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

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

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

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

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

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

Hence

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

Since $g(\lambda, \nu) \le f_0(\tilde{x})$ 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 $-\mathbf{R}_+$.

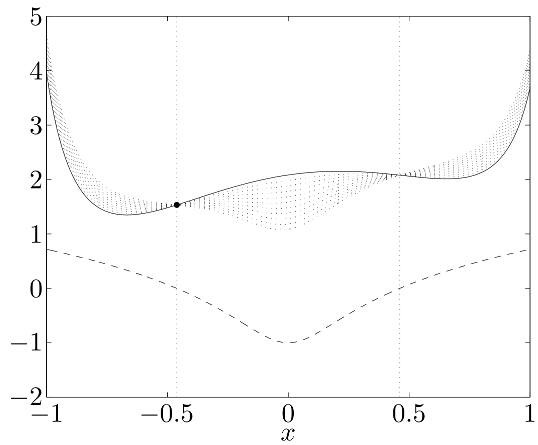

Figure 5.1 *Lower bound from a dual feasible point.* The solid curve shows the objective function f_0, and the dashed curve shows the constraint function f_1. The feasible set is the interval $[-0.46, 0.46]$, which is indicated by the two dotted vertical lines. The optimal point and value are $x^\star = -0.46$, $p^\star = 1.54$ (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 $L(x, \lambda) \leq f_0(x)$.

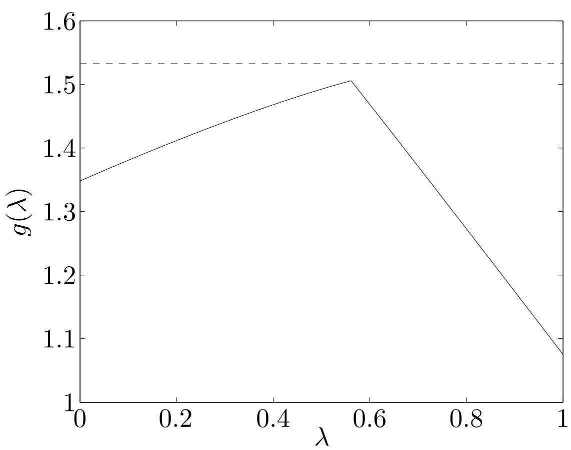

Figure 5.2 The dual function g for the problem in figure 5.1. Neither f_0 nor f_1 is convex, but the dual function is concave. The horizontal dashed line shows $p^\star$, the optimal value of the problem.

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,

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

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 $f_i(x) \leq 0$, 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 $\lambda_i \geq 0$, 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 $\lambda_i u \leq I_-(u)$ and $\nu_i u \leq I_0(u)$ 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 $\mathbf{R}^n \times$