

CHAPTER 6

BASICS OF SET-CONSTRAINED AND UNCONSTRAINED OPTIMIZATION

6.1 Introduction

In this chapter we consider the optimization problem

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

The function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ that we wish to minimize is a real-valued function called the *objective function* or *cost function*. The vector $\mathbf{x}$ is an n -vector of independent variables: $\mathbf{x} = [x_1, x_2, \dots, x_n]^\top \in \mathbb{R}^n$. The variables $x_1, \dots, x_n$ 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 $\mathbf{x}$ 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, in which case we seek *maximizers*. Minimizers and maximizers are also called *extremizers*. Maximization problems, however, can be represented equivalently in the minimization form above because maximizing f is equivalent to minimizing $-f$. Therefore, we can confine our attention to minimization problems without loss of generality.

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

The constraint " $\mathbf{x} \in \Omega$ " is called a *set constraint*. Often, the constraint set Ω takes the form $\Omega = \{\mathbf{x} : \mathbf{h}(\mathbf{x}) = \mathbf{0}, \mathbf{g}(\mathbf{x}) \leq \mathbf{0}\}$, where $\mathbf{h}$ and $\mathbf{g}$ 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 Suppose that $f : \mathbb{R}^n \rightarrow \mathbb{R}$ is a real-valued function defined on some set $\Omega \subset \mathbb{R}^n$. A point $\mathbf{x}^* \in \Omega$ is a *local minimizer* of f over Ω if there exists $\varepsilon > 0$ such that $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega \setminus \{\mathbf{x}^*\}$ and $\|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$. A point $\mathbf{x}^* \in \Omega$ is a *global minimizer* of f over Ω if $f(\mathbf{x}) \geq f(\mathbf{x}^*)$ for all $\mathbf{x} \in \Omega \setminus \{\mathbf{x}^*\}$. ■

If in the definitions above we replace " $\geq$ " with " $>$," then we have a *strict local minimizer* and a *strict global minimizer*, respectively. In Figure 6.1, we illustrate the definitions for $n = 1$.

If $\mathbf{x}^*$ is a global minimizer of f over Ω , we write $f(\mathbf{x}^*) = \min_{\mathbf{x} \in \Omega} f(\mathbf{x})$ and $\mathbf{x}^* = \arg \min_{\mathbf{x} \in \Omega} f(\mathbf{x})$. If the minimization is unconstrained, we simply write $\mathbf{x}^* = \arg \min_{\mathbf{x}} f(\mathbf{x})$ or $\mathbf{x}^* = \arg \min f(\mathbf{x})$. In other words, given a real-valued function f , the notation $\arg \min f(\mathbf{x})$ denotes the *argument* that minimizes the function f (a point in the domain of f), assuming that such a point is unique (if there is more than one such point, we pick one arbitrarily). For example, if $f : \mathbb{R} \rightarrow \mathbb{R}$ is given by $f(x) = (x+1)^2 + 3$, then $\arg \min f(x) = -1$. If we write $\arg \min_{\mathbf{x} \in \Omega}$, then we treat " $\mathbf{x} \in \Omega$ " to be a constraint for the minimization. For example, for the function f above, $\arg \min_{x \geq 0} f(x) = 0$.

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.

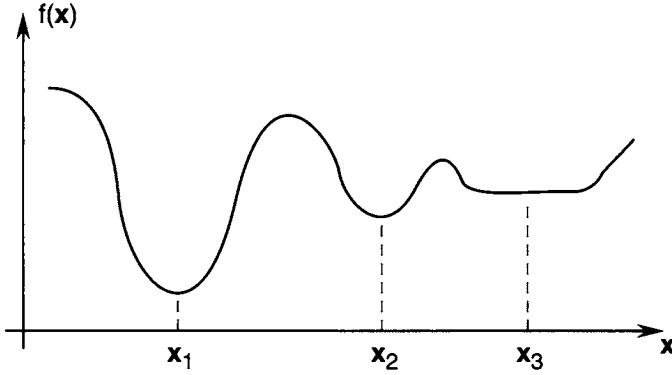


Figure 6.1 Examples of minimizers: x_1 : strict global minimizer; x_2 : strict local minimizer; x_3 : local (not strict) minimizer.

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 $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Recall that the first-order derivative of f , denoted Df , is

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

Note that the gradient ∇f is just the transpose of Df ; that is, $\nabla f = (Df)^\top$. The second derivative of $f : \mathbb{R}^n \rightarrow \mathbb{R}$ (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,

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

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*.

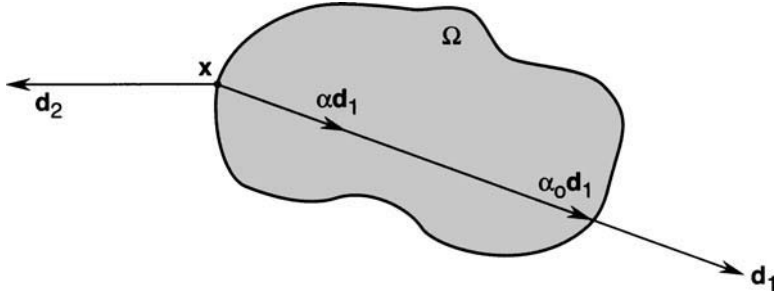


Figure 6.2 Two-dimensional illustration of feasible directions; $\mathbf{d}_1$ is a feasible direction, $\mathbf{d}_2$ is not a feasible direction.

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

Figure 6.2 illustrates the notion of feasible directions.

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

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

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

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

Applying the chain rule yields

$$\frac{\partial f}{\partial \mathbf{d}}(\mathbf{x}) = \left. \frac{d}{d\alpha} f(\mathbf{x} + \alpha\mathbf{d}) \right|_{\alpha=0} = \nabla f(\mathbf{x})^\top \mathbf{d} = \langle \nabla f(\mathbf{x}), \mathbf{d} \rangle = \mathbf{d}^\top \nabla f(\mathbf{x}).$$

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

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

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

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

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

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

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 $f \in C^1$ a real-valued function on Ω . If $\mathbf{x}^*$ is a local minimizer of f over Ω , then for any feasible direction $\mathbf{d}$ at $\mathbf{x}^*$, we have*

$$\mathbf{d}^\top \nabla f(\mathbf{x}^*) \geq 0.$$

□

Proof. Define

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

Note that $\mathbf{x}(0) = \mathbf{x}^*$. Define the composite function

$$\phi(\alpha) = f(\mathbf{x}(\alpha)).$$

Then, by Taylor's theorem,

$$f(\mathbf{x}^* + \alpha \mathbf{d}) - f(\mathbf{x}^*) = \phi(\alpha) - \phi(0) = \phi'(0)\alpha + o(\alpha) = \alpha \mathbf{d}^\top \nabla f(\mathbf{x}(0)) + o(\alpha),$$

where $\alpha \geq 0$ [recall the definition of $o(\alpha)$ ("little-oh of α ") in Part I]. Thus, if $\phi(\alpha) \geq \phi(0)$, that is, $f(\mathbf{x}^* + \alpha \mathbf{d}) \geq f(\mathbf{x}^*)$ for sufficiently small values of $\alpha > 0$ ($\mathbf{x}^*$ is a local minimizer), then we have to have $\mathbf{d}^\top \nabla f(\mathbf{x}^*) \geq 0$ (see Exercise 5.8). ■

Theorem 6.1 is illustrated in Figure 6.3.

An alternative way to express the FONC is

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

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

$$f(\mathbf{x}^*) \leq f(\mathbf{x}^* + \alpha \mathbf{d}).$$

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

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

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

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

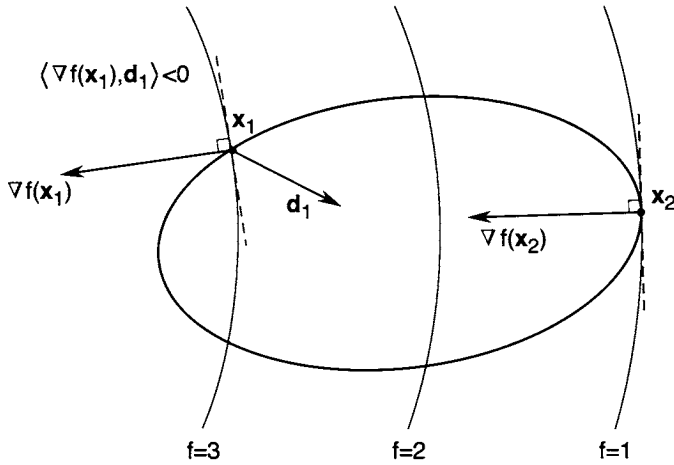


Figure 6.3 Illustration of the FONC for a constrained case; $\mathbf{x}_1$ does not satisfy the FONC, whereas $\mathbf{x}_2$ satisfies the FONC.

A special case of interest is when $\mathbf{x}^*$ 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 $f \in C^1$ a real-valued function on Ω . If $\mathbf{x}^*$ is a local minimizer of f over Ω and if $\mathbf{x}^*$ is an interior point of Ω , then*

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

□

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

Example 6.3 Consider the problem

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

- Is the first-order necessary condition (FONC) for a local minimizer satisfied at $\mathbf{x} = [1, 3]^\top$?
- Is the FONC for a local minimizer satisfied at $\mathbf{x} = [0, 3]^\top$?
- Is the FONC for a local minimizer satisfied at $\mathbf{x} = [1, 0]^\top$?

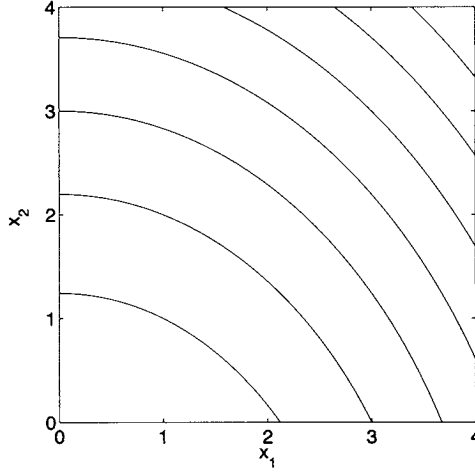


Figure 6.4 Level sets of the function in Example 6.3.

d. Is the FONC for a local minimizer satisfied at $\mathbf{x} = [0, 0]^\top$?

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

- a. At $\mathbf{x} = [1, 3]^\top$, we have $\nabla f(\mathbf{x}) = [2x_1, x_2 + 3]^\top = [2, 6]^\top$. The point $\mathbf{x} = [1, 3]^\top$ is an interior point of $\Omega = \{\mathbf{x} : x_1 \geq 0, x_2 \geq 0\}$. Hence, the FONC requires that $\nabla f(\mathbf{x}) = \mathbf{0}$. The point $\mathbf{x} = [1, 3]^\top$ does not satisfy the FONC for a local minimizer.
- b. At $\mathbf{x} = [0, 3]^\top$, we have $\nabla f(\mathbf{x}) = [0, 6]^\top$, and hence $\mathbf{d}^\top \nabla f(\mathbf{x}) = 6d_2$, where $\mathbf{d} = [d_1, d_2]^\top$. For $\mathbf{d}$ to be feasible at $\mathbf{x}$, we need $d_1 \geq 0$, and d_2 can take an arbitrary value in $\mathbb{R}$. The point $\mathbf{x} = [0, 3]^\top$ does not satisfy the FONC for a minimizer because d_2 is allowed to be less than zero. For example, $\mathbf{d} = [1, -1]^\top$ is a feasible direction, but $\mathbf{d}^\top \nabla f(\mathbf{x}) = -6 < 0$.
- c. At $\mathbf{x} = [1, 0]^\top$, we have $\nabla f(\mathbf{x}) = [2, 3]^\top$, and hence $\mathbf{d}^\top \nabla f(\mathbf{x}) = 2d_1 + 3d_2$. For $\mathbf{d}$ to be feasible, we need $d_2 \geq 0$, and d_1 can take an arbitrary value in $\mathbb{R}$. For example, $\mathbf{d} = [-5, 1]^\top$ is a feasible direction. But $\mathbf{d}^\top \nabla f(\mathbf{x}) = -7 < 0$. Thus, $\mathbf{x} = [1, 0]^\top$ does not satisfy the FONC for a local minimizer.
- d. At $\mathbf{x} = [0, 0]^\top$, we have $\nabla f(\mathbf{x}) = [0, 3]^\top$, and hence $\mathbf{d}^\top \nabla f(\mathbf{x}) = 3d_2$. For $\mathbf{d}$ to be feasible, we need $d_2 \geq 0$ and $d_1 \geq 0$. Hence, $\mathbf{x} = [0, 0]^\top$ 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

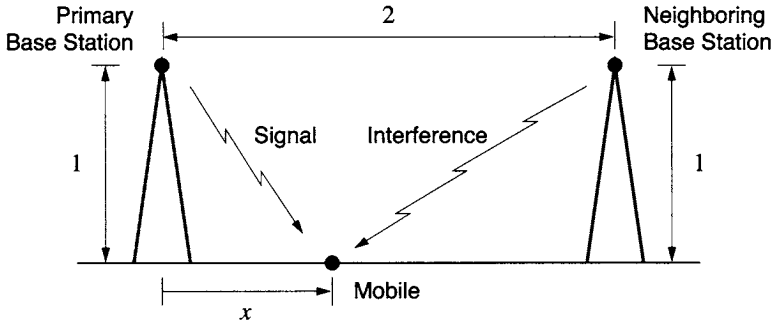


Figure 6.5 Simplified cellular wireless system in Example 6.4.

simpler). A mobile user (also called a *mobile*) is located at position x (see Figure 6.5).

There are two base station antennas, one for the primary base station and another for the neighboring base station. 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 base station). We are interested in finding the position of the mobile that maximizes the *signal-to-interference ratio*, which is the ratio of the signal power received from the primary base station to the signal power received from the neighboring base station.

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

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

We have

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

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 is easy to see that $x^* = 1 - \sqrt{2}$ is the optimal position. ■

The next example illustrates that in some problems the FONC is not helpful for eliminating candidate local minimizers. However, in such cases, there may be a recasting of the problem into an equivalent form that makes the FONC useful.

Example 6.5 Consider the set-constrained problem

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

where $\Omega = \{[x_1, x_2]^\top : x_1^2 + x_2^2 = 1\}$.

- a. Consider a point $\mathbf{x}^* \in \Omega$. Specify all feasible directions at $\mathbf{x}^*$.
- b. Which points in Ω satisfy the FONC for this set-constrained problem?
- c. Based on part b, is the FONC for this set-constrained problem useful for eliminating local-minimizer candidates?
- d. Suppose that we use polar coordinates to parameterize points $\mathbf{x} \in \Omega$ in terms of a single parameter θ :

$$x_1 = \cos \theta \quad x_2 = \sin \theta.$$

Now use the FONC for unconstrained problems (with respect to θ) to derive a necessary condition of this sort: If $\mathbf{x}^* \in \Omega$ is a local minimizer, then $\mathbf{d}^\top \nabla f(\mathbf{x}^*) = 0$ for all $\mathbf{d}$ satisfying a “certain condition.” Specify what this certain condition is.

Solution:

- a. There are no feasible directions at any $\mathbf{x}^*$.
- b. Because of part a, *all* points in Ω satisfy the FONC for this set-constrained problem.
- c. No, the FONC for this set-constrained problem is not useful for eliminating local-minimizer candidates.
- d. Write $h(\theta) = f(g(\theta))$, where $g : \mathbb{R} \rightarrow \mathbb{R}^2$ is given by the equations relating θ to $\mathbf{x} = [x_1, x_2]^\top$. Note that $Dg(\theta) = [-\sin \theta, \cos \theta]^\top$. Hence, by the chain rule,

$$h'(\theta) = Df(g(\theta))Dg(\theta) = Dg(\theta)^\top \nabla f(g(\theta)).$$

Notice that $Dg(\theta)$ is tangent to Ω at $\mathbf{x} = g(\theta)$. Alternatively, we could say that $Dg(\theta)$ is orthogonal to $\mathbf{x} = g(\theta)$.

Suppose that $\mathbf{x}^* \in \Omega$ is a local minimizer. Write $\mathbf{x}^* = g(\theta^*)$. Then θ^* is an unconstrained minimizer of h . By the FONC for unconstrained problems, $h'(\theta^*) = 0$, which implies that $\mathbf{d}^\top \nabla f(\mathbf{x}^*) = 0$ for all $\mathbf{d}$ tangent to Ω at $\mathbf{x}^*$ (or, alternatively, for all $\mathbf{d}$ orthogonal to $\mathbf{x}^*$). ■

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 , $\mathbf{x}^*$ a local minimizer of f over Ω , and $\mathbf{d}$ a feasible direction at $\mathbf{x}^*$. If $\mathbf{d}^\top \nabla f(\mathbf{x}^*) = 0$, then

$$\mathbf{d}^\top \mathbf{F}(\mathbf{x}^*) \mathbf{d} \geq 0,$$

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

Proof. We prove the result by contradiction. Suppose that there is a feasible direction $\mathbf{d}$ at $\mathbf{x}^*$ such that $\mathbf{d}^\top \nabla f(\mathbf{x}^*) = 0$ and $\mathbf{d}^\top \mathbf{F}(\mathbf{x}^*) \mathbf{d} < 0$. Let $\mathbf{x}(\alpha) = \mathbf{x}^* + \alpha \mathbf{d}$ and define the composite function $\phi(\alpha) = f(\mathbf{x}^* + \alpha \mathbf{d}) = f(\mathbf{x}(\alpha))$. Then, by Taylor's theorem,

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

where by assumption, $\phi'(0) = \mathbf{d}^\top \nabla f(\mathbf{x}^*) = 0$ and $\phi''(0) = \mathbf{d}^\top \mathbf{F}(\mathbf{x}^*) \mathbf{d} < 0$. For sufficiently small α ,

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

that is,

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

which contradicts the assumption that $\mathbf{x}^*$ is a local minimizer. Thus,

$$\phi''(0) = \mathbf{d}^\top \mathbf{F}(\mathbf{x}^*) \mathbf{d} \geq 0.$$

■

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

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

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

$$\mathbf{d}^\top \mathbf{F}(\mathbf{x}^*) \mathbf{d} \geq 0.$$

□

Proof. If $\mathbf{x}^*$ 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 6.6 Consider a function of one variable $f(x) = x^3$, $f : \mathbb{R} \rightarrow \mathbb{R}$. 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). ■

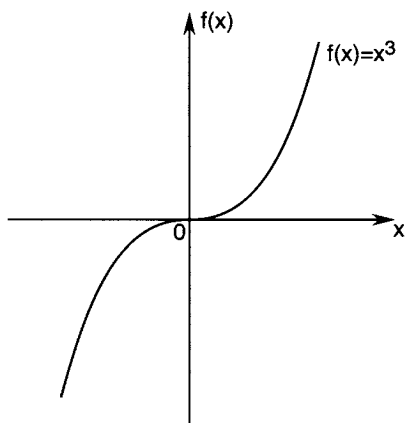


Figure 6.6 The point 0 satisfies the FONC and SONC but is not a minimizer.

Example 6.7 Consider a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, where $f(\mathbf{x}) = x_1^2 - x_2^2$. The FONC requires that $\nabla f(\mathbf{x}) = [2x_1, -2x_2]^\top = \mathbf{0}$. Thus, $\mathbf{x} = [0, 0]^\top$ satisfies the FONC. The Hessian matrix of f is

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

The Hessian matrix is indefinite; that is, for some $\mathbf{d}_1 \in \mathbb{R}^2$ we have $\mathbf{d}_1^\top \mathbf{F} \mathbf{d}_1 > 0$ (e.g., $\mathbf{d}_1 = [1, 0]^\top$) and for some $\mathbf{d}_2$ we have $\mathbf{d}_2^\top \mathbf{F} \mathbf{d}_2 < 0$ (e.g., $\mathbf{d}_2 = [0, 1]^\top$). Thus, $\mathbf{x} = [0, 0]^\top$ does not satisfy the SONC, and hence it is not a minimizer. The graph of $f(\mathbf{x}) = x_1^2 - x_2^2$ is shown in Figure 6.7. ■

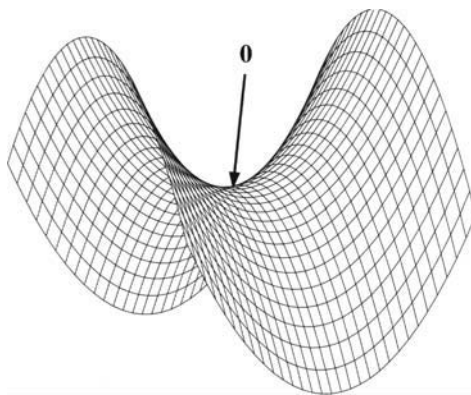


Figure 6.7 Graph of $f(\mathbf{x}) = x_1^2 - x_2^2$. The point 0 satisfies the FONC but not SONC; this point is not a minimizer.

We now derive sufficient conditions that imply that $\mathbf{x}^*$ is a local minimizer.

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

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

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

Proof. Because $f \in \mathcal{C}^2$, we have $\mathbf{F}(\mathbf{x}^*) = \mathbf{F}^\top(\mathbf{x}^*)$. Using assumption 2 and Rayleigh's inequality it follows that if $\mathbf{d} \neq \mathbf{0}$, then $0 < \lambda_{\min}(\mathbf{F}(\mathbf{x}^*))\|\mathbf{d}\|^2 \leq \mathbf{d}^\top \mathbf{F}(\mathbf{x}^*)\mathbf{d}$. By Taylor's theorem and assumption 1,

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

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

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

which completes the proof. ■

Example 6.8 Let $f(\mathbf{x}) = x_1^2 + x_2^2$. We have $\nabla f(\mathbf{x}) = [2x_1, 2x_2]^\top = \mathbf{0}$ if and only if $\mathbf{x} = [0, 0]^\top$. For all $\mathbf{x} \in \mathbb{R}^2$, we have

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

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

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 that $f \in \mathcal{C}^2$) 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.

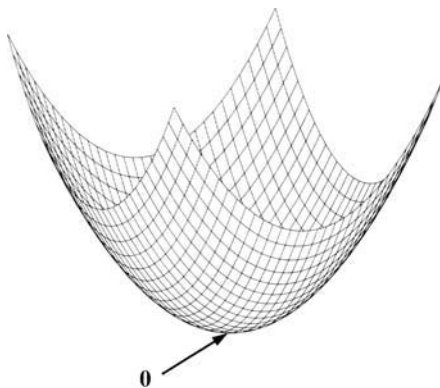


Figure 6.8 Graph of $f(\mathbf{x}) = x_1^2 + x_2^2$.

EXERCISES

6.1 Consider the problem

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

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

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

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

6.2 Find minimizers and maximizers of the function

$$f(x_1, x_2) = \frac{1}{3}x_1^3 - 4x_1 + \frac{1}{3}x_2^3 - 16x_2.$$

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

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

6.5 Consider the problem of minimizing $f : \mathbb{R} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^3$, over the constraint set Ω . Suppose that 0 is an *interior point* of Ω .

- Suppose that 0 is a local minimizer. By the FONC we know that $f'(0) = 0$ (where f' is the first derivative of f). By the SONC we know that $f''(0) \geq 0$ (where f'' is the second derivative of f). State and prove a *third-order necessary condition* (TONC) involving the third derivative at 0, $f'''(0)$.
- Give an example of f such that the FONC, SONC, and TONC (in part a) hold at the interior point 0, but 0 is not a local minimizer of f over Ω . (Show that your example is correct.)
- Suppose that f is a third-order polynomial. If 0 satisfies the FONC, SONC, and TONC (in part a), then is this *sufficient* for 0 to be a local minimizer?

6.6 Consider the problem of minimizing $f : \mathbb{R} \rightarrow \mathbb{R}$, $f \in \mathcal{C}^3$, over the constraint set $\Omega = [0, 1]$. Suppose that $\mathbf{x}^* = 0$ is a local minimizer.

- By the FONC we know that $f'(0) \geq 0$ (where f' is the first derivative of f). By the SONC we know that if $f'(0) = 0$, then $f''(0) \geq 0$ (where f'' is the second derivative of f). State and prove a *third-order necessary condition* involving the third derivative at 0, $f'''(0)$.
- Give an example of f such that the FONC, SONC, and TONC (in part a) hold at the point 0, but 0 is not a local minimizer of f over $\Omega = [0, 1]$.

6.7 Let $f : \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathbf{x}_0 \in \mathbb{R}^n$, and $\Omega \subset \mathbb{R}^n$. Show that

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

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

6.8 Consider the following function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$:

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

- Find the gradient and Hessian of f at the point $[1, 1]^\top$.
- Find the directional derivative of f at $[1, 1]^\top$ with respect to a unit vector in the direction of maximal rate of increase.
- Find a point that satisfies the FONC (interior case) for f . Does this point satisfy the SONC (for a minimizer)?

6.9 Consider the following function:

$$f(x_1, x_2) = x_1^2 x_2 + x_2^3 x_1.$$

- In what direction does the function f *decrease* most rapidly at the point $\mathbf{x}^{(0)} = [2, 1]^\top$?
- What is the rate of increase of f at the point $\mathbf{x}^{(0)}$ in the direction of maximum decrease of f ?
- Find the rate of increase of f at the point $\mathbf{x}^{(0)}$ in the direction $\mathbf{d} = [3, 4]^\top$.

6.10 Consider the following function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$:

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

- Find the directional derivative of f at $[0, 1]^\top$ in the direction $[1, 0]^\top$.
- 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.

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

- Does the point $[x_1, x_2]^\top = \mathbf{0}$ satisfy the first-order necessary condition for a minimizer? That is, if f is the objective function, is it true that $\mathbf{d}^\top \nabla f(\mathbf{0}) \geq 0$ for all feasible directions $\mathbf{d}$ at $\mathbf{0}$?
- Is the point $[x_1, x_2]^\top = \mathbf{0}$ a local minimizer, a strict local minimizer, a local maximizer, a strict local maximizer, or none of the above?

6.12 Consider the problem

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

where $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by $f(\mathbf{x}) = 5x_2$ with $\mathbf{x} = [x_1, x_2]^\top$, and $\Omega = \{\mathbf{x} = [x_1, x_2]^\top : x_1^2 + x_2 \geq 1\}$.

- Does the point $\mathbf{x}^* = [0, 1]^\top$ satisfy the first-order necessary condition?
- Does the point $\mathbf{x}^* = [0, 1]^\top$ satisfy the second-order necessary condition?
- Is the point $\mathbf{x}^* = [0, 1]^\top$ a local minimizer?

6.13 Consider the problem

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

where $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by $f(\mathbf{x}) = -3x_1$ with $\mathbf{x} = [x_1, x_2]^\top$, and $\Omega = \{\mathbf{x} = [x_1, x_2]^\top : x_1 + x_2^2 \leq 2\}$. Answer each of the following questions, showing complete justification.

- Does the point $\mathbf{x}^* = [2, 0]^\top$ satisfy the first-order necessary condition?
- Does the point $\mathbf{x}^* = [2, 0]^\top$ satisfy the second-order necessary condition?
- Is the point $\mathbf{x}^* = [2, 0]^\top$ a local minimizer?

6.14 Consider the problem

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

where $\Omega = \{\mathbf{x} \in \mathbb{R}^2 : x_1^2 + x_2^2 \geq 1\}$ and $f(\mathbf{x}) = x_2$.

- Find all point(s) satisfying the FONC.
- Which of the point(s) in part a satisfy the SONC?
- Which of the point(s) in part a are local minimizers?

6.15 Consider the problem

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

where $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by $f(\mathbf{x}) = 3x_1$ with $\mathbf{x} = [x_1, x_2]^\top$, and $\Omega = \{\mathbf{x} = [x_1, x_2]^\top : x_1 + x_2^2 \geq 2\}$. Answer each of the following questions, showing complete justification.

- Does the point $\mathbf{x}^* = [2, 0]^\top$ satisfy the first-order necessary condition?
 - Does the point $\mathbf{x}^* = [2, 0]^\top$ satisfy the second-order necessary condition?
 - Is the point $\mathbf{x}^* = [2, 0]^\top$ a local minimizer?
- Hint:* Draw a picture with the constraint set and level sets of f .

6.16 Consider the problem

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

where $\mathbf{x} = [x_1, x_2]^\top$, $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ is given by $f(\mathbf{x}) = 4x_1^2 - x_2^2$, and $\Omega = \{\mathbf{x} : x_1^2 + 2x_1 - x_2 \geq 0, x_1 \geq 0, x_2 \geq 0\}$.

- Does the point $\mathbf{x}^* = \mathbf{0} = [0, 0]^\top$ satisfy the first-order necessary condition?
- Does the point $\mathbf{x}^* = \mathbf{0}$ satisfy the second-order necessary condition?
- Is the point $\mathbf{x}^* = \mathbf{0}$ a local minimizer of the given problem?

6.17 Consider the problem

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

where $\Omega \subset \{\mathbf{x} \in \mathbb{R}^2 : x_1 > 0, x_2 > 0\}$ and $f: \Omega \rightarrow \mathbb{R}$ is given by $f(\mathbf{x}) = \log(x_1) + \log(x_2)$ with $\mathbf{x} = [x_1, x_2]^\top$, where “log” represents natural logarithm. Suppose that $\mathbf{x}^*$ is an optimal solution. Answer each of the following questions, showing complete justification.

- Is it possible that $\mathbf{x}^*$ is an interior point of Ω ?
- At what point(s) (if any) is the second-order necessary condition satisfied?

6.18 Suppose that we are given n real numbers, $x_1, \dots, x_n$. Find the number $\bar{x} \in \mathbb{R}$ such that the sum of the squared difference between $\bar{x}$ and the numbers above is minimized (assuming that the solution $\bar{x}$ exists).

6.19 An art collector stands at a distance of x feet from the wall, where a piece of art (picture) of height a feet is hung, b feet above his eyes, as shown in

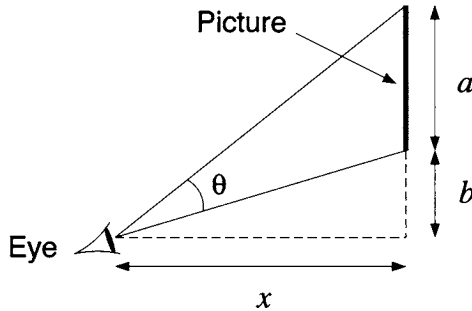


Figure 6.9 Art collector's eye position in Exercise 6.19.

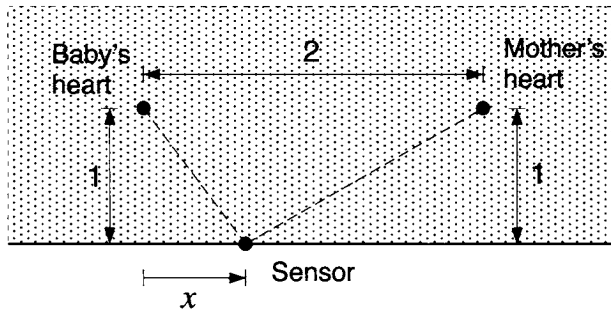


Figure 6.10 Simplified fetal heart monitoring system for Exercise 6.20.

Figure 6.9. Find the distance from the wall for which the angle θ subtended by the eye to the picture is maximized.

Hint: (1) Maximizing θ is equivalent to maximizing $\tan(\theta)$.

(2) If $\theta = \theta_2 - \theta_1$, then $\tan(\theta) = (\tan(\theta_2) - \tan(\theta_1)) / (1 + \tan(\theta_2)\tan(\theta_1))$.

6.20 Figure 6.10 shows a simplified model of a fetal heart monitoring system (the distances shown have been scaled down to make the calculations simpler). A heartbeat sensor is located at position x (see Figure 6.10).

The energy of the heartbeat signal measured by the sensor is the reciprocal of the squared distance from the source (baby's heart or mother's heart). Find the position of the sensor that maximizes the *signal-to-interference ratio*, which is the ratio of the signal energy from the baby's heart to the signal energy from the mother's heart.

6.21 An amphibian vehicle needs to travel from point A (on land) to point B (in water), as illustrated in Figure 6.11. The speeds at which the vehicle travels on land and water are v_1 and v_2 , respectively.

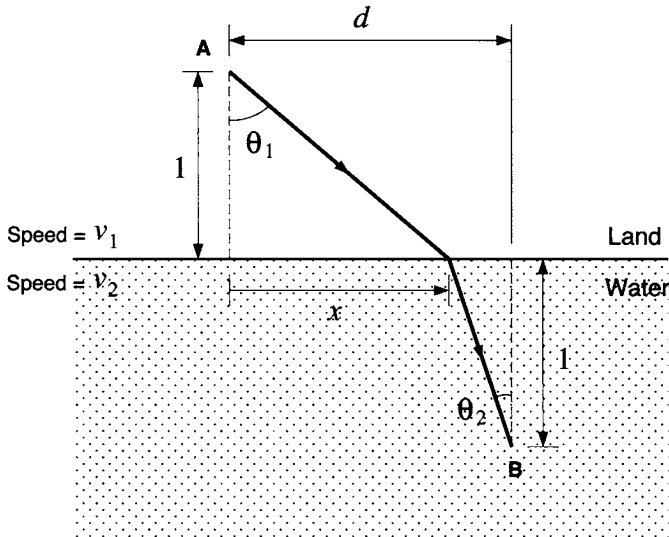


Figure 6.11 Path of amphibian vehicle in Exercise 6.21.

- a. Suppose that the vehicle traverses a path that minimizes the total time taken to travel from A to B. Use the first-order necessary condition to show that for the optimal path above, the angles θ_1 and θ_2 in Figure 6.11 satisfy Snell's law:

$$\frac{\sin \theta_1}{\sin \theta_2} = \frac{v_1}{v_2}.$$

- b. Does the minimizer for the problem in part a satisfy the second-order sufficient condition?

6.22 Suppose that you have a piece of land to sell and you have two buyers. If the first buyer receives a fraction x_1 of the piece of land, the buyer will pay you $U_1(x_1)$ dollars. Similarly, the second buyer will pay you $U_2(x_2)$ dollars for a fraction of x_2 of the land. Your goal is to sell parts of your land to the two buyers so that you maximize the total dollars you receive. (Other than the constraint that you can only sell whatever land you own, there are no restrictions on how much land you can sell to each buyer.)

- a. Formulate the problem as an optimization problem of the kind

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

by specifying f and Ω . Draw a picture of the constraint set.

- b. Suppose that $U_i(x_i) = a_i x_i$, $i = 1, 2$, where a_1 and a_2 are given positive constants such that $a_1 > a_2$. Find all feasible points that satisfy the first-order necessary condition, giving full justification.
- c. Among those points in the answer of part b, find all that also satisfy the second-order necessary condition.

6.23 Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ be defined by

$$f(\mathbf{x}) = (x_1 - x_2)^4 + x_1^2 - x_2^2 - 2x_1 + 2x_2 + 1,$$

where $\mathbf{x} = [x_1, x_2]^\top$. Suppose that we wish to minimize f over $\mathbb{R}^2$. Find all points satisfying the FONC. Do these points satisfy the SONC?

6.24 Show that if $\mathbf{d}$ is a feasible direction at a point $\mathbf{x} \in \Omega$, then for all $\beta > 0$, the vector $\beta \mathbf{d}$ is also a feasible direction at $\mathbf{x}$.

6.25 Let $\Omega = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{A}\mathbf{x} = \mathbf{b}\}$. Show that $\mathbf{d} \in \mathbb{R}^n$ is a feasible direction at $\mathbf{x} \in \Omega$ if and only if $\mathbf{A}\mathbf{d} = \mathbf{0}$.

6.26 Let $f : \mathbb{R}^2 \rightarrow \mathbb{R}$. Consider the problem

$$\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & x_1, x_2 \geq 0, \end{array}$$

where $\mathbf{x} = [x_1, x_2]^\top$. Suppose that $\nabla f(\mathbf{0}) \neq \mathbf{0}$, and

$$\frac{\partial f}{\partial x_1}(\mathbf{0}) \leq 0, \quad \frac{\partial f}{\partial x_2}(\mathbf{0}) \leq 0.$$

Show that $\mathbf{0}$ cannot be a minimizer for this problem.

6.27 Let $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{c} \neq \mathbf{0}$, and consider the problem of minimizing the function $f(\mathbf{x}) = \mathbf{c}^\top \mathbf{x}$ over a constraint set $\Omega \subset \mathbb{R}^n$. Show that we cannot have a solution lying in the interior of Ω .

6.28 Consider the problem

$$\begin{array}{ll} \text{maximize} & c_1 x_1 + c_2 x_2 \\ \text{subject to} & x_1 + x_2 \leq 1 \\ & x_1, x_2 \geq 0, \end{array}$$

where c_1 and c_2 are constants such that $c_1 > c_2 \geq 0$. This is a *linear programming* problem (see Part III). Assuming that the problem has an optimal feasible solution, use the first-order necessary condition to show that the *unique* optimal feasible solution $\mathbf{x}^*$ is $[1, 0]^\top$.

Hint: First show that $\mathbf{x}^*$ cannot lie in the interior of the constraint set. Then, show that $\mathbf{x}^*$ cannot lie on the line segments $L_1 = \{\mathbf{x} : x_1 = 0, 0 \leq x_2 < 1\}$, $L_2 = \{\mathbf{x} : 0 \leq x_1 < 1, x_2 = 0\}$, $L_3 = \{\mathbf{x} : 0 \leq x_1 < 1, x_2 = 1 - x_1\}$.

6.29 Line Fitting. Let $[x_1, y_1]^\top, \dots, [x_n, y_n]^\top$, $n \geq 2$, be points on the $\mathbb{R}^2$ plane (each $x_i, y_i \in \mathbb{R}$). We wish to find the straight line of “best fit” through these points (“best” in the sense that the average squared error is minimized); that is, we wish to find $a, b \in \mathbb{R}$ to minimize

$$f(a, b) = \frac{1}{n} \sum_{i=1}^n (ax_i + b - y_i)^2.$$

a. Let

$$\begin{aligned}\bar{X} &= \frac{1}{n} \sum_{i=1}^n x_i, \\ \bar{Y} &= \frac{1}{n} \sum_{i=1}^n y_i, \\ \overline{X^2} &= \frac{1}{n} \sum_{i=1}^n x_i^2, \\ \overline{Y^2} &= \frac{1}{n} \sum_{i=1}^n y_i^2, \\ \overline{XY} &= \frac{1}{n} \sum_{i=1}^n x_i y_i.\end{aligned}$$

Show that $f(a, b)$ can be written in the form $\mathbf{z}^\top \mathbf{Q} \mathbf{z} - 2\mathbf{c}^\top \mathbf{z} + d$, where $\mathbf{z} = [a, b]^\top$, $\mathbf{Q} = \mathbf{Q}^\top \in \mathbb{R}^{2 \times 2}$, $\mathbf{c} \in \mathbb{R}^2$ and $d \in \mathbb{R}$, and find expressions for $\mathbf{Q}$, $\mathbf{c}$, and d in terms of $\bar{X}$, $\bar{Y}$, $\overline{X^2}$, $\overline{Y^2}$, and $\overline{XY}$.

b. Assume that the x_i , $i = 1, \dots, n$, are not all equal. Find the parameters a^* and b^* for the line of best fit in terms of $\bar{X}$, $\bar{Y}$, $\overline{X^2}$, $\overline{Y^2}$, and $\overline{XY}$. Show that the point $[a^*, b^*]^\top$ is the only local minimizer of f .

Hint: $\overline{X^2} - (\bar{X})^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2$.

c. Show that if a^* and b^* are the parameters of the line of best fit, then $\bar{Y} = a^* \bar{X} + b^*$ (and hence once we have computed a^* , we can compute b^* using the formula $b^* = \bar{Y} - a^* \bar{X}$).

6.30 Suppose that we are given a set of vectors $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(p)}\}$, $\mathbf{x}^{(i)} \in \mathbb{R}^n$, $i = 1, \dots, p$. Find the vector $\bar{\mathbf{x}} \in \mathbb{R}^n$ such that the average squared distance (norm) between $\bar{\mathbf{x}}$ and $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(p)}$,

$$\frac{1}{p} \sum_{i=1}^p \|\bar{\mathbf{x}} - \mathbf{x}^{(i)}\|^2,$$

is minimized. Use the SOSC to prove that the vector $\bar{\mathbf{x}}$ found above is a strict local minimizer. How is $\bar{\mathbf{x}}$ related to the centroid (or center of gravity) of the given set of points $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(p)}\}$?

6.31 Consider a function $f : \Omega \rightarrow \mathbb{R}$, where $\Omega \subset \mathbb{R}^n$ is a convex set and $f \in \mathcal{C}^1$. Given $\mathbf{x}^* \in \Omega$, suppose that there exists $c > 0$ such that $\mathbf{d}^\top \nabla f(\mathbf{x}^*) \geq c\|\mathbf{d}\|$ for all feasible directions $\mathbf{d}$ at $\mathbf{x}^*$. Show that $\mathbf{x}^*$ is a strict local minimizer of f over Ω .

6.32 Prove the following generalization of the second-order sufficient condition:

Theorem: Let Ω be a convex subset of $\mathbb{R}^n$, $f \in \mathcal{C}^2$ a real-valued function on Ω , and $\mathbf{x}^*$ a point in Ω . Suppose that there exists $c \in \mathbb{R}$, $c > 0$, such that for all feasible directions $\mathbf{d}$ at $\mathbf{x}^*$ ($\mathbf{d} \neq \mathbf{0}$), the following hold:

1. $\mathbf{d}^\top \nabla f(\mathbf{x}^*) \geq 0$.
2. $\mathbf{d}^\top \mathbf{F}(\mathbf{x}^*)\mathbf{d} \geq c\|\mathbf{d}\|^2$.

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

6.33 Consider the quadratic function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{Q}\mathbf{x} - \mathbf{x}^\top \mathbf{b},$$

where $\mathbf{Q} = \mathbf{Q}^\top > 0$. Show that $\mathbf{x}^*$ minimizes f if and only if $\mathbf{x}^*$ satisfies the FONC.

6.34 Consider the linear system $x_{k+1} = ax_k + bu_{k+1}$, $k \geq 0$, where $x_i \in \mathbb{R}$, $u_i \in \mathbb{R}$, and the initial condition is $x_0 = 0$. Find the values of the control inputs $u_1, \dots, u_n$ to minimize

$$-qx_n + r \sum_{i=1}^n u_i^2,$$

where $q, r > 0$ are given constants. This can be interpreted as desiring to make x_n as large as possible but at the same time desiring to make the total input energy $\sum_{i=1}^n u_i^2$ as small as possible. The constants q and r reflect the relative weights of these two objectives.