CHAPTER 13
Optimization

Consider a real-valued function f defined on an arbitrary subset X_0 of an arbitrary set X. We say that $x_0 \in X_0$ minimizes f on X_0 if

$$f(x_0) \le f(x) \text{ for all } x \in X_0.$$

For example, the problem of minimizing $f(x) = x \ln x$ for $x > 0$ falls into this framework if we put $X = \mathbb{R}$ and $X_0 = (0, \infty)$.

Sometimes, however, we want to restrict attention to a subset of X_0 characterized by a finite number of inequalities of the form $h_i(x) \le 0$, $i = 1, \ldots, m$, where each h_i is a real-valued function defined on X_0. We say that $x_0 \in X_0$ minimizes f subject to the constraints $h_i(x) \le 0$ for $x \in X_0$ if

$$f(x_0) \le f(x) \text{ for all } x \in X_0 \text{ with } h_i(x) \le 0,\ i = 1, \ldots, m.$$

For example, the problem of minimizing $f(x,y) = x \ln x + y \ln y$ for $x > 0$ and $y > 0$ satisfying $x^2 + y^2 \le 1$ falls into this framework if we put $X = \mathbb{R}^2$ and let X_0 denote the strictly positive first quadrant.

Notation. We put

$$H(x) := \begin{bmatrix} h_1(x) \\ \vdots \\ h_m(x) \end{bmatrix},$$

and write $H(x) \le 0$ to mean that $h_i(x) \le 0$ for $i = 1, \ldots, m$.

13.1. Introduction to Lagrange Multipliers

The generalization of the first quadrant in two-dimensional space to m-dimensional space is

$$\mathbb{R}^m_+ := \{\lambda \in \mathbb{R}^m : \lambda_i \ge 0, i = 1, \ldots, m\}.$$

The **Lagrangian** for the minimization problem subject to inequality constraints is the function $L : \mathbb{R}^m_+ \times X_0 \to \mathbb{R}$, defined by

$$L(\lambda, x) := f(x) + \lambda^{\mathsf{T}} H(x). \tag{13.1}$$

As we shall see, there is a close connection between unconstrained minimization of the Lagrangian over X_0 and minimization of f subject to the constraint $H(x) \le 0$ for

$x \in X_0$. In this section, we show that minimizing the Lagrangian is sufficient to minimize f subject to the constraint. In a later section, we show that in order for x_0 to minimize f subject to the constraint, it is necessary for a suitable Lagrange multiplier λ_0 to exist and that x_0 minimize $L(\lambda_0, x)$ for $x \in X_0$ (this result requires an additional assumption). Finally, as we know from calculus, the solution of minimization problems can often be made easier by using derivatives. For this reason, we will also study the derivative of the Lagrangian with respect to x.

Theorem 13.1 (Inequality Constraints). *If there exists a $\lambda_0 \in \mathbb{R}^m_+$ and an $x_0 \in X_0$ such that*

$$H(x_0) \le 0 \quad \text{and} \quad \lambda_0^{\mathsf{T}} H(x_0) = 0, \tag{13.2}$$

and such that

$$L(\lambda_0, x_0) \le L(\lambda_0, x), \quad \text{for all } x \in X_0, \tag{13.3}$$

then

$$f(x_0) \le f(x), \quad \text{for all } x \in X_0 \text{ with } H(x) \le 0. \tag{13.4}$$

Proof. Since $\lambda_0 \in \mathbb{R}^m_+$, for every $x \in X_0$ such that $H(x) \le 0$, $\lambda_0^{\mathsf{T}} H(x) \le 0$. Hence, we can write

$$\begin{aligned} f(x) &\ge f(x) + \lambda_0^{\mathsf{T}} H(x) \\ &= L(\lambda_0, x) \\ &\ge L(\lambda_0, x_0), && \text{by (13.3)}, \\ &= f(x_0) + \lambda_0^{\mathsf{T}} H(x_0) \\ &= f(x_0), && \text{by (13.2)}. \end{aligned}$$ □

In the problems you are asked to prove the following two theorems, which extend Theorem 13.1 to the case of equality constraints.

Theorem 13.2 (Equality Constraints). *Let $g_1, \ldots, g_k$ each map X_0 into $\mathbb{R}$, and put $G(x) := [g_1(x), \ldots, g_k(x)]^{\mathsf{T}}$. Put $L(\mu, x) := f(x) + \mu^{\mathsf{T}} G(x)$. If there exists a $\mu_0 \in \mathbb{R}^k$ and an $x_0 \in X_0$ such that $G(x_0) = 0$, and*

$$L(\mu_0, x_0) \le L(\mu_0, x), \quad \text{for all } x \in X_0,$$

then

$$f(x_0) \le f(x), \quad \text{for all } x \in X_0 \text{ with } G(x) = 0.$$

Theorem 13.3 (Mixed Constraints). *Let $H: X_0 \to \mathbb{R}^m$ and let $G: X_0 \to \mathbb{R}^k$. Put $L(\lambda, \mu, x) := f(x) + \lambda^{\mathsf{T}} H(x) + \mu^{\mathsf{T}} G(x)$. If there exists a $\lambda_0 \in \mathbb{R}^m_+$, a $\mu_0 \in \mathbb{R}^k$, and an $x_0 \in X_0$ such that* (13.2) *holds, $G(x_0) = 0$, and*

$$L(\lambda_0, \mu_0, x_0) \le L(\lambda_0, \mu_0, x), \quad \text{for all } x \in X_0,$$

then

$$f(x_0) \le f(x), \quad \text{for all } x \in X_0 \text{ with } H(x) \le 0 \text{ and } G(x) = 0.$$

13.2. Convex Functions

A real-valued function f is said to be **convex** if the line joining $(x, f(x))$ and $(y, f(y))$ lies above the function values $f(z)$ when z lies on the line joining x and y. A quick sketch of the graphs of functions like $f(x) = e^x$ and $f(x) = x^2$ illustrates the idea of a convex function. Functions of two or more variables can also be convex. For example, the bowl-shaped function $f(x,y) = x^2 + y^2$ is convex.

A little reflection shows that we need to make our definition of convex function more precise. For example, we must assume that f is defined for all points on the line joining x and y. This observation leads to the definition of a **convex set**. A subset C of a real or complex vector space is said to be convex if for all points $x, y \in C$, the line joining x and y also lies in C. More precisely, the line joining x and y is the set of points of the form

$$\lambda x + (1-\lambda)y, \quad 0 \le \lambda \le 1.$$

When $\lambda = 0$, this point reduces to y, and when $\lambda = 1$, this point reduces to x. Note that the above formula can also be written as $y + \lambda(x-y)$.

We can now give a precise definition of a **convex function**. A real-valued function f defined on a convex subset C of a real or complex vector space is said to be convex if

$$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda) f(y)$$

for all $x, y \in C$ and all $0 \le \lambda \le 1$. If the above inequality is strict for $0 < \lambda < 1$, we say that f is **strictly convex**. Note that the above inequality can also be written as

$$f(y + \lambda(x-y)) \le f(y) + \lambda[f(x) - f(y)].$$

Interchanging the roles of x and y yields

$$f(x + \lambda(y-x)) \le f(x) + \lambda[f(y) - f(x)].$$

If we rearrange this inequality, we find that

$$\frac{f(x + \lambda(y-x)) - f(x)}{\lambda} \le f(y) - f(x).$$

Let us put

$$(D^+ f)(x, y-x) := \lim_{\lambda \downarrow 0} \frac{f(x + \lambda(y-x)) - f(x)}{\lambda}. \tag{13.5}$$

Assuming this limit exists, it is called the **one-sided Gâteaux derivative** of f at x in the direction $y - x$. As we show below, for a convex function f, this limit always exists either as a finite number or as $-\infty$ Thus, for a convex function,

$$(D^+ f)(x, y-x) \le f(y) - f(x),$$

which we can rewrite as

$$f(y) \ge f(x) + (D^+ f)(x, y-x). \tag{13.6}$$

Theorem 13.4. *Let f be a real-valued function defined on a convex subset C of a real or complex vector space. If f is convex, then* (13.6) *holds for all $x, y \in C$. Conversely, if the derivatives in* (13.6) *exist and* (13.6) *holds for all $x, y \in C$, then f is convex; if the inequality in* (13.6) *is strict for $x, y \in C$ with $x \ne y$, then f is strictly convex.*

Proof. The foregoing analysis establishes the forward part of the theorem. The proof of the converse part is left to Problem 13.8. □

If there is a particular $x \in C$ for which the derivative in (13.6) is nonnegative for all $y \in C$, then $f(y) \ge f(x)$ for all $y \in C$; i.e., x is a **global minimizer** of f on C. The converse is also true; if x is a global minimizer, then the limit in (13.5) is nonnegative. We summarize this finding precisely in the following theorem.

Theorem 13.5. *Let f be a convex function defined on a convex subset C of a real or complex vector space. A point $x \in C$ satisfies $(D^+ f)(x, y-x) \ge 0$ for all $y \in C$ if and only if x is a global minimizer of f on C.*

Existence of the Limit in (13.5)

If we can show that the quotient in (13.5) is a nonincreasing function of λ, then as $\lambda \downarrow 0$, either the quotients tend to $-\infty$ or they are bounded below. In the latter case, the quotients tend to a finite limit (see Problem 13.13 for precise details). To show that the quotients are nonincreasing, we must show that for $0 < \lambda_1 < \lambda_2 \le 1$,

$$\frac{f(x + \lambda_1(y-x)) - f(x)}{\lambda_1} \le \frac{f(x + \lambda_2(y-x)) - f(x)}{\lambda_2}. \tag{13.7}$$

Put $y_1 := x + \lambda_1(y-x)$ and $y_2 := x + \lambda_2(y-x)$. Rearrange the second definition as $y - x = (y_2 - x)/\lambda_2$ and substitute this into the first definition so that

$$y_1 = x + \frac{\lambda_1}{\lambda_2}(y_2 - x).$$

Since $0 < \lambda_1/\lambda_2 < 1$, and since f is convex,

$$f(y_1) \le f(x) + \frac{\lambda_1}{\lambda_2}[f(y_2) - f(x)],$$

which we can rearrange as (13.7) as required.

13.3. Lagrange Multipliers and Derivatives

We return to the optimization problems considered in Section 13.1.

13.3.1. Sufficient Conditions

Theorem 13.6. *Let X_0 be a convex subset of an arbitrary real or complex vector space X. Assume that each of the real-valued functions f, $h_1, \ldots, h_m$ is convex on X_0. Suppose there is a point $x_0 \in X_0$ and a $\lambda_0 \in \mathbb{R}^m_+$ such that*

$$H(x_0) \le 0 \qquad \text{and} \qquad \lambda_0^{\mathrm{T}} H(x_0) = 0. \tag{13.8}$$

Suppose also that the Lagrangian in (13.1) *satisfies $(D_x^+ L)(\lambda_0, x_0, y - x_0) \ge 0$ for all $y \in X_0$. Then $f(x_0) \le f(y)$ for all $y \in X_0$ with $H(y) \le 0$.*

Proof. This is a simple application of Theorem 13.1 and the convexity property (13.6) applied to the Lagrangian. First note that (13.8) is the same as (13.2). It then suffices to establish that (13.3) holds. Since f and the h_i are convex, the Lagrangian $L(\lambda, x)$ is convex in x for $\lambda \in \mathbb{R}^m_+$ (Problem 13.14). Hence,

$$L(\lambda_0, y) \ge L(\lambda_0, x_0) + (D_x^+ L)(\lambda, x_0, y - x_0)$$

holds for all $y \in X_0$. By hypothesis the Gâteaux derivative is nonnegative, and so (13.3) follows as required. □

Theorem 13.7. *Let X_0 be a convex subset of a real vector space X. Assume that each of the real-valued functions f, $h_1, \ldots, h_m$ is convex on X_0. In addition, let $G(x) = Ax + b$, where $b \in \mathbb{R}^k$ and $A\colon X \to \mathbb{R}^k$ is linear. Suppose there exist $x_0 \in X_0$, $\lambda_0 \in \mathbb{R}^m_+$, and $\mu_0 \in \mathbb{R}^k$ such that*

$$H(x_0) \le 0, \quad \lambda_0^{\mathrm{T}} H(x_0) = 0, \quad \text{and} \quad G(x_0) = 0.$$

Suppose also that the Lagrangian

$$L(\lambda, \mu, x) := f(x) + \lambda^{\mathrm{T}} H(x) + \mu^{\mathrm{T}} G(x)$$

satisfies $(D_x^+ L)(\lambda_0, \mu_0, x_0, y - x_0) \ge 0$ for all $y \in X_0$. Then $f(x_0) \le f(y)$ for all $y \in X_0$ with $H(y) \le 0$ and $G(y) = 0$.

Proof. See Problem 13.14. □