(X^{-1}) \\ \text{subject to} & a_i^T X a_i \le 1, \quad i = 1, \ldots, m, \end{array}$$

with $\mathbf{dom}\, f_0 = \mathbf{S}^n_{++}$. We assume that the vectors $a_1, \ldots, a_m$ span $\mathbf{R}^n$ (which implies that the problem is bounded below).

(a) Show that the matrix

$$X_{\text{sim}} = \left(\sum_{k=1}^m a_k a_k^T\right)^{-1},$$

is feasible. *Hint.* Show that

$$\begin{bmatrix} \sum_{k=1}^m a_k a_k^T & a_i \\ a_i^T & 1 \end{bmatrix} \succeq 0,$$

and use Schur complements (§A.5.5) to prove that $a_i^T X a_i \le 1$ for $i = 1, \ldots, m$.

(b) Now we establish a bound on how suboptimal the feasible point X_{sim} is, via the dual problem,

$$\begin{array}{ll} \text{maximize} & \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - \mathbf{1}^T\lambda + n \\ \text{subject to} & \lambda \succeq 0, \end{array}$$

with the implicit constraint $\sum_{i=1}^m \lambda_i a_i a_i^T \succ 0$. (This dual is derived on page 222.) To derive a bound, we restrict our attention to dual variables of the form $\lambda = t\mathbf{1}$, where $t > 0$. Find (analytically) the optimal value of t, and evaluate the dual objective at this λ. Use this to prove that the volume of the ellipsoid $\{u \mid u^T X_{\text{sim}} u \le 1\}$ is no more than a factor $(m/n)^{n/2}$ more than the volume of the minimum volume ellipsoid.

5.10 *Optimal experiment design.* The following problems arise in experiment design (see §7.5).

(a) *D-optimal design.*

$$\begin{array}{ll} \text{minimize} & \log\det\left(\sum_{i=1}^p x_i v_i v_i^T\right)^{-1} \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1. \end{array}$$

(b) *A-optimal design.*

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}\left(\sum_{i=1}^p x_i v_i v_i^T\right)^{-1} \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1. \end{array}$$

The domain of both problems is $\{x \mid \sum_{i=1}^p x_i v_i v_i^T \succ 0\}$. The variable is $x \in \mathbf{R}^p$; the vectors $v_1, \ldots, v_p \in \mathbf{R}^n$ are given.

Derive dual problems by first introducing a new variable $X \in \mathbf{S}^n$ and an equality constraint $X = \sum_{i=1}^p x_i v_i v_i^T$, and then applying Lagrange duality. Simplify the dual problems as much as you can.

5.11 Derive a dual problem for

$$\text{minimize} \quad \sum_{i=1}^N \|A_i x + b_i\|_2 + (1/2)\|x - x_0\|_2^2.$$

The problem data are $A_i \in \mathbf{R}^{m_i \times n}$, $b_i \in \mathbf{R}^{m_i}$, and $x_0 \in \mathbf{R}^n$. First introduce new variables $y_i \in \mathbf{R}^{m_i}$ and equality constraints $y_i = A_i x + b_i$.

5.12 *Analytic centering.* Derive a dual problem for

$$\text{minimize} \quad -\sum_{i=1}^m \log(b_i - a_i^T x)$$

with domain $\{x \mid a_i^T x < b_i,\ i = 1, \ldots, m\}$. First introduce new variables y_i and equality constraints $y_i = b_i - a_i^T x$.

(The solution of this problem is called the *analytic center* of the linear inequalities $a_i^T x \leq b_i$, $i = 1, \ldots, m$. Analytic centers have geometric applications (see §8.5.3), and play an important role in barrier methods (see chapter 11).)

5.13 *Lagrangian relaxation of Boolean LP.* A *Boolean linear program* is an optimization problem of the form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i \in \{0, 1\}, \quad i = 1, \ldots, n, \end{array}$$

and is, in general, very difficult to solve. In exercise 4.15 we studied the LP relaxation of this problem,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & 0 \leq x_i \leq 1, \quad i = 1, \ldots, n, \end{array} \tag{5.107}$$

which is far easier to solve, and gives a lower bound on the optimal value of the Boolean LP. In this problem we derive another lower bound for the Boolean LP, and work out the relation between the two lower bounds.

(a) *Lagrangian relaxation.* The Boolean LP can be reformulated as the problem

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \\ & x_i(1 - x_i) = 0, \quad i = 1, \ldots, n, \end{array}$$

which has quadratic equality constraints. Find the Lagrange dual of this problem. The optimal value of the dual problem (which is convex) gives a lower bound on the optimal value of the Boolean LP. This method of finding a lower bound on the optimal value is called *Lagrangian relaxation.*

(b) Show that the lower bound obtained via Lagrangian relaxation, and via the LP relaxation (5.107), are the same. *Hint.* Derive the dual of the LP relaxation (5.107).

5.14 *A penalty method for equality constraints.* We consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax = b, \end{array} \tag{5.108}$$

where $f_0 : \mathbf{R}^n \to \mathbf{R}$ is convex and differentiable, and $A \in \mathbf{R}^{m \times n}$ with $\mathbf{rank}\, A = m$.

In a *quadratic penalty method*, we form an auxiliary function

$$\phi(x) = f_0(x) + \alpha \|Ax - b\|_2^2,$$

where $\alpha > 0$ is a parameter. This auxiliary function consists of the objective plus the *penalty term* $\alpha\|Ax-b\|_2^2$. The idea is that a minimizer of the auxiliary function, $\tilde{x}$, should be an approximate solution of the original problem. Intuition suggests that the larger the penalty weight α, the better the approximation $\tilde{x}$ to a solution of the original problem.

Suppose $\tilde{x}$ is a minimizer of ϕ. Show how to find, from $\tilde{x}$, a dual feasible point for (5.108). Find the corresponding lower bound on the optimal value of (5.108).

5.15 Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m, \end{array} \tag{5.109}$$

where the functions $f_i : \mathbf{R}^n \to \mathbf{R}$ are differentiable and convex. Let $h_1, \ldots, h_m : \mathbf{R} \to \mathbf{R}$ be increasing differentiable convex functions. Show that

$$\phi(x) = f_0(x) + \sum_{i=1}^m h_i(f_i(x))$$

is convex. Suppose $\tilde{x}$ minimizes ϕ. Show how to find from $\tilde{x}$ a feasible point for the dual of (5.109). Find the corresponding lower bound on the optimal value of (5.109).

5.16 *An exact penalty method for inequality constraints.* Consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m, \end{array} \tag{5.110}$$

where the functions $f_i : \mathbf{R}^n \to \mathbf{R}$ are differentiable and convex. In an exact penalty method, we solve the auxiliary problem

$$\text{minimize} \quad \phi(x) = f_0(x) + \alpha \max_{i=1,\ldots,m} \max\{0, f_i(x)\}, \tag{5.111}$$

where $\alpha > 0$ is a parameter. The second term in ϕ penalizes deviations of x from feasibility. The method is called an *exact* penalty method if for sufficiently large α, solutions of the auxiliary problem (5.111) also solve the original problem (5.110).

(a) Show that ϕ is convex.

(b) The auxiliary problem can be expressed as

$$\begin{array}{ll} \text{minimize} & f_0(x) + \alpha y \\ \text{subject to} & f_i(x) \leq y, \quad i = 1, \ldots, m \\ & 0 \leq y \end{array}$$

where the variables are x and $y \in \mathbf{R}$. Find the Lagrange dual of this problem, and express it in terms of the Lagrange dual function g of (5.110).

(c) Use the result in (b) to prove the following property. Suppose $\lambda^\star$ is an optimal solution of the Lagrange dual of (5.110), and that strong duality holds. If $\alpha > \mathbf{1}^T\lambda^\star$, then any solution of the auxiliary problem (5.111) is also an optimal solution of (5.110).

5.17 *Robust linear programming with polyhedral uncertainty.* Consider the robust LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \sup_{a \in \mathcal{P}_i} a^T x \leq b_i, \quad i = 1, \ldots, m, \end{array}$$

with variable $x \in \mathbf{R}^n$, where $\mathcal{P}_i = \{a \mid C_i a \preceq d_i\}$. The problem data are $c \in \mathbf{R}^n$, $C_i \in \mathbf{R}^{m_i \times n}$, $d_i \in \mathbf{R}^{m_i}$, and $b \in \mathbf{R}^m$. We assume the polyhedra $\mathcal{P}_i$ are nonempty.

Show that this problem is equivalent to the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & d_i^T z_i \leq b_i, \quad i = 1, \ldots, m \\ & C_i^T z_i = x, \quad i = 1, \ldots, m \\ & z_i \succeq 0, \quad i = 1, \ldots, m \end{array}$$

with variables $x \in \mathbf{R}^n$ and $z_i \in \mathbf{R}^{m_i}$, $i = 1, \ldots, m$. *Hint.* Find the dual of the problem of maximizing $a_i^T x$ over $a_i \in \mathcal{P}_i$ (with variable a_i).

5.18 *Separating hyperplane between two polyhedra.* Formulate the following problem as an LP or an LP feasibility problem. Find a separating hyperplane that strictly separates two polyhedra

$$\mathcal{P}_1 = \{x \mid Ax \preceq b\}, \qquad \mathcal{P}_2 = \{x \mid Cx \preceq d\},$$

i.e., find a vector $a \in \mathbf{R}^n$ and a scalar γ such that

$$a^T x > \gamma \text{ for } x \in \mathcal{P}_1, \qquad a^T x < \gamma \text{ for } x \in \mathcal{P}_2.$$

You can assume that $\mathcal{P}_1$ and $\mathcal{P}_2$ do not intersect.

Hint. The vector a and scalar γ must satisfy

$$\inf_{x \in \mathcal{P}_1} a^T x > \gamma > \sup_{x \in \mathcal{P}_2} a^T x.$$

Use LP duality to simplify the infimum and supremum in these conditions.

5.19 *The sum of the largest elements of a vector.* Define $f : \mathbf{R}^n \to \mathbf{R}$ as

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

where r is an integer between 1 and n, and $x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[r]}$ are the components of x sorted in decreasing order. In other words, $f(x)$ is the sum of the r largest elements of x. In this problem we study the constraint

$$f(x) \leq \alpha.$$

As we have seen in chapter 3, page 80, this is a convex constraint, and equivalent to a set of $n!/(r!(n-r)!)$ linear inequalities

$$x_{i_1} + \cdots + x_{i_r} \leq \alpha, \quad 1 \leq i_1 < i_2 < \cdots < i_r \leq n.$$

The purpose of this problem is to derive a more compact representation.

(a) Given a vector $x \in \mathbf{R}^n$, show that $f(x)$ is equal to the optimal value of the LP

$$\begin{array}{ll} \text{maximize} & x^T y \\ \text{subject to} & 0 \preceq y \preceq \mathbf{1} \\ & \mathbf{1}^T y = r \end{array}$$

with $y \in \mathbf{R}^n$ as variable.

(b) Derive the dual of the LP in part (a). Show that it can be written as

$$\begin{array}{ll} \text{minimize} & rt + \mathbf{1}^T u \\ \text{subject to} & t\mathbf{1} + u \succeq x \\ & u \succeq 0, \end{array}$$

where the variables are $t \in \mathbf{R}$, $u \in \mathbf{R}^n$. By duality this LP has the same optimal value as the LP in (a), *i.e.*, $f(x)$. We therefore have the following result: x satisfies $f(x) \leq \alpha$ if and only if there exist $t \in \mathbf{R}$, $u \in \mathbf{R}^n$ such that

$$rt + \mathbf{1}^T u \leq \alpha, \qquad t\mathbf{1} + u \succeq x, \qquad u \succeq 0.$$

These conditions form a set of $2n+1$ linear inequalities in the $2n+1$ variables x, u, t.

(c) As an application, we consider an extension of the classical Markowitz portfolio optimization problem

$$\begin{array}{ll} \text{minimize} & x^T \Sigma x \\ \text{subject to} & \overline{p}^T x \geq r_{\min} \\ & \mathbf{1}^T x = 1, \quad x \succeq 0 \end{array}$$

discussed in chapter 4, page 155. The variable is the portfolio $x \in \mathbf{R}^n$; $\overline{p}$ and Σ are the mean and covariance matrix of the price change vector p.

Suppose we add a *diversification constraint*, requiring that no more than 80% of the total budget can be invested in any 10% of the assets. This constraint can be expressed as

$$\sum_{i=1}^{\lfloor 0.1n \rfloor} x_{[i]} \leq 0.8.$$

Formulate the portfolio optimization problem with diversification constraint as a QP.

5.20 *Dual of channel capacity problem.* Derive a dual for the problem

$$\begin{array}{ll} \text{minimize} & -c^T x + \sum_{i=1}^m y_i \log y_i \\ \text{subject to} & Px = y \\ & x \succeq 0, \quad \mathbf{1}^T x = 1, \end{array}$$

where $P \in \mathbf{R}^{m \times n}$ has nonnegative elements, and its columns add up to one (*i.e.*, $P^T \mathbf{1} = \mathbf{1}$). The variables are $x \in \mathbf{R}^n$, $y \in \mathbf{R}^m$. (For $c_j = \sum_{i=1}^m p_{ij} \log p_{ij}$, the optimal value is, up to a factor $\log 2$, the negative of the capacity of a discrete memoryless channel with channel transition probability matrix P; see exercise 4.57.)

Simplify the dual problem as much as possible.

Strong duality and Slater's condition

5.21 *A convex problem in which strong duality fails.* Consider the optimization problem

$$\begin{array}{ll} \text{minimize} & e^{-x} \\ \text{subject to} & x^2/y \leq 0 \end{array}$$

with variables x and y, and domain $\mathcal{D} = \{(x, y) \mid y > 0\}$.

(a) Verify that this is a convex optimization problem. Find the optimal value.

(b) Give the Lagrange dual problem, and find the optimal solution $\lambda^\star$ and optimal value $d^\star$ of the dual problem. What is the optimal duality gap?

(c) Does Slater's condition hold for this problem?

(d) What is the optimal value $p^\star(u)$ of the perturbed problem

$$\begin{array}{ll} \text{minimize} & e^{-x} \\ \text{subject to} & x^2/y \leq u \end{array}$$

as a function of u? Verify that the global sensitivity inequality

$$p^\star(u) \geq p^\star(0) - \lambda^\star u$$

does not hold.

5.22 *Geometric interpretation of duality.* For each of the following optimization problems, draw a sketch of the sets

$$\begin{array}{rcl} \mathcal{G} & = & \{(u, t) \mid \exists x \in \mathcal{D},\ f_0(x) = t,\ f_1(x) = u\}, \\ \mathcal{A} & = & \{(u, t) \mid \exists x \in \mathcal{D},\ f_0(x) \leq t,\ f_1(x) \leq u\}, \end{array}$$

give the dual problem, and solve the primal and dual problems. Is the problem convex? Is Slater's condition satisfied? Does strong duality hold?

The domain of the problem is $\mathbf{R}$ unless otherwise stated.

(a) Minimize x subject to $x^2 \leq 1$.

(b) Minimize x subject to $x^2 \leq 0$.

(c) Minimize x subject to $|x| \leq 0$.

(d) Minimize x subject to $f_1(x) \leq 0$ where

$$f_1(x) = \begin{cases} -x + 2 & x \geq 1 \\ x & -1 \leq x \leq 1 \\ -x - 2 & x \leq -1. \end{cases}$$

(e) Minimize x^3 subject to $-x + 1 \leq 0$.

(f) Minimize x^3 subject to $-x + 1 \leq 0$ with domain $\mathcal{D} = \mathbf{R}_+$.

5.23 *Strong duality in linear programming.* We prove that strong duality holds for the LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq b \end{array}$$

and its dual

$$\begin{array}{ll} \text{maximize} & -b^T z \\ \text{subject to} & A^T z + c = 0, \quad z \succeq 0, \end{array}$$

provided at least one of the problems is feasible. In other words, the only possible exception to strong duality occurs when $p^\star = \infty$ and $d^\star = -\infty$.

(a) Suppose $p^\star$ is finite and $x^\star$ is an optimal solution. (If finite, the optimal value of an LP is attained.) Let $I \subseteq \{1, 2, \ldots, m\}$ be the set of active constraints at $x^\star$:

$$a_i^T x^\star = b_i, \quad i \in I, \qquad a_i^T x^\star < b_i, \quad i \notin I.$$

Show that there exists a $z \in \mathbf{R}^m$ that satisfies

$$z_i \geq 0, \quad i \in I, \qquad z_i = 0, \quad i \notin I, \qquad \sum_{i \in I} z_i a_i + c = 0.$$

Show that z is dual optimal with objective value $c^T x^\star$.

Hint. Assume there exists no such z, *i.e.*, $-c \notin \{\sum_{i \in I} z_i a_i \mid z_i \geq 0\}$. Reduce this to a contradiction by applying the strict separating hyperplane theorem of example 2.20, page 49. Alternatively, you can use Farkas' lemma (see §5.8.3).

(b) Suppose $p^\star = \infty$ and the dual problem is feasible. Show that $d^\star = \infty$. *Hint.* Show that there exists a nonzero $v \in \mathbf{R}^m$ such that $A^T v = 0$, $v \succeq 0$, $b^T v < 0$. If the dual is feasible, it is unbounded in the direction v.

(c) Consider the example

$$\begin{array}{ll} \text{minimize} & x \\ \text{subject to} & \begin{bmatrix} 0 \\ 1 \end{bmatrix} x \preceq \begin{bmatrix} -1 \\ 1 \end{bmatrix}. \end{array}$$

Formulate the dual LP, and solve the primal and dual problems. Show that $p^\star = \infty$ and $d^\star = -\infty$.

5.24 *Weak max-min inequality.* Show that the weak max-min inequality

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) \leq \inf_{w \in W} \sup_{z \in Z} f(w, z)$$

always holds, with no assumptions on $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$, $W \subseteq \mathbf{R}^n$, or $Z \subseteq \mathbf{R}^m$.

5.25 [BL00, page 95] *Convex-concave functions and the saddle-point property.* We derive conditions under which the saddle-point property

$$\sup_{z \in Z} \inf_{w \in W} f(w, z) = \inf_{w \in W} \sup_{z \in Z} f(w, z) \tag{5.112}$$

holds, where $f : \mathbf{R}^n \times \mathbf{R}^m \to \mathbf{R}$, $W \times Z \subseteq \mathbf{dom}\, f$, and W and Z are nonempty. We will assume that the function

$$g_z(w) = \begin{cases} f(w, z) & w \in W \\ \infty & \text{otherwise} \end{cases}$$

is closed and convex for all $z \in Z$, and the function

$$h_w(z) = \begin{cases} -f(w, z) & z \in Z \\ \infty & \text{otherwise} \end{cases}$$

is closed and convex for all $w \in W$.

(a) The righthand side of (5.112) can be expressed as $p(0)$, where

$$p(u) = \inf_{w \in W} \sup_{z \in Z} (f(w, z) + u^T z).$$

Show that p is a convex function.

(b) Show that the conjugate of p is given by

$$p^*(v) = \begin{cases} -\inf_{w \in W} f(w,v) & v \in Z \\ \infty & \text{otherwise.} \end{cases}$$

(c) Show that the conjugate of p^* is given by

$$p^{**}(u) = \sup_{z \in Z} \inf_{w \in W} (f(w,z) + u^T z).$$

Combining this with (a), we can express the max-min equality (5.112) as $p^{**}(0) = p(0)$.

(d) From exercises 3.28 and 3.39 (d), we know that $p^{**}(0) = p(0)$ if $0 \in \mathbf{int\,dom}\, p$. Conclude that this is the case if W and Z are bounded.

(e) As another consequence of exercises 3.28 and 3.39, we have $p^{**}(0) = p(0)$ if $0 \in \mathbf{dom}\, p$ and p is closed. Show that p is closed if the sublevel sets of g_z are bounded.

Optimality conditions

5.26 Consider the QCQP

$$\begin{array}{ll} \text{minimize} & x_1^2 + x_2^2 \\ \text{subject to} & (x_1-1)^2 + (x_2-1)^2 \le 1 \\ & (x_1-1)^2 + (x_2+1)^2 \le 1 \end{array}$$

with variable $x \in \mathbf{R}^2$.

(a) Sketch the feasible set and level sets of the objective. Find the optimal point $x^\star$ and optimal value $p^\star$.

(b) Give the KKT conditions. Do there exist Lagrange multipliers $\lambda_1^\star$ and $\lambda_2^\star$ that prove that $x^\star$ is optimal?

(c) Derive and solve the Lagrange dual problem. Does strong duality hold?

5.27 *Equality constrained least-squares.* Consider the equality constrained least-squares problem

$$\begin{array}{ll} \text{minimize} & \|Ax - b\|_2^2 \\ \text{subject to} & Gx = h \end{array}$$

where $A \in \mathbf{R}^{m \times n}$ with $\mathbf{rank}\, A = n$, and $G \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, G = p$.
Give the KKT conditions, and derive expressions for the primal solution $x^\star$ and the dual solution $\nu^\star$.

5.28 Prove (without using any linear programming code) that the optimal solution of the LP

$$\begin{array}{ll} \text{minimize} & 47x_1 + 93x_2 + 17x_3 - 93x_4 \\ \text{subject to} & \begin{bmatrix} -1 & -6 & 1 & 3 \\ -1 & -2 & 7 & 1 \\ 0 & 3 & -10 & -1 \\ -6 & -11 & -2 & 12 \\ 1 & 6 & -1 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \preceq \begin{bmatrix} -3 \\ 5 \\ -8 \\ -7 \\ 4 \end{bmatrix} \end{array}$$

is unique, and given by $x^\star = (1,1,1,1)$.

5.29 The problem

$$\begin{array}{ll} \text{minimize} & -3x_1^2 + x_2^2 + 2x_3^2 + 2(x_1 + x_2 + x_3) \\ \text{subject to} & x_1^2 + x_2^2 + x_3^2 = 1, \end{array}$$

is a special case of (5.32), so strong duality holds even though the problem is not convex. Derive the KKT conditions. Find all solutions x, ν that satisfy the KKT conditions. Which pair corresponds to the optimum?

5.30 Derive the KKT conditions for the problem

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}\, X - \log\det X \\ \text{subject to} & Xs = y, \end{array}$$

with variable $X \in \mathbf{S}^n$ and domain $\mathbf{S}^n_{++}$. $y \in \mathbf{R}^n$ and $s \in \mathbf{R}^n$ are given, with $s^T y = 1$. Verify that the optimal solution is given by

$$X^\star = I + yy^T - \frac{1}{s^T s} ss^T.$$

5.31 *Supporting hyperplane interpretation of KKT conditions.* Consider a convex problem with no equality constraints,

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \le 0, \quad i = 1, \dots, m. \end{array}$$

Assume that $x^\star \in \mathbf{R}^n$ and $\lambda^\star \in \mathbf{R}^m$ satisfy the KKT conditions

$$\begin{array}{rcll} f_i(x^\star) & \le & 0, & i = 1, \dots, m \\ \lambda_i^\star & \ge & 0, & i = 1, \dots, m \\ \lambda_i^\star f_i(x^\star) & = & 0, & i = 1, \dots, m \\ \nabla f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) & = & 0. & \end{array}$$

Show that

$$\nabla f_0(x^\star)^T (x - x^\star) \ge 0$$

for all feasible x. In other words the KKT conditions imply the simple optimality criterion of §4.2.3.

Perturbation and sensitivity analysis

5.32 *Optimal value of perturbed problem.* Let $f_0, f_1, \dots, f_m : \mathbf{R}^n \to \mathbf{R}$ be convex. Show that the function

$$p^\star(u, v) = \inf\{f_0(x) \mid \exists x \in \mathcal{D},\ f_i(x) \le u_i,\ i = 1, \dots, m,\ Ax - b = v\}$$

is convex. This function is the optimal cost of the perturbed problem, as a function of the perturbations u and v (see §5.6.1).

5.33 *Parametrized ℓ_1-norm approximation.* Consider the ℓ_1-norm minimization problem

$$\text{minimize} \quad \|Ax + b + \epsilon d\|_1$$

with variable $x \in \mathbf{R}^3$, and

$$A = \begin{bmatrix} -2 & 7 & 1 \\ -5 & -1 & 3 \\ -7 & 3 & -5 \\ -1 & 4 & -4 \\ 1 & 5 & 5 \\ 2 & -5 & -1 \end{bmatrix}, \qquad b = \begin{bmatrix} -4 \\ 3 \\ 9 \\ 0 \\ -11 \\ 5 \end{bmatrix}, \qquad d = \begin{bmatrix} -10 \\ -13 \\ -27 \\ -10 \\ -7 \\ 14 \end{bmatrix}.$$

We denote by $p^\star(\epsilon)$ the optimal value as a function of ϵ.

(a) Suppose $\epsilon = 0$. Prove that $x^\star = \mathbf{1}$ is optimal. Are there any other optimal points?

(b) Show that $p^\star(\epsilon)$ is affine on an interval that includes $\epsilon = 0$.

5.34 Consider the pair of primal and dual LPs

$$\begin{array}{ll} \text{minimize} & (c+\epsilon d)^T x \\ \text{subject to} & Ax \preceq b + \epsilon f \end{array}$$

and

$$\begin{array}{ll} \text{maximize} & -(b+\epsilon f)^T z \\ \text{subject to} & A^T z + c + \epsilon d = 0 \\ & z \succeq 0 \end{array}$$

where

$$A = \begin{bmatrix} -4 & 12 & -2 & 1 \\ -17 & 12 & 7 & 11 \\ 1 & 0 & -6 & 1 \\ 3 & 3 & 22 & -1 \\ -11 & 2 & -1 & -8 \end{bmatrix}, \qquad b = \begin{bmatrix} 8 \\ 13 \\ -4 \\ 27 \\ -18 \end{bmatrix}, \qquad f = \begin{bmatrix} 6 \\ 15 \\ -13 \\ 48 \\ 8 \end{bmatrix},$$

$c = (49, -34, -50, -5)$, $d = (3, 8, 21, 25)$, and ϵ is a parameter.

(a) Prove that $x^\star = (1,1,1,1)$ is optimal when $\epsilon = 0$, by constructing a dual optimal point $z^\star$ that has the same objective value as $x^\star$. Are there any other primal or dual optimal solutions?

(b) Give an explicit expression for the optimal value $p^\star(\epsilon)$ as a function of ϵ on an interval that contains $\epsilon = 0$. Specify the interval on which your expression is valid. Also give explicit expressions for the primal solution $x^\star(\epsilon)$ and the dual solution $z^\star(\epsilon)$ as a function of ϵ, on the same interval.

Hint. First calculate $x^\star(\epsilon)$ and $z^\star(\epsilon)$, assuming that the primal and dual constraints that are active at the optimum for $\epsilon = 0$, remain active at the optimum for values of ϵ around 0. Then verify that this assumption is correct.

5.35 *Sensitivity analysis for GPs.* Consider a GP

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \le 1, & i = 1, \ldots, m \\ & h_i(x) = 1, & i = 1, \ldots, p, \end{array}$$

where $f_0, \ldots, f_m$ are posynomials, $h_1, \ldots, h_p$ are monomials, and the domain of the problem is $\mathbf{R}^n_{++}$. We define the perturbed GP as

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \le e^{u_i}, & i = 1, \ldots, m \\ & h_i(x) = e^{v_i}, & i = 1, \ldots, p, \end{array}$$

and we denote the optimal value of the perturbed GP as $p^\star(u, v)$. We can think of u_i and v_i as relative, or fractional, perturbations of the constraints. For example, $u_1 = -0.01$ corresponds to tightening the first inequality constraint by (approximately) 1%.

Let $\lambda^\star$ and $\nu^\star$ be optimal dual variables for the convex form GP

$$\begin{array}{lll} \text{minimize} & \log f_0(y) & \\ \text{subject to} & \log f_i(y) \le 0, & i = 1, \ldots, m \\ & \log h_i(y) = 0, & i = 1, \ldots, p, \end{array}$$

with variables $y_i = \log x_i$. Assuming that $p^\star(u, v)$ is differentiable at $u = 0$, $v = 0$, relate $\lambda^\star$ and $\nu^\star$ to the derivatives of $p^\star(u, v)$ at $u = 0$, $v = 0$. Justify the statement "Relaxing the ith constraint by α percent will give an improvement in the objective of around $\alpha\lambda_i^\star$ percent, for α small."

Theorems of alternatives

5.36 *Alternatives for linear equalities.* Consider the linear equations $Ax = b$, where $A \in \mathbf{R}^{m\times n}$. From linear algebra we know that this equation has a solution if and only $b \in \mathcal{R}(A)$, which occurs if and only if $b \perp \mathcal{N}(A^T)$. In other words, $Ax = b$ has a solution if and only if there exists no $y \in \mathbf{R}^m$ such that $A^T y = 0$ and $b^T y \neq 0$.

Derive this result from the theorems of alternatives in §5.8.2.

5.37 [BT97] *Existence of equilibrium distribution in finite state Markov chain.* Let $P \in \mathbf{R}^{n\times n}$ be a matrix that satisfies

$$p_{ij} \geq 0, \quad i,j = 1,\ldots,n, \qquad P^T\mathbf{1} = \mathbf{1},$$

i.e., the coefficients are nonnegative and the columns sum to one. Use Farkas' lemma to prove there exists a $y \in \mathbf{R}^n$ such that

$$Py = y, \qquad y \succeq 0, \qquad \mathbf{1}^T y = 1.$$

(We can interpret y as an equilibrium distribution of the Markov chain with n states and transition probability matrix P.)

5.38 [BT97] *Option pricing.* We apply the results of example 5.10, page 263, to a simple problem with three assets: a riskless asset with fixed return $r > 1$ over the investment period of interest (for example, a bond), a stock, and an option on the stock. The option gives us the right to purchase the stock at the end of the period, for a predetermined price K.

We consider two scenarios. In the first scenario, the price of the stock goes up from S at the beginning of the period, to Su at the end of the period, where $u > r$. In this scenario, we exercise the option only if $Su > K$, in which case we make a profit of $Su - K$. Otherwise, we do not exercise the option, and make zero profit. The value of the option at the end of the period, in the first scenario, is therefore $\max\{0, Su - K\}$.

In the second scenario, the price of the stock goes down from S to Sd, where $d < 1$. The value at the end of the period is $\max\{0, Sd - K\}$.

In the notation of example 5.10,

$$V = \begin{bmatrix} r & uS & \max\{0, Su - K\} \\ r & dS & \max\{0, Sd - K\} \end{bmatrix}, \qquad p_1 = 1, \qquad p_2 = S, \qquad p_3 = C,$$

where C is the price of the option.

Show that for given r, S, K, u, d, the option price C is uniquely determined by the no-arbitrage condition. In other words, the market for the option is complete.

Generalized inequalities

5.39 *SDP relaxations of two-way partitioning problem.* We consider the two-way partitioning problem (5.7), described on page 219,

$$\begin{array}{ll} \text{minimize} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1,\ldots,n, \end{array} \tag{5.113}$$

with variable $x \in \mathbf{R}^n$. The Lagrange dual of this (nonconvex) problem is given by the SDP

$$\begin{array}{ll} \text{maximize} & -\mathbf{1}^T \nu \\ \text{subject to} & W + \mathbf{diag}(\nu) \succeq 0 \end{array} \tag{5.114}$$

with variable $\nu \in \mathbf{R}^n$. The optimal value of this SDP gives a lower bound on the optimal value of the partitioning problem (5.113). In this exercise we derive another SDP that gives a lower bound on the optimal value of the two-way partitioning problem, and explore the connection between the two SDPs.

(a) *Two-way partitioning problem in matrix form.* Show that the two-way partitioning problem can be cast as

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(WX) \\ \text{subject to} & X \succeq 0, \quad \mathbf{rank}\, X = 1 \\ & X_{ii} = 1, \quad i = 1, \ldots, n, \end{array}$$

with variable $X \in \mathbf{S}^n$. *Hint.* Show that if X is feasible, then it has the form $X = xx^T$, where $x \in \mathbf{R}^n$ satisfies $x_i \in \{-1, 1\}$ (and vice versa).

(b) *SDP relaxation of two-way partitioning problem.* Using the formulation in part (a), we can form the relaxation

$$\begin{array}{ll} \text{minimize} & \mathbf{tr}(WX) \\ \text{subject to} & X \succeq 0 \\ & X_{ii} = 1, \quad i = 1, \ldots, n, \end{array} \tag{5.115}$$

with variable $X \in \mathbf{S}^n$. This problem is an SDP, and therefore can be solved efficiently. Explain why its optimal value gives a lower bound on the optimal value of the two-way partitioning problem (5.113). What can you say if an optimal point $X^\star$ for this SDP has rank one?

(c) We now have two SDPs that give a lower bound on the optimal value of the two-way partitioning problem (5.113): the SDP relaxation (5.115) found in part (b), and the Lagrange dual of the two-way partitioning problem, given in (5.114). What is the relation between the two SDPs? What can you say about the lower bounds found by them? *Hint:* Relate the two SDPs via duality.

5.40 *E-optimal experiment design.* A variation on the two optimal experiment design problems of exercise 5.10 is the *E-optimal design* problem

$$\begin{array}{ll} \text{minimize} & \lambda_{\max} \left(\sum_{i=1}^p x_i v_i v_i^T \right)^{-1} \\ \text{subject to} & x \succeq 0, \quad \mathbf{1}^T x = 1. \end{array}$$

(See also §7.5.) Derive a dual for this problem, by first reformulating it as

$$\begin{array}{ll} \text{minimize} & 1/t \\ \text{subject to} & \sum_{i=1}^p x_i v_i v_i^T \succeq tI \\ & x \succeq 0, \quad \mathbf{1}^T x = 1, \end{array}$$

with variables $t \in \mathbf{R}$, $x \in \mathbf{R}^p$ and domain $\mathbf{R}_{++} \times \mathbf{R}^p$, and applying Lagrange duality. Simplify the dual problem as much as you can.

5.41 *Dual of fastest mixing Markov chain problem.* On page 174, we encountered the SDP

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & -tI \preceq P - (1/n)\mathbf{1}\mathbf{1}^T \preceq tI \\ & P\mathbf{1} = \mathbf{1} \\ & P_{ij} \geq 0, \quad i, j = 1, \ldots, n \\ & P_{ij} = 0 \text{ for } (i, j) \notin \mathcal{E}, \end{array}$$

with variables $t \in \mathbf{R}$, $P \in \mathbf{S}^n$.

Show that the dual of this problem can be expressed as

$$\begin{array}{ll} \text{maximize} & \mathbf{1}^T z - (1/n)\mathbf{1}^T Y \mathbf{1} \\ \text{subject to} & \|Y\|_{2*} \leq 1 \\ & (z_i + z_j) \leq Y_{ij} \text{ for } (i, j) \in \mathcal{E} \end{array}$$

with variables $z \in \mathbf{R}^n$ and $Y \in \mathbf{S}^n$. The norm $\|\cdot\|_{2*}$ is the dual of the spectral norm on $\mathbf{S}^n$: $\|Y\|_{2*} = \sum_{i=1}^n |\lambda_i(Y)|$, the sum of the absolute values of the eigenvalues of Y. (See §A.1.6, page 637.)

5.42 *Lagrange dual of conic form problem in inequality form.* Find the Lagrange dual problem of the conic form problem in inequality form

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax \preceq_K b \end{array}$$

where $A \in \mathbf{R}^{m\times n}$, $b \in \mathbf{R}^m$, and K is a proper cone in $\mathbf{R}^m$. Make any implicit equality constraints explicit.

5.43 *Dual of SOCP.* Show that the dual of the SOCP

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \dots, m, \end{array}$$

with variables $x \in \mathbf{R}^n$, can be expressed as

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^m (b_i^T u_i - d_i v_i) \\ \text{subject to} & \sum_{i=1}^m (A_i^T u_i - c_i v_i) + f = 0 \\ & \|u_i\|_2 \leq v_i, \quad i = 1, \dots, m, \end{array}$$

with variables $u_i \in \mathbf{R}^{n_i}$, $v_i \in \mathbf{R}$, $i = 1, \dots, m$. The problem data are $f \in \mathbf{R}^n$, $A_i \in \mathbf{R}^{n_i \times n}$, $b_i \in \mathbf{R}^{n_i}$, $c_i \in \mathbf{R}$ and $d_i \in \mathbf{R}$, $i = 1, \dots, m$.

Derive the dual in the following two ways.

(a) Introduce new variables $y_i \in \mathbf{R}^{n_i}$ and $t_i \in \mathbf{R}$ and equalities $y_i = A_i x + b_i$, $t_i = c_i^T x + d_i$, and derive the Lagrange dual.

(b) Start from the conic formulation of the SOCP and use the conic dual. Use the fact that the second-order cone is self-dual.

5.44 *Strong alternatives for nonstrict LMIs.* In example 5.14, page 270, we mentioned that the system

$$Z \succeq 0, \qquad \mathbf{tr}(GZ) > 0, \qquad \mathbf{tr}(F_i Z) = 0, \quad i = 1, \dots, n, \tag{5.116}$$

is a strong alternative for the nonstrict LMI

$$F(x) = x_1 F_1 + \cdots + x_n F_n + G \preceq 0, \tag{5.117}$$

if the matrices F_i satisfy

$$\sum_{i=1}^n v_i F_i \succeq 0 \implies \sum_{i=1}^n v_i F_i = 0. \tag{5.118}$$

In this exercise we prove this result, and give an example to illustrate that the systems are not always strong alternatives.

(a) Suppose (5.118) holds, and that the optimal value of the auxiliary SDP

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & F(x) \preceq sI \end{array}$$

is positive. Show that the optimal value is attained. If follows from the discussion in §5.9.4 that the systems (5.117) and (5.116) are strong alternatives.

Hint. The proof simplifies if you assume, without loss of generality, that the matrices $F_1, \dots, F_n$ are independent, so (5.118) may be replaced by $\sum_{i=1}^n v_i F_i \succeq 0 \Rightarrow v = 0$.

(b) Take $n = 1$, and

$$G = \left[\begin{array}{cc} 0 & 1 \\ 1 & 0 \end{array}\right], \qquad F_1 = \left[\begin{array}{cc} 0 & 0 \\ 0 & 1 \end{array}\right].$$

Show that (5.117) and (5.116) are both infeasible.