1 Optimization

An *optimization problem* has the standard form

$$\begin{aligned} &\text{minimize} && f(x) \\ &\text{subject to} && h(x) = b \\ &&& x \in X. \end{aligned} \tag{1.1}$$

It consists of a vector $x \in \mathbb{R}^n$ of *decision variables*, an *objective function* $f : \mathbb{R}^n \to \mathbb{R}$, a *functional constraint* $h(x) = b$ where $h : \mathbb{R}^n \to \mathbb{R}^m$ and $b \in \mathbb{R}^m$, and a *regional constraint* $x \in X$ where $X \subseteq \mathbb{R}^n$. The set $X(b) = \{x \in X : h(x) = b\}$ is called the *feasible set*, and a problem is called *feasible* if $X(b)$ is non-empty and bounded if $f(x)$ is bounded from below on $X(b)$. A vector x^* is called optimal if it is in the feasible set and minimizes f among all vectors in the feasible set. The assumption that the functional constraint holds with equality is without loss of generality: an inequality constraint like $g(x) \leqslant b$ can be re-written as $g(x) + z = b$, where z is a new *slack variable* with the additional regional constraint $z \geqslant 0$.

1.1 Lagrangian Methods

A well-known method for solving constrained optimization problems is the method of Lagrange multipliers. The idea behind this method is to reduce constrained optimization to unconstrained optimization, and to take the (functional) constraints into account by augmenting the objective function with a weighted sum of them. To this end, define the *Lagrangian* associated with (1.1) as

$$L(x, \lambda) = f(x) - \lambda^T(h(x) - b), \tag{1.2}$$

where $\lambda \in \mathbb{R}^m$ is a vector of *Lagrange multipliers*.

The following result provides a condition under which minimizing the Lagrangian, subject only to the regional constraints, yields a solution to the original constrained problem. The result is easy to prove, yet extremely useful in practice.

THEOREM 1.1 (Lagrangian Sufficiency Theorem). *Let $x \in X$ and $\lambda \in \mathbb{R}^m$ such that $L(x, \lambda) = \inf_{x' \in X} L(x', \lambda)$ and $h(x) = b$. Then x is an optimal solution of* (1.1).

Proof. We have that

$$\begin{aligned} \min_{x' \in X(b)} f(x') &= \min_{x' \in X(b)} [f(x') - \lambda^T(h(x') - b)] \\ &\geqslant \min_{x' \in X} [f(x') - \lambda^T(h(x') - b)] \\ &= f(x) - \lambda^T(h(x) - b) = f(x). \end{aligned}$$

Equality in the first line holds because $h(x') - b = 0$ when $x' \in X(b)$. The inequality on the second line holds because the minimum is taken over a larger set. In the third line we finally use that x minimizes L and that $h(x) = b$. □

Two remarks are in order. First, a vector λ of Lagrange multipliers satisfying the conditions of the theorem is not guaranteed to exist in general, but it does exist for a large class of problems. Second, the theorem appears to be useful mainly for showing that a given solution x is optimal. In certain cases, however, it can also be used to find an optimal solution. Our general strategy in these cases will be to minimize $L(x, \lambda)$ for all values of λ, in order to obtain a minimizer $x^*(\lambda)$ that depends on λ, and then find λ^* such that $x^*(\lambda^*)$ satisfies the constraints. Let us apply this strategy to a concrete example.

Example 1.2. Consider minimizing $x_1^2 + x_2^2$ subject to $a_1x_1 + a_2x_2 = b$ and $x_1, x_2 \geqslant 0$ for some $a_1, a_2, b \geqslant 0$. The Lagrangian is

$$L((x_1, x_2), \lambda) = x_1^2 + x_2^2 - \lambda(a_1x_1 + a_2x_2 - b),$$

and taking partial derivaties reveals that it has a unique stationary point at $(x_1, x_2) = (\lambda a_1/2, \lambda a_2/2)$. We now choose λ such that the constraint $a_1x_1+a_2x_2 = b$ is satisfied at this point, which happens for $\lambda = 2b/(a_1^2+a_2^2)$. Since $\partial^2 L/\partial^2 x_1^2 > 0$, $\partial^2 L/\partial^2 x_2^2 > 0$, and $\partial^2 L/(\partial x_1 \partial x_2) = 0$ for this value of λ, we have found a minimum with value $b^2/(a_1^2+a_2^2)$ at $(x_1, x_2) = (a_1 b, a_2 b)/(a_1^2 + a_2^2)$.

More generally, to

$$\text{minimize } f(x) \text{ subject to } h(x) \leqslant b,\ x \in X, \tag{1.3}$$

we proceed as follows:

1. Introduce a vector z of slack variables to obtain the equivalent problem

$$\text{minimize } f(x) \text{ subject to } h(x) + z = b,\ x \in X,\ z \geqslant 0.$$

2. Compute the Lagrangian $L(x, z, \lambda) = f(x) - \lambda^T(h(x) + z - b)$.
3. Define the set

$$Y = \{\lambda \in \mathbb{R}^m : \inf_{x \in X, z \geqslant 0} L(x, z, \lambda) > -\infty\}.$$

4. For each $\lambda \in Y$, minimize $L(x, z, \lambda)$ subject only to the regional constraints, i.e., find $x^*(\lambda), z^*(\lambda)$ satisfying

$$L(x^*(\lambda), z^*(\lambda), \lambda) = \inf_{x \in X, z \geqslant 0} L(x, z, \lambda). \tag{1.4}$$

5. Find $\lambda^* \in Y$ such that $(x^*(\lambda^*), z^*(\lambda^*))$ is feasible, i.e., such that $x^*(\lambda^*) \in X$, $z^*(\lambda^*) \geqslant 0$, and $h(x^*(\lambda^*)) + z^*(\lambda^*) = b$. By Theorem 1.1, $x^*(\lambda^*)$ is optimal for (1.3).

1.2 The Lagrange Dual

Another useful concept that arises from the method of Lagrange multipliers is that of a dual problem. Denote by $\phi(b) = \inf_{x \in X(b)} f(x)$ the solution of (1.1), and define the (Lagrange) dual function $g : \mathbb{R}^m \to \mathbb{R}$ as the minimum value of the Lagrangian over X, i.e.,

$$g(\lambda) = \inf_{x \in X} L(x, \lambda).$$

Then, for all $\lambda \in \mathbb{R}^m$,

$$\inf_{x \in X(b)} f(x) = \inf_{x \in X(b)} L(x, \lambda) \geqslant \inf_{x \in X} L(x, \lambda) = g(\lambda), \tag{1.5}$$

i.e., the dual function provides a lower bound on the optimal value of (1.1). Since this holds for every value of λ, it is interesting to choose λ to make the lower bound as large as possible. This motivates the *dual problem* to

$$\begin{aligned} &\text{maximize} && g(\lambda) \\ &\text{subject to} && \lambda \in Y, \end{aligned}$$

where $Y = \{\lambda \in \mathbb{R}^m : g(\lambda) > -\infty\}$. In this context (1.1) is then referred to as the *primal problem*. Equation (1.5) is a proof of the *weak duality theorem*, which states that

$$\inf_{x \in X(b)} f(x) \geqslant \max_{\lambda \in Y} g(\lambda).$$

The primal problem (1.1) is said to satisfy *strong duality* if this holds with equality, i.e., if there exists λ such that

$$\phi(b) = g(\lambda).$$

If this is the case, then (1.1) can be solved using the method of Lagrangian multipliers. We can of course just try the method and see whether it works, as we did for Example 1.2. For certain important classes of optimization problems, however, it can be guaranteed that strong duality always holds.

1.3 Supporting Hyperplanes

A geometric interpretation of the dual function can be given in terms of ϕ. Fix $b \in \mathbb{R}^m$ and consider ϕ as a function of $c \in \mathbb{R}^m$. Further consider the hyperplane given by $\alpha : \mathbb{R}^m \to \mathbb{R}$ with

$$\alpha(c) = \beta - \lambda^T(b - c).$$

This hyperplane has intercept β at b and slope λ. We can now try to find $\phi(b)$ as follows:

1. For each λ, find $\beta_\lambda = \max\{\beta : \alpha(c) \leqslant \phi(c) \text{ for all } c \in \mathbb{R}^m\}$.
2. Choose λ to maximize β_λ.

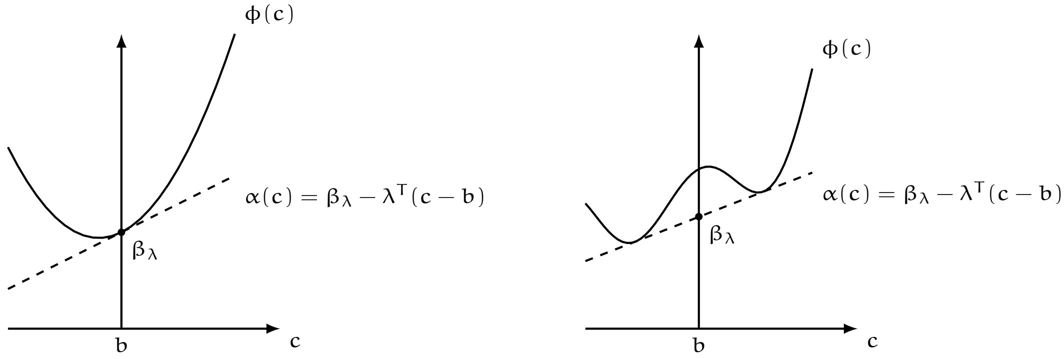

Figure 1.1: Geometric interpretation of the dual with optimal value $g(\lambda) = \beta_\lambda$. In the situation on the left strong duality holds, and $\beta_\lambda = \phi(b)$. In the situation on the right, strong duality does not hold, and $\beta_\lambda < \phi(b)$.

This approach is illustrated in Figure 1.1. We always have that $\beta_\lambda \leqslant \phi(b)$. In the situation on the left of Figure 1.1, this condition holds with equality because there is a tangent to ϕ at b. In fact,

$$\begin{aligned} g(\lambda) &= \inf_{x\in X} L(x,\lambda) \\ &= \inf_{c\in\mathbb{R}^m} \inf_{x\in X(c)} \big(f(x) - \lambda^T(h(x)-b)\big) \\ &= \inf_{c\in\mathbb{R}^m} \big(\phi(c) - \lambda^T(c-b)\big) \\ &= \sup\big\{\beta : \beta - \lambda^T(b-c) \leqslant \phi(c) \text{ for all } c \in \mathbb{R}^m\big\} \\ &= \beta_\lambda \end{aligned}$$

We again see the weak duality result as $\max_\lambda \beta_\lambda \leqslant \phi(b)$, but we also obtain a condition for strong duality. Call a hyperplane $\alpha : \mathbb{R}^m \to \mathbb{R}$ a *supporting hyperplane* to ϕ at b if $\alpha(c) = \phi(b) - \lambda^T(b-c)$ and $\phi(c) \geqslant \phi(b) - \lambda^T(b-c)$ for all $c \in \mathbb{R}^m$.

THEOREM 1.3. *The following are equivalent:*

1. *there exists a (non-vertical) supporting hyperplane to* ϕ *at* b*;*
2. *the problem satisfies strong duality.*

Proof. Suppose there exists a supporting hyperplane to ϕ at b. This means that there exists $\lambda \in \mathbb{R}^m$ such that for all $c \in \mathbb{R}^m$,

$$\phi(b) - \lambda^T(b-c) \leqslant \phi(c).$$

This implies that

$$\begin{aligned} \phi(b) &\leqslant \inf_{c\in\mathbb{R}^m} \big(\phi(c) - \lambda^T(c-b)\big) \\ &= \inf_{c\in\mathbb{R}^m} \inf_{x\in X(c)} \big(f(x) - \lambda^T(h(x)-b)\big) \\ &= \inf_{x\in X} L(x,\lambda) \\ &= g(\lambda). \end{aligned}$$

However, by (1.5) we have that $\phi(b) \geqslant g(\lambda)$. Hence $\phi(b) = g(\lambda)$, and strong duality holds.

Now suppose that the problem satisfies strong duality. Then there exists $\lambda \in \mathbb{R}^m$ such that for all $x \in X$,

$$\phi(b) \leqslant L(x, \lambda) = f(x) - \lambda^T(h(x) - b)$$

Minimizing the right hand side over $x \in X(c)$ yields that for all $c \in \mathbb{R}^m$

$$\phi(b) \leqslant \phi(c) - \lambda^T(c - b),$$

and hence

$$\phi(b) - \lambda^T(b - c) \leqslant \phi(c).$$

This describes a supporting hyperplane to ϕ at b. □

In the next lecture we will see that a supporting hyperplane exists for all $b \in \mathbb{R}^m$ if $\phi(b)$ is a convex function of b, and we will give sufficient conditions for this to be the case.