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, $\{1, \ldots, n\}$: A feasible x corresponds to the partition

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

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 $p^\star$

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

Entropy maximization

Consider the entropy maximization problem

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

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 $X \in \mathbf{S}^n$,

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

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).

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.

Lagrange dual of QCQP

We consider the QCQP

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

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

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

Minimum volume covering ellipsoid

We consider the problem (5.14):

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

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

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

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 $i = 1, \dots, m$. This is always satisfied, so strong duality always obtains between (5.14) and the dual problem (5.31).

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 $x_i^\star$.

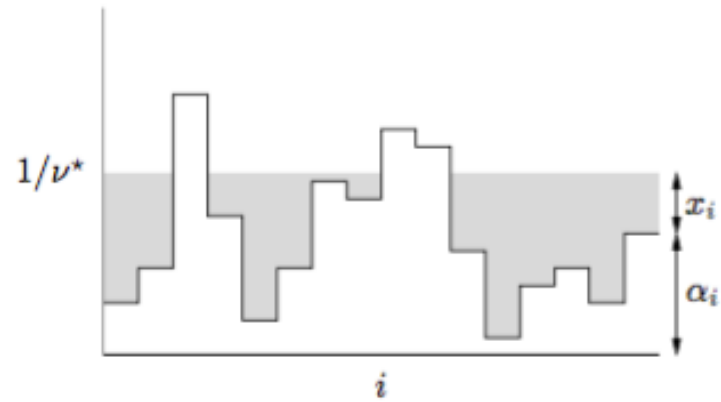

Figure 5.7 Illustration of water-filling algorithm. The height of each patch is given by α_i. The region is flooded to a level $1/\nu^\star$ which uses a total quantity of water equal to one. The height of the water (shown shaded) above each patch is the optimal value of $x_i^\star$.

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.
