6

Basics of Set-Constrained and Unconstrained Optimization

6.1 INTRODUCTION

In this chapter, we consider the optimization problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega. \end{array}$$

The function $f : \mathbb{R}^n \to \mathbb{R}$ that we wish to minimize is a real-valued function, and is called the *objective function*, or *cost function*. The vector x is an n-vector of independent variables, that is, $x = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n$. The variables $x_1, \ldots, x_n$ are often referred to as *decision variables*. The set Ω is a subset of $\mathbb{R}^n$, called the *constraint set* or *feasible set*.

The optimization problem above can be viewed as a decision problem that involves finding the "best" vector x of the decision variables over all possible vectors in Ω. By the "best" vector we mean the one that results in the smallest value of the objective function. This vector is called the *minimizer* of f over Ω. It is possible that there may be many minimizers. In this case, finding any of the minimizers will suffice.

There are also optimization problems that require maximization of the objective function. These problems, however, can be represented in the above form because maximizing f is equivalent to minimizing $-f$. Therefore, we can confine our attention to minimization problems without loss of generality.

The above problem is a general form of a *constrained* optimization problem, because the decision variables are constrained to be in the constraint set Ω. If $\Omega = \mathbb{R}^n$, then we refer to the problem as an *unconstrained* optimization problem. In this chapter, we discuss basic properties of the general optimization problem above,

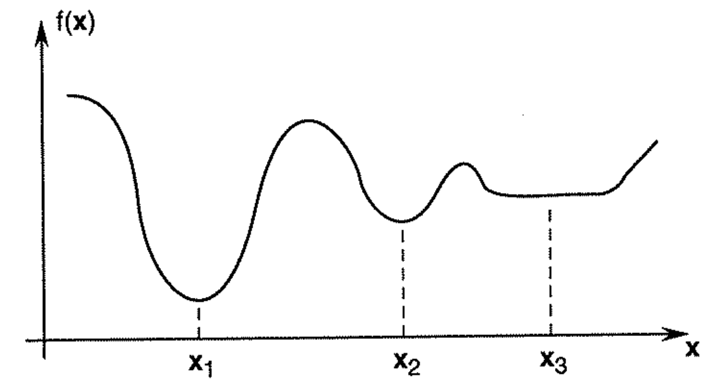

Figure 6.1 Examples of minimizers: x_1: strict global minimizer; x_2: strict local minimizer; x_3: local (not strict) minimizer

which includes the unconstrained case. In the remaining chapters of this part, we deal with iterative algorithms for solving unconstrained optimization problems.

The constraint "$x \in \Omega$" is called a *set constraint*. Often, the constraint set Ω takes the form $\Omega = \{x : h(x) = 0,\ g(x) \leq 0\}$, where h and g are given functions. We refer to such constraints as *functional constraints*. The remainder of this chapter deals with general set constraints, including the special case where $\Omega = \mathbb{R}^n$. The case where $\Omega = \mathbb{R}^n$ is called the *unconstrained* case. In Parts III and IV, we consider constrained optimization problems with functional constraints.

In considering the general optimization problem above, we distinguish between two kinds of minimizers, as specified by the following definitions.

Definition 6.1 ***Local minimizer.*** Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function defined on some set $\Omega \subset \mathbb{R}^n$. A point $x^* \in \Omega$ is a *local minimizer* of f over Ω if there exists $\varepsilon > 0$ such that $f(x) \geq f(x^*)$ for all $x \in \Omega \setminus \{x^*\}$ and $\|x - x^*\| < \varepsilon$.

Global minimizer. A point $x^* \in \Omega$ is a *global minimizer* of f over Ω if $f(x) \geq f(x^*)$ for all $x \in \Omega \setminus \{x^*\}$. ■

If, in the above definitions, we replace "$\geq$" with "$>$", then we have a *strict local minimizer* and a *strict global minimizer*, respectively.

In Figure 6.1, we graphically illustrate the above definitions for $n = 1$.

Given a real-valued function f, the notation $\arg\min f(x)$ denotes the argument that minimizes the function f (a point in the domain of f), assuming such a point is unique. For example, if $f : \mathbb{R} \to \mathbb{R}$ is given by $f(x) = (x+1)^2 + 3$, then $\arg\min f(x) = -1$. If we write $\arg\min_{x\in\Omega}$, then we treat Ω as the domain of f. For example, for the function f above, $\arg\min_{x\geq 0} f(x) = 0$. In general, we can think of $\arg\min_{x\in\Omega} f(x)$ as the global minimizer of f over Ω (assuming it exists and is unique).

Strictly speaking, an optimization problem is solved only when a global minimizer is found. However, global minimizers are, in general, difficult to find. Therefore, in practice, we often have to be satisfied with finding local minimizers.

6.2 CONDITIONS FOR LOCAL MINIMIZERS

In this section, we derive conditions for a point x^* to be a local minimizer. We use derivatives of a function $f : \mathbb{R}^n \to \mathbb{R}$. Recall that the first-order derivative of f, denoted Df, is

$$Df \triangleq \left[\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right].$$

Note that the gradient ∇f is just the transpose of Df; that is, $\nabla f = (Df)^T$. The second derivative of $f : \mathbb{R}^n \to \mathbb{R}$ (also called the *Hessian* of f) is

$$F(x) \triangleq D^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1}(x) \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}.$$

Example 6.1 Let $f(x_1, x_2) = 5x_1 + 8x_2 + x_1x_2 - x_1^2 - 2x_2^2$. Then,

$$Df(x) = (\nabla f(x))^T = \left[\frac{\partial f}{\partial x_1}(x), \frac{\partial f}{\partial x_2}(x)\right] = [5 + x_2 - 2x_1, 8 + x_1 - 4x_2],$$

and

$$F(x) = D^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) \\ \frac{\partial^2 f}{\partial x_1 \partial x_2}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ 1 & -4 \end{bmatrix}.$$

■

Given an optimization problem with constraint set Ω, a minimizer may lie either in the interior or on the boundary of Ω. To study the case where it lies on the boundary, we need the notion of *feasible directions*.

Definition 6.2 ***Feasible direction.*** A vector $d \in \mathbb{R}^n$, $d \neq 0$, is a *feasible direction* at $x \in \Omega$ if there exists $\alpha_0 > 0$ such that $x + \alpha d \in \Omega$ for all $\alpha \in [0, \alpha_0]$. ■

Figure 6.2 illustrates the notion of feasible directions.

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued function and let d be a feasible direction at $x \in \Omega$. The *directional derivative of f in the direction d*, denoted $\partial f / \partial d$, is the real-valued function defined by

$$\frac{\partial f}{\partial d}(x) = \lim_{\alpha \to 0} \frac{f(x + \alpha d) - f(x)}{\alpha}.$$

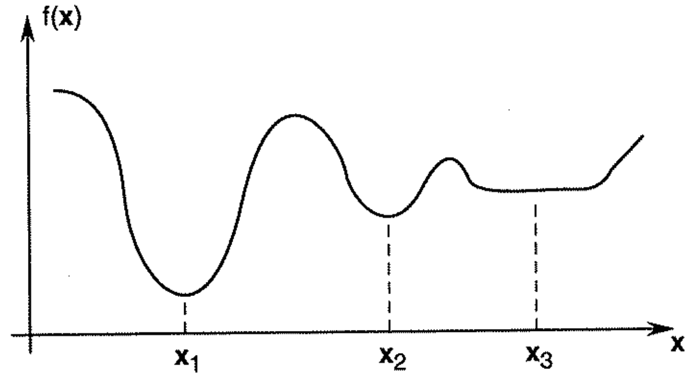

Figure 6.1 Examples of minimizers: x_1: strict global minimizer; x_2: strict local minimizer; x_3: local (not strict) minimizer

which includes the unconstrained case. In the remaining chapters of this part, we deal with iterative algorithms for solving unconstrained optimization problems.

The constraint "$x \in \Omega$" is called a *set constraint*. Often, the constraint set Ω takes the form $\Omega = \{x : h(x) = 0,\ g(x) \leq 0\}$, where h and g are given functions. We refer to such constraints as *functional constraints*. The remainder of this chapter deals with general set constraints, including the special case where $\Omega = \mathbb{R}^n$. The case where $\Omega = \mathbb{R}^n$ is called the *unconstrained* case. In Parts III and IV, we consider constrained optimization problems with functional constraints.

In considering the general optimization problem above, we distinguish between two kinds of minimizers, as specified by the following definitions.

Definition 6.1 ***Local minimizer.*** Suppose that $f : \mathbb{R}^n \to \mathbb{R}$ is a real-valued function defined on some set $\Omega \subset \mathbb{R}^n$. A point $x^* \in \Omega$ is a *local minimizer* of f over Ω if there exists $\varepsilon > 0$ such that $f(x) \geq f(x^*)$ for all $x \in \Omega \setminus \{x^*\}$ and $\|x - x^*\| < \varepsilon$.

Global minimizer. A point $x^* \in \Omega$ is a *global minimizer* of f over Ω if $f(x) \geq f(x^*)$ for all $x \in \Omega \setminus \{x^*\}$. ∎

If, in the above definitions, we replace "$\geq$" with "$>$", then we have a *strict local minimizer* and a *strict global minimizer*, respectively.

In Figure 6.1, we graphically illustrate the above definitions for $n = 1$.

Given a real-valued function f, the notation $\arg\min f(x)$ denotes the argument that minimizes the function f (a point in the domain of f), assuming such a point is unique. For example, if $f : \mathbb{R} \to \mathbb{R}$ is given by $f(x) = (x + 1)^2 + 3$, then $\arg\min f(x) = -1$. If we write $\arg\min_{x\in\Omega}$, then we treat Ω as the domain of f. For example, for the function f above, $\arg\min_{x\geq 0} f(x) = 0$. In general, we can think of $\arg\min_{x\in\Omega} f(x)$ as the global minimizer of f over Ω (assuming it exists and is unique).

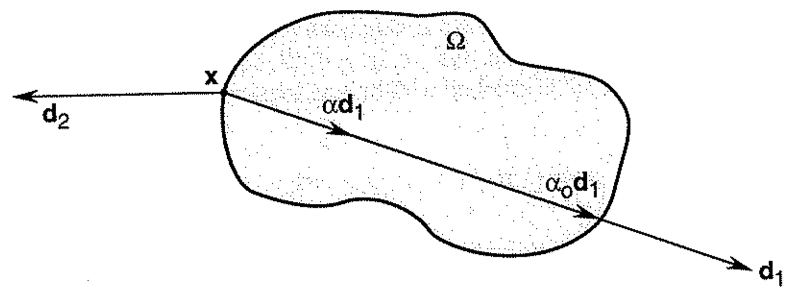

Figure 6.2 Two-dimensional illustration of feasible directions; $\boldsymbol{d}_1$ is a feasible direction, $\boldsymbol{d}_2$ is not a feasible direction

If $\|\boldsymbol{d}\| = 1$, then $\partial f/\partial \boldsymbol{d}$ is the rate of increase of f at $\boldsymbol{x}$ in the direction $\boldsymbol{d}$. To compute the above directional derivative, suppose that $\boldsymbol{x}$ and $\boldsymbol{d}$ are given. Then, $f(\boldsymbol{x} + \alpha \boldsymbol{d})$ is a function of α, and

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \frac{d}{d\alpha} f(\boldsymbol{x} + \alpha \boldsymbol{d})\bigg|_{\alpha=0}.$$

Applying the chain rule yields

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \frac{d}{d\alpha} f(\boldsymbol{x} + \alpha \boldsymbol{d})\bigg|_{\alpha=0} = \nabla f(\boldsymbol{x})^T \boldsymbol{d} = \langle \nabla f(\boldsymbol{x}), \boldsymbol{d} \rangle = \boldsymbol{d}^T \nabla f(\boldsymbol{x}).$$

In summary, if $\boldsymbol{d}$ is a unit vector, that is, $\|\boldsymbol{d}\| = 1$, then $\langle \nabla f(\boldsymbol{x}), \boldsymbol{d} \rangle$ is the rate of increase of f at the point $\boldsymbol{x}$ in the direction $\boldsymbol{d}$.

Example 6.2 Define $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(\boldsymbol{x}) = x_1 x_2 x_3$, and let

$$\boldsymbol{d} = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right]^T.$$

The directional derivative of f in the direction $\boldsymbol{d}$ is

$$\frac{\partial f}{\partial \boldsymbol{d}}(\boldsymbol{x}) = \nabla f(\boldsymbol{x})^T \boldsymbol{d} = [x_2 x_3, x_1 x_3, x_1 x_2] \begin{bmatrix} 1/2 \\ 1/2 \\ 1/\sqrt{2} \end{bmatrix} = \frac{x_2 x_3 + x_1 x_3 + \sqrt{2} x_1 x_2}{2}.$$

Note that because $\|\boldsymbol{d}\| = 1$, the above is also the rate of increase of f at $\boldsymbol{x}$ in the direction $\boldsymbol{d}$. ∎

We are now ready to state and prove the following theorem.

Theorem 6.1 ***First-Order Necessary Condition (FONC).*** *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-valued function on Ω. If $\boldsymbol{x}^*$ is a local minimizer of f over Ω, then for any feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$, we have*

$$\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) \geq 0.$$

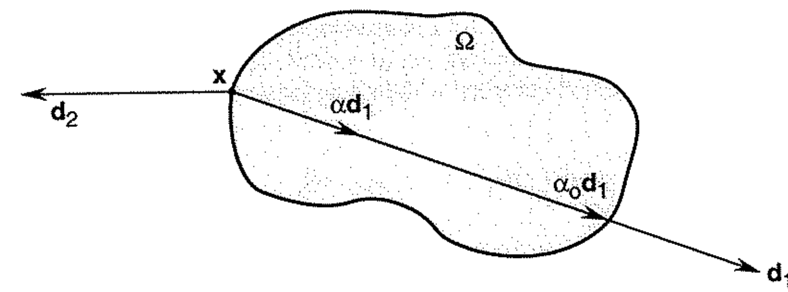

Figure 6.2 Two-dimensional illustration of feasible directions; d_1 is a feasible direction, d_2 is not a feasible direction

If $\|d\| = 1$, then $\partial f/\partial d$ is the rate of increase of f at x in the direction d. To compute the above directional derivative, suppose that x and d are given. Then, $f(x + \alpha d)$ is a function of α, and

$$\frac{\partial f}{\partial d}(x) = \frac{d}{d\alpha} f(x + \alpha d)\Big|_{\alpha=0}.$$

Applying the chain rule yields

$$\frac{\partial f}{\partial d}(x) = \frac{d}{d\alpha} f(x + \alpha d)\Big|_{\alpha=0} = \nabla f(x)^T d = \langle \nabla f(x), d\rangle = d^T \nabla f(x).$$

In summary, if d is a unit vector, that is, $\|d\| = 1$, then $\langle \nabla f(x), d\rangle$ is the rate of increase of f at the point x in the direction d.

Example 6.2 Define $f : \mathbb{R}^3 \to \mathbb{R}$ by $f(x) = x_1 x_2 x_3$, and let

$$d = \left[\frac{1}{2}, \frac{1}{2}, \frac{1}{\sqrt{2}}\right]^T.$$

The directional derivative of f in the direction d is

$$\frac{\partial f}{\partial d}(x) = \nabla f(x)^T d = [x_2x_3, x_1x_3, x_1x_2] \begin{bmatrix} 1/2 \\ 1/2 \\ 1/\sqrt{2} \end{bmatrix} = \frac{x_2x_3 + x_1x_3 + \sqrt{2}x_1x_2}{2}.$$

Note that because $\|d\| = 1$, the above is also the rate of increase of f at x in the direction d. ∎

We are now ready to state and prove the following theorem.

Theorem 6.1 ***First-Order Necessary Condition (FONC).*** *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-valued function on Ω. If x^* is a local minimizer of f over Ω, then for any feasible direction d at x^*, we have*

$$d^T \nabla f(x^*) \ge 0.$$

□

Proof. Define

$$x(\alpha) = x^* + \alpha d \in \Omega.$$

Note that $x(0) = x^*$. Define the composite function

$$\phi(\alpha) = f(x(\alpha)).$$

Then, by Taylor's theorem,

$$f(x^* + \alpha d) - f(x^*) = \phi(\alpha) - \phi(0) = \phi'(0)\alpha + o(\alpha) = \alpha d^T \nabla f(x(0)) + o(\alpha),$$

where $\alpha \ge 0$ (recall the definition of $o(\alpha)$ ("little-oh of α") in Part I). Thus, if $\phi(\alpha) \ge \phi(0)$, that is, $f(x^* + \alpha d) \ge f(x^*)$ for sufficiently small values of $\alpha > 0$ (x^* is a local minimizer), then we have to have $d^T \nabla f(x^*) \ge 0$ (see Exercise 5.7). ∎

The above theorem is graphically illustrated in Figure 6.3.

An alternative way to express the FONC is:

$$\frac{\partial f}{\partial d}(x^*) \ge 0$$

for all feasible directions d. In other words, if x^* is a local minimizer, then the rate of increase of f at x^* in any feasible direction d in Ω is nonnegative. Using directional derivatives, an alternative proof of Theorem 6.1 is as follows. Suppose that x^* is a local minimizer. Then, for any feasible direction d, there exists $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha})$,

$$f(x^*) \le f(x^* + \alpha d).$$

Hence, for all $\alpha \in (0, \bar{\alpha})$, we have

$$\frac{f(x^* + \alpha d) - f(x^*)}{\alpha} \ge 0.$$

Taking the limit as $\alpha \to 0$, we conclude that

$$\frac{\partial f}{\partial d}(x^*) \ge 0.$$

A special case of interest is when x^* is an interior point of Ω (see Section 4.4). In this case, any direction is feasible, and we have the following result.

Corollary 6.1 ***Interior case.*** *Let Ω be a subset of $\mathbb{R}^n$ and $f \in \mathcal{C}^1$ a real-valued function on Ω. If x^* is a local minimizer of f over Ω and if x^* is an interior point of Ω, then*

$$\nabla f(x^*) = \mathbf{0}.$$

□

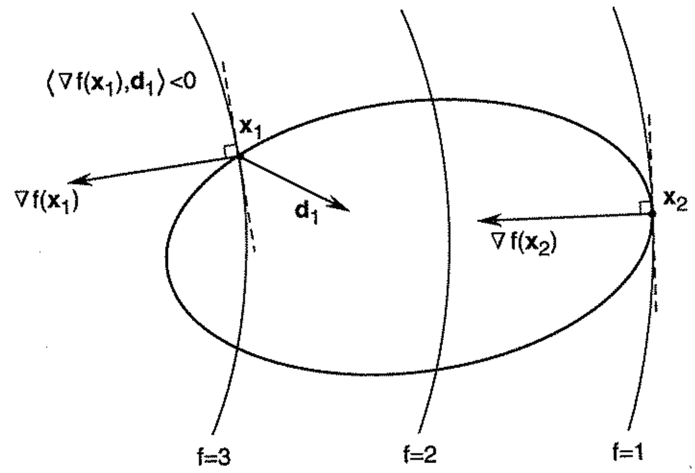

Figure 6.3 Illustration of the FONC for the constrained case; x_1 does not satisfy the FONC, x_2 satisfies the FONC

Proof. Suppose that f has a local minimizer x^* that is an interior point of Ω. Because x^* is an interior point of Ω, the set of feasible directions at x^* is the whole of $\mathbb{R}^n$. Thus, for any $d \in \mathbb{R}^n$, $d^T \nabla f(x^*) \geq 0$ and $-d^T \nabla f(x^*) \geq 0$. Hence, $d^T \nabla f(x^*) = 0$ for all $d \in \mathbb{R}^n$, which implies that $\nabla f(x^*) = \mathbf{0}$. ∎

Example 6.3 Consider the problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + 0.5x_2^2 + 3x_2 + 4.5 \\ \text{subject to} & x_1, x_2 \geq 0. \end{array}$$

Questions:

a. Is the first-order necessary condition (FONC) for a local minimizer satisfied at $x = [1, 3]^T$?

b. Is the FONC for a local minimizer satisfied at $x = [0, 3]^T$?

c. Is the FONC for a local minimizer satisfied at $x = [1, 0]^T$?

d. Is the FONC for a local minimizer satisfied at $x = [0, 0]^T$?

Answers: First, let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x) = x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$, where $x = [x_1, x_2]^T$. A plot of the level sets of f is shown in Figure 6.4.

a. At $x = [1, 3]^T$, we have $\nabla f(x) = [2x_1, x_2 + 3]^T = [2, 6]^T$. The point $x = [1, 3]^T$ is an interior point of $\Omega = \{x : x_1 \geq 0, x_2 \geq 0\}$. Hence, the FONC requires $\nabla f(x) = \mathbf{0}$. The point $x = [1, 3]^T$ does not satisfy the FONC for a local minimizer.

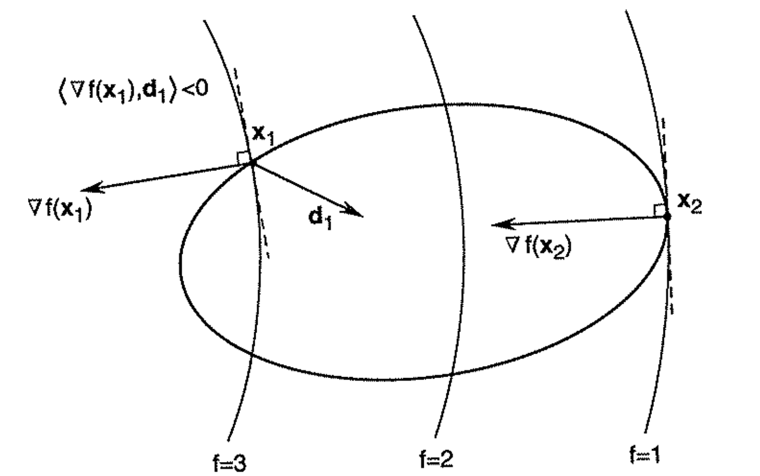

Figure 6.3 Illustration of the FONC for the constrained case; x_1 does not satisfy the FONC, x_2 satisfies the FONC

Proof. Suppose that f has a local minimizer x^* that is an interior point of Ω. Because x^* is an interior point of Ω, the set of feasible directions at x^* is the whole of $\mathbb{R}^n$. Thus, for any $d \in \mathbb{R}^n$, $d^T \nabla f(x^*) \geq 0$ and $-d^T \nabla f(x^*) \geq 0$. Hence, $d^T \nabla f(x^*) = 0$ for all $d \in \mathbb{R}^n$, which implies that $\nabla f(x^*) = 0$. ∎

Example 6.3 Consider the problem

$$\begin{array}{ll} \text{minimize} & x_1^2 + 0.5x_2^2 + 3x_2 + 4.5 \\ \text{subject to} & x_1, x_2 \geq 0. \end{array}$$

Questions:

a. Is the first-order necessary condition (FONC) for a local minimizer satisfied at $x = [1, 3]^T$?

b. Is the FONC for a local minimizer satisfied at $x = [0, 3]^T$?

c. Is the FONC for a local minimizer satisfied at $x = [1, 0]^T$?

d. Is the FONC for a local minimizer satisfied at $x = [0, 0]^T$?

Answers: First, let $f : \mathbb{R}^2 \to \mathbb{R}$ be defined by $f(x) = x_1^2 + 0.5x_2^2 + 3x_2 + 4.5$, where $x = [x_1, x_2]^T$. A plot of the level sets of f is shown in Figure 6.4.

a. At $x = [1, 3]^T$, we have $\nabla f(x) = [2x_1, x_2 + 3]^T = [2, 6]^T$. The point $x = [1, 3]^T$ is an interior point of $\Omega = \{x : x_1 \geq 0, x_2 \geq 0\}$. Hence, the FONC requires $\nabla f(x) = 0$. The point $x = [1, 3]^T$ does not satisfy the FONC for a local minimizer.

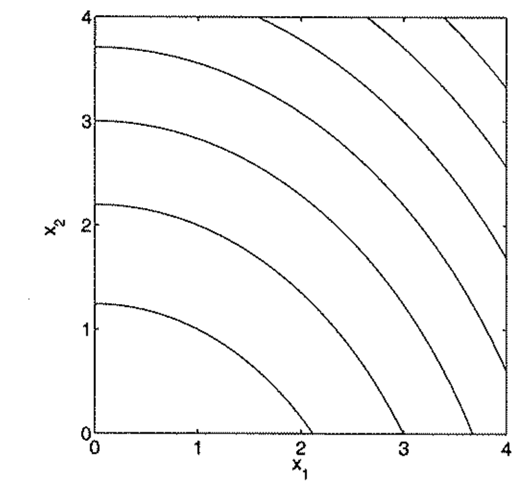

Figure 6.4 Level sets of the function in Example 6.3

b. At $x = [0, 3]^T$, we have $\nabla f(x) = [0, 6]^T$, and hence $d^T \nabla f(x) = 6d_2$, where $d = [d_1, d_2]^T$. For d to be feasible at x, we need $d_1 \geq 0$, and d_2 can take an arbitrary value in $\mathbb{R}$. The point $x = [0, 3]^T$ does not satisfy the FONC for a minimizer because d_2 is allowed to be less than zero. For example, $d = [1, -1]^T$ is a feasible direction, but $d^T \nabla f(x) = -6 < 0$.

c. At $x = [1, 0]^T$, we have $\nabla f(x) = [2, 3]^T$, and hence $d^T \nabla f(x) = 2d_1 + 3d_2$. For d to be feasible, we need $d_2 \geq 0$, and d_1 can take an arbitrary value in $\mathbb{R}$. For example, $d = [-5, 1]^T$ is a feasible direction. But $d^T \nabla f(x) = -7 < 0$. Thus, $x = [1, 0]^T$ does not satisfy the FONC for a local minimizer.

d. At $x = [0, 0]^T$, we have $\nabla f(x) = [0, 3]^T$, and hence $d^T \nabla f(x) = 3d_2$. For d to be feasible, we need $d_2 \geq 0$ and $d_1 \geq 0$. Hence, $x = [0, 0]^T$ satisfies the FONC for a local minimizer.

∎

Example 6.4 Figure 6.5 shows a simplified model of a cellular wireless system (the distances shown have been scaled down to make the calculations simpler). A mobile user (also called a "mobile") is located at position x (see Figure 6.5).

There are two basestation antennas, one for the primary basestation and another for the neighboring basestation. Both antennas are transmitting signals to the mobile user, at equal power. However, the power of the received signal as measured by the mobile is the reciprocal of the squared distance from the associated antenna (primary or neighboring basestation). We are interested in finding the position of the mobile that maximizes the *signal-to-interference ratio*, which is the ratio of the received signal power from the primary basestation to the received signal power from the neighboring basestation.

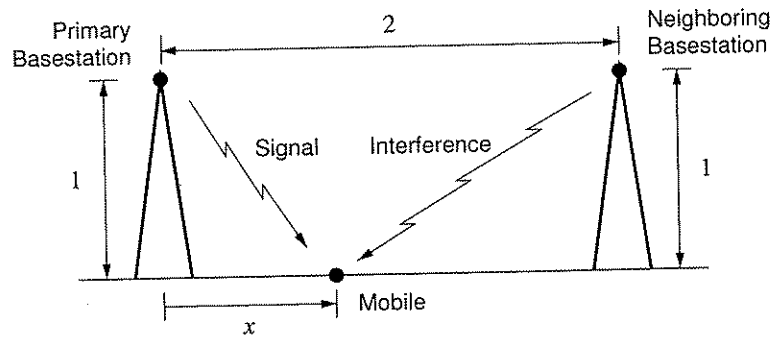

Figure 6.5 Simplified cellular wireless system in Example 6.4

We use the FONC to solve this problem. The squared distance from the mobile to the primary antenna is $1 + x^2$, while the squared distance from the mobile to the neighboring antenna is $1 + (2 - x)^2$. Therefore, the signal-to-interference ratio is

$$f(x) = \frac{1 + x^2}{1 + (2 - x)^2}.$$

We have

$$\begin{aligned} f'(x) &= \frac{-2x(1 + (2 - x)^2) - 2(2 - x)(1 + x^2)}{1 + (2 - x)^2} \\ &= \frac{4(x^2 - 2x - 1)}{1 + (2 - x)^2}. \end{aligned}$$

By the FONC, at the optimal position x^*, we have $f'(x^*) = 0$. Hence, either $x^* = 1 - \sqrt{2}$ or $x^* = 1 + \sqrt{2}$. Evaluating the objective function at these two candidate points, it easy to see that $x^* = 1 - \sqrt{2}$ is the optimal position. ∎

We now derive a second-order necessary condition that is satisfied by a local minimizer.

Theorem 6.2 ***Second-Order Necessary Condition (SONC).*** *Let $\Omega \subset \mathbb{R}^n$, $f \in \mathcal{C}^2$ a function on Ω, $\boldsymbol{x}^*$ a local minimizer of f over Ω, and $\boldsymbol{d}$ a feasible direction at $\boldsymbol{x}^*$. If $\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) = 0$, then*

$$\boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*) \boldsymbol{d} \geq 0,$$

where $\boldsymbol{F}$ is the Hessian of f. □

Proof. We prove the result by contradiction. Suppose that there is a feasible direction $\boldsymbol{d}$ at $\boldsymbol{x}^*$ such that $\boldsymbol{d}^T \nabla f(\boldsymbol{x}^*) = 0$ and $\boldsymbol{d}^T \boldsymbol{F}(\boldsymbol{x}^*) \boldsymbol{d} < 0$. Let $\boldsymbol{x}(\alpha) = \boldsymbol{x}^* + \alpha \boldsymbol{d}$ and define the composite function $\phi(\alpha) = f(\boldsymbol{x}^* + \alpha \boldsymbol{d}) = f(\boldsymbol{x}(\alpha))$. Then, by Taylor's theorem

$$\phi(\alpha) = \phi(0) + \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2),$$

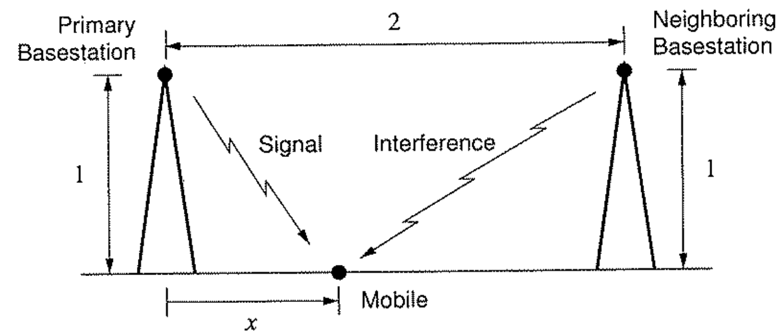

Figure 6.5 Simplified cellular wireless system in Example 6.4

We use the FONC to solve this problem. The squared distance from the mobile to the primary antenna is $1 + x^2$, while the squared distance from the mobile to the neighboring antenna is $1 + (2 - x)^2$. Therefore, the signal-to-interference ratio is

$$f(x) = \frac{1 + x^2}{1 + (2 - x)^2}.$$

We have

$$\begin{aligned} f'(x) &= \frac{-2x(1 + (2-x)^2) - 2(2-x)(1+x^2)}{1 + (2-x)^2} \\ &= \frac{4(x^2 - 2x - 1)}{1 + (2-x)^2}. \end{aligned}$$

By the FONC, at the optimal position x^*, we have $f'(x^*) = 0$. Hence, either $x^* = 1 - \sqrt{2}$ or $x^* = 1 + \sqrt{2}$. Evaluating the objective function at these two candidate points, it easy to see that $x^* = 1 - \sqrt{2}$ is the optimal position. ∎

We now derive a second-order necessary condition that is satisfied by a local minimizer.

Theorem 6.2 ***Second-Order Necessary Condition (SONC).*** *Let $\Omega \subset \mathbb{R}^n$, $f \in \mathcal{C}^2$ a function on Ω, x^* a local minimizer of f over Ω, and d a feasible direction at x^*. If $d^T \nabla f(x^*) = 0$, then*

$$d^T F(x^*) d \geq 0,$$

where F is the Hessian of f. □

Proof. We prove the result by contradiction. Suppose that there is a feasible direction d at x^* such that $d^T \nabla f(x^*) = 0$ and $d^T F(x^*) d < 0$. Let $x(\alpha) = x^* + \alpha d$ and define the composite function $\phi(\alpha) = f(x^* + \alpha d) = f(x(\alpha))$. Then, by Taylor's theorem

$$\phi(\alpha) = \phi(0) + \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2),$$

where by assumption $\phi'(0) = d^T \nabla f(x^*) = 0$, and $\phi''(0) = d^T F(x^*) d < 0$. For sufficiently small α,

$$\phi(\alpha) - \phi(0) = \phi''(0)\frac{\alpha^2}{2} + o(\alpha^2) < 0,$$

that is,

$$f(x^* + \alpha d) < f(x^*),$$

which contradicts the assumption that x^* is a local minimizer. Thus,

$$\phi''(0) = d^T F(x^*) d \geq 0.$$

∎

Corollary 6.2 ***Interior Case.*** *Let x^* be an interior point of $\Omega \subset \mathbb{R}^n$. If x^* is a local minimizer of $f : \Omega \to \mathbb{R}$, $f \in \mathcal{C}^2$, then*

$$\nabla f(x^*) = \mathbf{0},$$

and $F(x^*)$ is positive semidefinite ($F(x^*) \geq 0$); that is, for all $d \in \mathbb{R}^n$,

$$d^T F(x^*) d \geq 0.$$

□

Proof. If x^* is an interior point then all directions are feasible. The result then follows from Corollary 6.1 and Theorem 6.2. ∎

In the examples below, we show that the necessary conditions are *not* sufficient.

Example 6.5 Consider a function of one variable $f(x) = x^3$, $f : \mathbb{R} \to \mathbb{R}$. Because $f'(0) = 0$, and $f''(0) = 0$, the point $x = 0$ satisfies both the FONC and SONC. However, $x = 0$ is not a minimizer (see Figure 6.6). ∎

Example 6.6 Consider a function $f : \mathbb{R}^2 \to \mathbb{R}$, where $f(x) = x_1^2 - x_2^2$. The FONC requires that $\nabla f(x) = [2x_1, -2x_2]^T = \mathbf{0}$. Thus, $x = [0, 0]^T$ satisfies the FONC. The Hessian matrix of f is

$$F(x) = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}.$$

The Hessian matrix is indefinite; that is, for some $d_1 \in \mathbb{R}^2$ we have $d_1^T F d_1 > 0$, e.g., $d_1 = [1, 0]^T$; and, for some d_2, we have $d_2^T F d_2 < 0$, e.g., $d_2 = [0, 1]^T$. Thus, $x = [0, 0]^T$ does not satisfy the SONC, and hence it is not a minimizer. The graph of $f(x) = x_1^2 - x_2^2$ is shown in Figure 6.7. ∎

We now derive sufficient conditions that imply that x^* is a local minimizer.

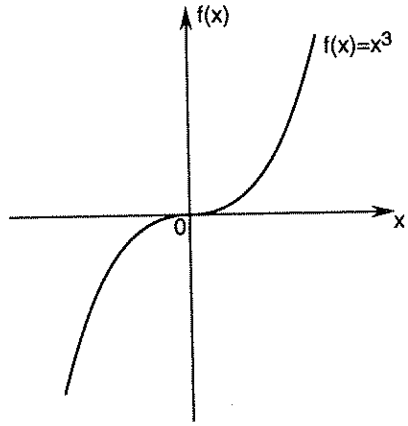

Figure 6.6 The point 0 satisfies the FONC and SONC, but is not a minimizer

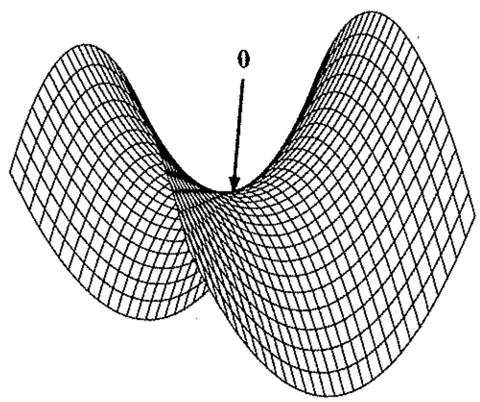

Figure 6.7 Graph of $f(\boldsymbol{x}) = x_1^2 - x_2^2$. The point $\mathbf{0}$ satisfies the FONC but not SONC; this point is not a minimizer

Theorem 6.3 ***Second-Order Sufficient Condition (SOSC), Interior Case.*** *Let $f \in \mathcal{C}^2$ be defined on a region in which $\boldsymbol{x}^*$ is an interior point. Suppose that*

1. $\nabla f(\boldsymbol{x}^*) = \mathbf{0}$; *and*
2. $\boldsymbol{F}(\boldsymbol{x}^*) > 0$.

Then, $\boldsymbol{x}^*$ is a strict local minimizer of f. □

Proof. Because $f \in \mathcal{C}^2$, we have $\boldsymbol{F}(\boldsymbol{x}^*) = \boldsymbol{F}^T(\boldsymbol{x}^*)$. Using assumption 2 and Rayleigh's inequality it follows that if $\boldsymbol{d} \neq \mathbf{0}$, then $0 < \lambda_{\min}(\boldsymbol{F}(\boldsymbol{x}^*))\|\boldsymbol{d}\|^2 \leq \boldsymbol{d}^T\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d}$. By Taylor's theorem and assumption 1,

$$f(\boldsymbol{x}^* + \boldsymbol{d}) - f(\boldsymbol{x}^*) = \frac{1}{2}\boldsymbol{d}^T\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} + o(\|\boldsymbol{d}\|^2) \geq \frac{\lambda_{\min}(\boldsymbol{F}(\boldsymbol{x}^*))}{2}\|\boldsymbol{d}\|^2 + o(\|\boldsymbol{d}\|^2).$$

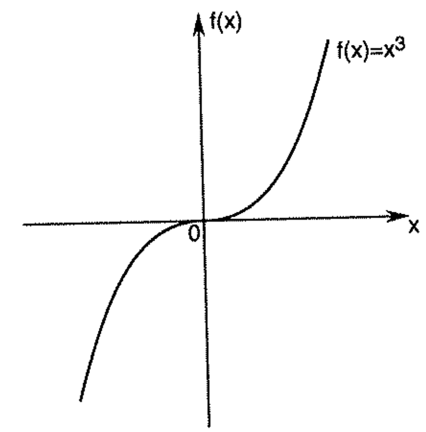

Figure 6.6 The point 0 satisfies the FONC and SONC, but is not a minimizer

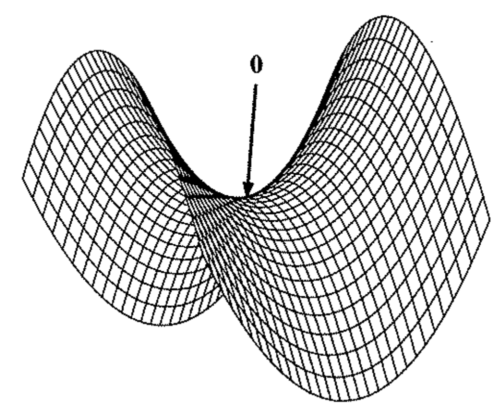

Figure 6.7 Graph of $f(\boldsymbol{x}) = x_1^2 - x_2^2$. The point $\boldsymbol{0}$ satisfies the FONC but not SONC; this point is not a minimizer

Theorem 6.3 ***Second-Order Sufficient Condition (SOSC), Interior Case.*** *Let $f \in \mathcal{C}^2$ be defined on a region in which $\boldsymbol{x}^*$ is an interior point. Suppose that*

1. $\nabla f(\boldsymbol{x}^*) = \boldsymbol{0}$; *and*
2. $\boldsymbol{F}(\boldsymbol{x}^*) > 0$.

Then, $\boldsymbol{x}^*$ is a strict local minimizer of f. □

Proof. Because $f \in \mathcal{C}^2$, we have $\boldsymbol{F}(\boldsymbol{x}^*) = \boldsymbol{F}^T(\boldsymbol{x}^*)$. Using assumption 2 and Rayleigh's inequality it follows that if $\boldsymbol{d} \neq \boldsymbol{0}$, then $0 < \lambda_{\min}(\boldsymbol{F}(\boldsymbol{x}^*))\|\boldsymbol{d}\|^2 \leq \boldsymbol{d}^T\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d}$. By Taylor's theorem and assumption 1,

$$f(\boldsymbol{x}^* + \boldsymbol{d}) - f(\boldsymbol{x}^*) = \frac{1}{2}\boldsymbol{d}^T\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{d} + o(\|\boldsymbol{d}\|^2) \geq \frac{\lambda_{\min}(\boldsymbol{F}(\boldsymbol{x}^*))}{2}\|\boldsymbol{d}\|^2 + o(\|\boldsymbol{d}\|^2).$$

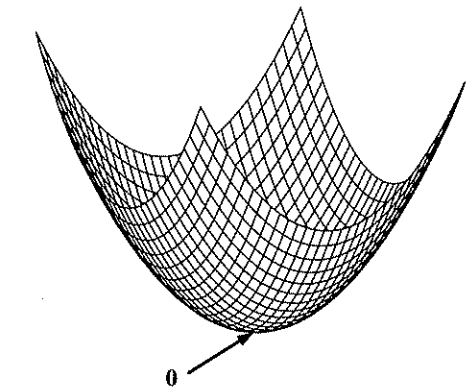

Figure 6.8 Graph of $f(\boldsymbol{x}) = x_1^2 + x_2^2$

Hence, for all $\boldsymbol{d}$ such that $\|\boldsymbol{d}\|$ is sufficiently small,

$$f(\boldsymbol{x}^* + \boldsymbol{d}) > f(\boldsymbol{x}^*),$$

and the proof is completed. ∎

Example 6.7 Let $f(\boldsymbol{x}) = x_1^2 + x_2^2$. We have $\nabla f(\boldsymbol{x}) = [2x_1, 2x_2]^T = \boldsymbol{0}$ if and only if $\boldsymbol{x} = [0, 0]^T$. For all $\boldsymbol{x} \in \mathbb{R}^2$, we have

$$\boldsymbol{F}(\boldsymbol{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} > 0.$$

The point $\boldsymbol{x} = [0, 0]^T$ satisfies the FONC, SONC, and SOSC. It is a strict local minimizer. Actually $\boldsymbol{x} = [0, 0]^T$ is a strict global minimizer. Figure 6.8 shows the graph of $f(\boldsymbol{x}) = x_1^2 + x_2^2$. ∎

In this chapter, we presented a theoretical basis for the solution of nonlinear unconstrained problems. In the following chapters, we are concerned with iterative methods of solving such problems. Such methods are of great importance in practice. Indeed, suppose that one is confronted with a highly nonlinear function of 20 variables. Then, the FONC requires the solution of 20 nonlinear simultaneous equations for 20 variables. These equations, being nonlinear, will normally have multiple solutions. In addition, we would have to compute 210 second derivatives (provided $f \in \mathcal{C}^2$) to use the SONC or SOSC. We begin our discussion of iterative methods in the next chapter with search methods for functions of one variable.

EXERCISES

6.1 Consider the problem

$$\text{minimize} \quad f(\boldsymbol{x})$$

$$\text{subject to} \qquad \boldsymbol{x} \in \Omega,$$

where $f \in \mathcal{C}^2$. For each of the following specifications for Ω, $\boldsymbol{x}^*$, and f, determine if the given point $\boldsymbol{x}^*$ is: (i) definitely a local minimizer; (ii) definitely not a local minimizer; or (iii) possibly a local minimizer. Fully justify your answer.

a. $f : \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{\boldsymbol{x} = [x_1, x_2]^T : x_1 \geq 1\}$, $\boldsymbol{x}^* = [1, 2]^T$, and gradient $\nabla f(\boldsymbol{x}^*) = [1, 1]^T$.

b. $f : \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{\boldsymbol{x} = [x_1, x_2]^T : x_1 \geq 1, x_2 \geq 2\}$, $\boldsymbol{x}^* = [1, 2]^T$, and gradient $\nabla f(\boldsymbol{x}^*) = [1, 0]^T$.

c. $f : \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{\boldsymbol{x} = [x_1, x_2]^T : x_1 \geq 0, x_2 \geq 0\}$, $\boldsymbol{x}^* = [1, 2]^T$, gradient $\nabla f(\boldsymbol{x}^*) = [0, 0]^T$, and Hessian $\boldsymbol{F}(\boldsymbol{x}^*) = \boldsymbol{I}$ (identity matrix).

d. $f : \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{\boldsymbol{x} = [x_1, x_2]^T : x_1 \geq 1, x_2 \geq 2\}$, $\boldsymbol{x}^* = [1, 2]^T$, gradient $\nabla f(\boldsymbol{x}^*) = [1, 0]^T$, and Hessian

$$\boldsymbol{F}(\boldsymbol{x}^*) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

6.2 Show that if $\boldsymbol{x}^*$ is a global minimizer of f over Ω, and $\boldsymbol{x}^* \in \Omega' \subset \Omega$, then $\boldsymbol{x}^*$ is a global minimizer of f over Ω'.

6.3 Suppose that $\boldsymbol{x}^*$ is a local minimizer of f over Ω, and $\Omega \subset \Omega'$. Show that if $\boldsymbol{x}^*$ is an interior point of Ω, then $\boldsymbol{x}^*$ is a local minimizer of f over Ω'. Show that the same conclusion cannot be made if $\boldsymbol{x}^*$ is not an interior point of Ω.

6.4 Let $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{x}_0 \in \mathbb{R}^n$, and $\Omega \subset \mathbb{R}^n$. Show that

$$\boldsymbol{x}_0 + \arg\min_{\boldsymbol{x} \in \Omega} f(\boldsymbol{x}) = \arg\min_{\boldsymbol{y} \in \Omega'} f(\boldsymbol{y}),$$

where $\Omega' = \{\boldsymbol{y} : \boldsymbol{y} - \boldsymbol{x}_0 \in \Omega\}$.

6.5 Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given below:

$$f(\boldsymbol{x}) = \boldsymbol{x}^T \begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix} \boldsymbol{x} + \boldsymbol{x}^T \begin{bmatrix} 3 \\ 5 \end{bmatrix} + 6.$$

a. Find the gradient and Hessian of f at the point $[1, 1]^T$.

b. Find the directional derivative of f at $[1, 1]^T$ with respect to a unit vector in the direction of maximal rate of increase.

c. Find a point that satisfies the FONC (interior case) for f. Does this point satisfy the SONC (for a minimizer)?

subject to $\quad x \in \Omega,$

where $f \in \mathcal{C}^2$. For each of the following specifications for Ω, x^*, and f, determine if the given point x^* is: (i) definitely a local minimizer; (ii) definitely not a local minimizer; or (iii) possibly a local minimizer. Fully justify your answer.

a. $f : \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{x = [x_1, x_2]^T : x_1 \geq 1\}$, $x^* = [1, 2]^T$, and gradient $\nabla f(x^*) = [1, 1]^T$.

b. $f : \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{x = [x_1, x_2]^T : x_1 \geq 1, x_2 \geq 2\}$, $x^* = [1, 2]^T$, and gradient $\nabla f(x^*) = [1, 0]^T$.

c. $f : \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{x = [x_1, x_2]^T : x_1 \geq 0, x_2 \geq 0\}$, $x^* = [1, 2]^T$, gradient $\nabla f(x^*) = [0, 0]^T$, and Hessian $F(x^*) = I$ (identity matrix).

d. $f : \mathbb{R}^2 \to \mathbb{R}$, $\Omega = \{x = [x_1, x_2]^T : x_1 \geq 1, x_2 \geq 2\}$, $x^* = [1, 2]^T$, gradient $\nabla f(x^*) = [1, 0]^T$, and Hessian

$$F(x^*) = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

6.2 Show that if x^* is a global minimizer of f over Ω, and $x^* \in \Omega' \subset \Omega$, then x^* is a global minimizer of f over Ω'.

6.3 Suppose that x^* is a local minimizer of f over Ω, and $\Omega \subset \Omega'$. Show that if x^* is an interior point of Ω, then x^* is a local minimizer of f over Ω'. Show that the same conclusion cannot be made if x^* is not an interior point of Ω.

6.4 Let $f : \mathbb{R}^n \to \mathbb{R}$, $x_0 \in \mathbb{R}^n$, and $\Omega \subset \mathbb{R}^n$. Show that

$$x_0 + \arg\min_{x \in \Omega} f(x) = \arg\min_{y \in \Omega'} f(y),$$

where $\Omega' = \{y : y - x_0 \in \Omega\}$.

6.5 Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given below:

$$f(x) = x^T \begin{bmatrix} 1 & 2 \\ 4 & 7 \end{bmatrix} x + x^T \begin{bmatrix} 3 \\ 5 \end{bmatrix} + 6.$$

a. Find the gradient and Hessian of f at the point $[1, 1]^T$.

b. Find the directional derivative of f at $[1, 1]^T$ with respect to a unit vector in the direction of maximal rate of increase.

c. Find a point that satisfies the FONC (interior case) for f. Does this point satisfy the SONC (for a minimizer)?

6.6 Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ given below:

$$f(x) = x^T \begin{bmatrix} 2 & 5 \\ -1 & 1 \end{bmatrix} x + x^T \begin{bmatrix} 3 \\ 4 \end{bmatrix} + 7.$$

a. Find the directional derivative of f at $[0, 1]^T$ in the direction $[1, 0]^T$.

b. Find all points that satisfy the first-order necessary condition for f.

Does f have a minimizer? If it does, then find all minimizer(s); otherwise explain why it does not.

6.7 Consider the problem

$$\begin{array}{ll} \text{minimize} & -x_2^2 \\ \text{subject to} & |x_2| \leq x_1^2 \\ & x_1 \geq 0, \end{array}$$

where $x_1, x_2 \in \mathbb{R}$.

a. Does the point $[x_1, x_1]^T = \mathbf{0}$ satisfy the first-order necessary condition for a minimizer? That is, if f is the objective function, is it true that $d^T \nabla f(\mathbf{0}) \geq 0$ for all feasible directions d at $\mathbf{0}$?

b. Is the point $[x_1, x_1]^T = \mathbf{0}$ a local minimizer, a strict local minimizer, a local maximizer, a strict local maximizer, or none of the above?

6.8 Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega, \end{array}$$

where $f : \mathbb{R}^2 \to \mathbb{R}$ is given by $f(x) = 5x_2$ with $x = [x_1, x_2]^T$, and $\Omega = \{x = [x_1, x_2]^T : x_1^2 + x_2 \geq 1\}$. Answer each of the following questions, showing complete justification.

a. Does the point $x^* = [0, 1]^T$ satisfy the first-order necessary condition?

b. Does the point $x^* = [0, 1]^T$ satisfy the second-order necessary condition?

c. Is the point $x^* = [0, 1]^T$ a local minimizer?

6.9 Consider the problem

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega, \end{array}$$

19

Problems with Equality Constraints

19.1 INTRODUCTION

In this part, we discuss methods for solving a class of nonlinear constrained optimization problems that can be formulated as:

$$\begin{array}{lll} \text{minimize} & f(\boldsymbol{x}) & \\ \text{subject to} & h_i(\boldsymbol{x}) = 0, & i = 1, \ldots, m \\ & g_j(\boldsymbol{x}) \leq 0, & j = 1, \ldots, p, \end{array}$$

where $\boldsymbol{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $h_i : \mathbb{R}^n \to \mathbb{R}$, $g_j : \mathbb{R}^n \to \mathbb{R}$, and $m \leq n$. In vector notation, the problem above can be represented in the following *standard form*:

$$\begin{array}{ll} \text{minimize} & f(\boldsymbol{x}) \\ \text{subject to} & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0} \\ & \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}, \end{array}$$

where $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, and $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^p$. As usual, we adopt the following terminology.

Definition 19.1 Any point satisfying the constraints is called a *feasible point*. The set of all feasible points

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{0}, \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{0}\}$$

is called the *feasible set*. ∎

Optimization problems of the above form are not new to us. Indeed, linear programming problems of the form

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

which we studied in Part III, are of this type.

As we remarked in Part II, there is no loss of generality by considering only minimization problems. For if we are confronted with a maximization problem, it can be easily transformed into the minimization problem by observing that

$$\text{maximize } f(\boldsymbol{x}) = \text{minimize } -f(\boldsymbol{x}).$$

We illustrate the problems we study in this part by considering the following simple numerical example.

Example 19.1

$$\begin{array}{ll} \text{minimize} & (x_1 - 1)^2 + x_2 - 2 \\ \text{subject to} & x_2 - x_1 = 1, \\ & x_1 + x_2 \leq 2. \end{array}$$

This problem is already in the standard form given earlier, with $f(x_1, x_2) = (x_1 - 1)^2 + x_2 - 2$, $h(x_1, x_2) = x_2 - x_1 - 1$, and $g(x_1, x_2) = x_1 + x_2 - 2$. This problem turns out to be simple enough to be solved graphically (see Figure 19.1). In the figure the set of points that satisfy the constraints (the feasible set) is marked by the heavy solid line. The inverted parabolas represent level sets of the objective function f—the lower the level set, the smaller the objective function value. Therefore, the solution can be obtained by finding the lowest level set that intersects the feasible set. In this case, the minimizer lies on the level set with $f = -1/4$. The minimizer of the objective function is $\boldsymbol{x}^* = [1/2, 3/2]^T$. ∎

In the remainder of this chapter, we discuss constrained optimization problems with only equality constraints. The general constrained optimization problem is discussed in the chapters to follow.

19.2 PROBLEM FORMULATION

The class of optimization problems we analyze in this chapter is

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

where $\boldsymbol{x} \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{h} : \mathbb{R}^n \to \mathbb{R}^m$, $\boldsymbol{h} = [h_1, \ldots, h_m]^T$, and $m \leq n$. We assume that the function $\boldsymbol{h}$ is continuously differentiable, that is, $\boldsymbol{h} \in \mathcal{C}^1$.

Optimization problems of the above form are not new to us. Indeed, linear programming problems of the form

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

which we studied in Part III, are of this type.

As we remarked in Part II, there is no loss of generality by considering only minimization problems. For if we are confronted with a maximization problem, it can be easily transformed into the minimization problem by observing that

$$\text{maximize } f(x) = \text{minimize } -f(x).$$

We illustrate the problems we study in this part by considering the following simple numerical example.

Example 19.1

$$\begin{array}{ll} \text{minimize} & (x_1 - 1)^2 + x_2 - 2 \\ \text{subject to} & x_2 - x_1 = 1, \\ & x_1 + x_2 \leq 2. \end{array}$$

This problem is already in the standard form given earlier, with $f(x_1, x_2) = (x_1 - 1)^2 + x_2 - 2$, $h(x_1, x_2) = x_2 - x_1 - 1$, and $g(x_1, x_2) = x_1 + x_2 - 2$. This problem turns out to be simple enough to be solved graphically (see Figure 19.1). In the figure the set of points that satisfy the constraints (the feasible set) is marked by the heavy solid line. The inverted parabolas represent level sets of the objective function f—the lower the level set, the smaller the objective function value. Therefore, the solution can be obtained by finding the lowest level set that intersects the feasible set. In this case, the minimizer lies on the level set with $f = -1/4$. The minimizer of the objective function is $x^* = [1/2, 3/2]^T$. ∎

In the remainder of this chapter, we discuss constrained optimization problems with only equality constraints. The general constrained optimization problem is discussed in the chapters to follow.

19.2 PROBLEM FORMULATION

The class of optimization problems we analyze in this chapter is

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

where $x \in \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$, $h : \mathbb{R}^n \to \mathbb{R}^m$, $h = [h_1, \ldots, h_m]^T$, and $m \leq n$. We assume that the function h is continuously differentiable, that is, $h \in \mathcal{C}^1$.

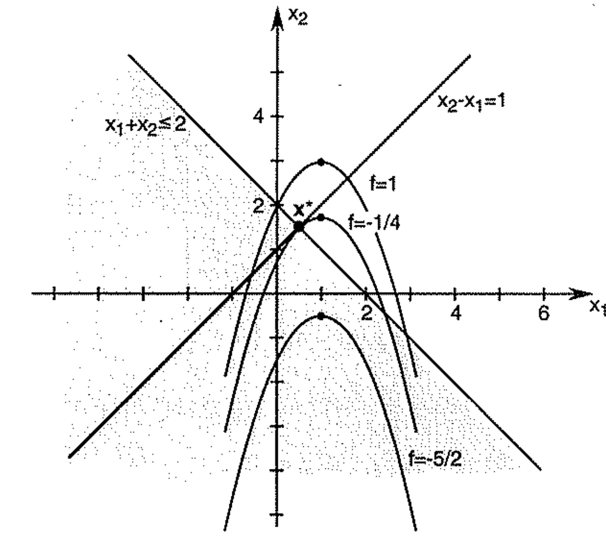

Figure 19.1 Graphical solution to the problem in Example 19.1

We introduce the following definition.

Definition 19.2 A point x^* satisfying the constraints $h_1(x^*) = 0, \ldots, h_m(x^*) = 0$ is said to be a *regular point* of the constraints if the gradient vectors $\nabla h_1(x^*), \ldots, \nabla h_m(x^*)$ are linearly independent. ∎

Let $Dh(x^*)$ be the Jacobian matrix of $h = [h_1, \ldots, h_m]^T$ at x^*, given by

$$Dh(x^*) = \begin{bmatrix} Dh_1(x^*) \\ \vdots \\ Dh_m(x^*) \end{bmatrix} = \begin{bmatrix} \nabla h_1(x^*)^T \\ \vdots \\ \nabla h_m(x^*)^T \end{bmatrix}.$$

Then, x^* is regular if and only if $\operatorname{rank} Dh(x^*) = m$, that is, the Jacobian matrix is of full rank.

The set of equality constraints $h_1(x) = 0, \ldots, h_m(x) = 0$, $h_i : \mathbb{R}^n \to \mathbb{R}$, describes a surface

$$S = \{x \in \mathbb{R}^n : h_1(x) = 0, \ldots, h_m(x) = 0\}.$$

Assuming the points in S are regular, the dimension of the surface S is $n - m$.

Example 19.2 Let $n = 3$ and $m = 1$ (i.e., we are operating in $\mathbb{R}^3$). Assuming that all points in S are regular, the set S is a two-dimensional surface. For example, let

$$h_1(x) = x_2 - x_3^2 = 0.$$

Note that $\nabla h_1(x) = [0, 1, -2x_3]^T$, and hence for any $x \in \mathbb{R}^3$, $\nabla h_1(x) \neq 0$. In this case,

$$\dim S = \dim\{x : h_1(x) = 0\} = n - m = 2.$$