Chapter 2

BASIC PRINCIPLES

2.1 Introduction

Nonlinear programming is based on a collection of definitions, theorems, and principles that must be clearly understood if the available nonlinear programming methods are to be used effectively.

This chapter begins with the definition of the gradient vector, the Hessian matrix, and the various types of extrema (maxima and minima). The conditions that must hold at the solution point are then discussed and techniques for the characterization of the extrema are described. Subsequently, the classes of convex and concave functions are introduced. These provide a natural formulation for the theory of global convergence.

Throughout the chapter, we focus our attention on the nonlinear optimization problem

$$\begin{aligned}&\text{minimize } f = f(\mathbf{x})\\&\text{subject to:} \quad \mathbf{x} \in \mathcal{R}\end{aligned}$$

where $f(\mathbf{x})$ is a real-valued function and $\mathcal{R} \subset E^n$ is the feasible region.

2.2 Gradient Information

In many optimization methods, gradient information pertaining to the objective function is required. This information consists of the first and second derivatives of $f(\mathbf{x})$ with respect to the n variables.

If $f(\mathbf{x}) \in C^1$, that is, if $f(\mathbf{x})$ has continuous first-order partial derivatives, the *gradient* of $f(\mathbf{x})$ is defined as

$$\begin{aligned}\mathbf{g}(\mathbf{x}) &= \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \cdots \quad \frac{\partial f}{\partial x_n} \right]^T\\&= \nabla f(\mathbf{x})\end{aligned} \tag{2.1}$$

where

$$\nabla = [\, \frac{\partial}{\partial x_1} \quad \frac{\partial}{\partial x_2} \quad \cdots \quad \frac{\partial}{\partial x_n} \,]^T \tag{2.2}$$

If $f(\mathbf{x}) \in C^2$, that is, if $f(\mathbf{x})$ has continuous second-order partial derivatives, the *Hessian*[1] of $f(\mathbf{x})$ is defined as

$$\mathbf{H}(\mathbf{x}) = \nabla \mathbf{g}^T = \nabla \{\nabla^T f(\mathbf{x})\} \tag{2.3}$$

Hence Eqs. (2.1) – (2.3) give

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

For a function $f(\mathbf{x}) \in C^2$

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$$

since differentiation is a linear operation and hence $\mathbf{H}(\mathbf{x})$ is an $n \times n$ square symmetric matrix.

The gradient and Hessian at a point $\mathbf{x} = \mathbf{x}_k$ are represented by $\mathbf{g}(\mathbf{x}_k)$ and $\mathbf{H}(\mathbf{x}_k)$ or by the simplified notation $\mathbf{g}_k$ and $\mathbf{H}_k$, respectively. Sometimes, when confusion is not likely to arise, $\mathbf{g}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$ are simplified to $\mathbf{g}$ and $\mathbf{H}$.

The gradient and Hessian tend to simplify the optimization process considerably. Nevertheless, in certain applications it may be uneconomic, time-consuming, or impossible to deduce and compute the partial derivatives of $f(\mathbf{x})$. For these applications, methods are preferred that do not require gradient information.

Gradient methods, namely, methods based on gradient information may use only $\mathbf{g}(\mathbf{x})$ or both $\mathbf{g}(\mathbf{x})$ and $\mathbf{H}(\mathbf{x})$. In the latter case, the inversion of matrix $\mathbf{H}(\mathbf{x})$ may be required which tends to introduce numerical inaccuracies and is time-consuming. Such methods are often avoided.

2.3 The Taylor Series

Some of the nonlinear programming procedures and methods utilize linear or quadratic approximations for the objective function and the equality and inequality constraints, namely, $f(\mathbf{x})$, $a_i(\mathbf{x})$, and $c_j(\mathbf{x})$ in Eq. (1.4). Such

[1]For the sake of simplicity, the gradient vector and Hessian matrix will be referred to as the gradient and Hessian, respectively, henceforth.

approximations can be obtained by using the Taylor series. If $f(\mathbf{x})$ is a function of two variables x_1 and x_2 such that $f(\mathbf{x}) \in C^P$ where $P \to \infty$, that is, $f(\mathbf{x})$ has continuous partial derivatives of all orders, then the value of function $f(\mathbf{x})$ at point $[x_1 + \delta_1,\ \ x_2 + \delta_2]$ is given by the Taylor series as

$$\begin{aligned} f(x_1 + \delta_1,\ x_2 + \delta_2) = {} & f(x_1,\ x_2) + \frac{\partial f}{\partial x_1}\delta_1 + \frac{\partial f}{\partial x_2}\delta_2 \\ & + \frac{1}{2}\left(\frac{\partial^2 f}{\partial x_1^2}\,\delta_1^2 + \frac{2\partial^2 f}{\partial x_1 \partial x_2}\,\delta_1\delta_2 + \frac{\partial^2 f}{\partial x_2^2}\,\delta_2^2\right) \\ & + O(\|\boldsymbol{\delta}\|^3) \end{aligned} \tag{2.4a}$$

where

$$\boldsymbol{\delta} = [\delta_1\ \delta_2]^T$$

$O(\|\boldsymbol{\delta}\|^3)$ is the *remainder*, and $\|\boldsymbol{\delta}\|$ is the Euclidean norm of $\boldsymbol{\delta}$ given by

$$\|\boldsymbol{\delta}\| = \sqrt{\boldsymbol{\delta}^T\boldsymbol{\delta}}$$

The notation $\phi(x) = O(x)$ *denotes that* $\phi(x)$ *approaches zero at least as fast as* x as x approaches zero, that is, there exists a constant $K \geq 0$ such that

$$\left|\frac{\phi(x)}{x}\right| \leq K \quad \text{as } x \to 0$$

The remainder term in Eq. (2.4a) can also be expressed as $o(\|\boldsymbol{\delta}\|^2)$ where the notation $\phi(x) = o(x)$ *denotes that* $\phi(x)$ *approaches zero faster than* x *as* x *approaches zero*, that is,

$$\left|\frac{\phi(x)}{x}\right| \to 0 \quad \text{as } x \to 0$$

If $f(\mathbf{x})$ is a function of n variables, then the Taylor series of $f(\mathbf{x})$ at point $[x_1 + \delta_1,\ x_2 + \delta_2,\ \ldots]$ is given by

$$\begin{aligned} f(x_1 + \delta_1,\ x_2 + \delta_2,\ \ldots) = {} & f(x_1,\ x_2,\ \ldots) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\,\delta_i \\ & + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \delta_i\,\frac{\partial^2 f}{\partial x_i \partial x_j}\,\delta_j \\ & + o(\|\boldsymbol{\delta}\|^2) \end{aligned} \tag{2.4b}$$

Alternatively, on using matrix notation

$$f(\mathbf{x} + \boldsymbol{\delta}) = f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T\boldsymbol{\delta} + \tfrac{1}{2}\boldsymbol{\delta}^T\mathbf{H}(\mathbf{x})\boldsymbol{\delta} + o(\|\boldsymbol{\delta}\|^2) \tag{2.4c}$$

where $\mathbf{g}(\mathbf{x})$ is the gradient, and $\mathbf{H}(\mathbf{x})$ is the Hessian at point $\mathbf{x}$.

As $\|\boldsymbol{\delta}\| \to 0$, second- and higher-order terms can be neglected and a *linear approximation* can be obtained for $f(\mathbf{x} + \boldsymbol{\delta})$ as

$$f(\mathbf{x} + \boldsymbol{\delta}) \approx f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T \boldsymbol{\delta} \tag{2.4d}$$

Similarly, a *quadratic approximation* for $f(\mathbf{x} + \boldsymbol{\delta})$ can be obtained as

$$f(\mathbf{x} + \boldsymbol{\delta}) \approx f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T \boldsymbol{\delta} + \tfrac{1}{2}\boldsymbol{\delta}^T \mathbf{H}(\mathbf{x})\boldsymbol{\delta} \tag{2.4e}$$

Another form of the Taylor series, which includes an expression for the remainder term, is

$$\begin{aligned} f(\mathbf{x} + \boldsymbol{\delta}) &= f(\mathbf{x}) \\ &\quad + \sum_{1 \le k_1 + k_2 + \cdots + k_n \le P} \frac{\partial^{k_1 + k_2 + \cdots + k_n} f(\mathbf{x})}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_n^{k_n}} \prod_{i=1}^{n} \frac{\delta_i^{k_i}}{k_i!} \\ &\quad + \sum_{k_1 + k_2 + \cdots + k_n = P+1} \frac{\partial^{P+1} f(\mathbf{x} + \alpha\boldsymbol{\delta})}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_n^{k_n}} \prod_{i=1}^{n} \frac{\delta_i^{k_i}}{k_i!} \end{aligned} \tag{2.4f}$$

where $0 \le \alpha \le 1$ and

$$\sum_{1 \le k_1 + k_2 + \cdots + k_n \le P} \frac{\partial^{k_1 + k_2 + \cdots + k_n} f(\mathbf{x})}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_n^{k_n}} \prod_{i=1}^{n} \frac{\delta_i^{k_i}}{k_i!}$$

is the sum of terms taken over all possible combinations of $k_1,\ k_2,\ \ldots,\ k_n$ that add up to a number in the range 1 to P. (See Chap. 4 of Protter and Morrey [1] for proof.) This representation of the Taylor series is completely general and, therefore, it can be used to obtain cubic and higher-order approximations for $f(\mathbf{x} + \boldsymbol{\delta})$. Furthermore, it can be used to obtain linear, quadratic, cubic, and higher-order *exact closed-form* expressions for $f(\mathbf{x} + \boldsymbol{\delta})$. If $f(\mathbf{x}) \in C^1$ and $P = 0$, Eq. (2.4f) gives

$$f(\mathbf{x} + \boldsymbol{\delta}) = f(\mathbf{x}) + \mathbf{g}(\mathbf{x} + \alpha\boldsymbol{\delta})^T \boldsymbol{\delta} \tag{2.4g}$$

and if $f(\mathbf{x}) \in C^2$ and $P = 1$, then

$$f(\mathbf{x} + \boldsymbol{\delta}) = f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T \boldsymbol{\delta} + \tfrac{1}{2}\boldsymbol{\delta}^T \mathbf{H}(\mathbf{x} + \alpha\boldsymbol{\delta})\boldsymbol{\delta} \tag{2.4h}$$

where $0 \le \alpha \le 1$. Eq. (2.4g) is usually referred to as the *mean-value theorem for differentiation*.

Yet another form of the Taylor series can be obtained by regrouping the terms in Eq. (2.4f) as

$$\begin{aligned} f(\mathbf{x} + \boldsymbol{\delta}) &= f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T \boldsymbol{\delta} + \tfrac{1}{2}\boldsymbol{\delta}^T \mathbf{H}(\mathbf{x})\boldsymbol{\delta} + \tfrac{1}{3!} D^3 f(\mathbf{x}) \\ &\quad + \cdots + \frac{1}{(r-1)!} D^{r-1} f(\mathbf{x}) + \cdots \end{aligned} \tag{2.4i}$$

where

$$D^r f(\mathbf{x}) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_r=1}^{n} \left\{ \delta_{i_1} \delta_{i_2} \cdots \delta_{i_r} \frac{\partial^r f(\mathbf{x})}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_r}} \right\}$$

2.4 Types of Extrema

The *extrema* of a function are its minima and maxima. Points at which a function has minima (maxima) are said to be *minimizers* (*maximizers*). Several types of minimizers (maximizers) can be distinguished, namely, local or global and weak or strong.

Definition 2.1 A point $\mathbf{x}^* \in \mathcal{R}$, where $\mathcal{R}$ is the feasible region, is said to be a *weak local minimizer* of $f(\mathbf{x})$ if there exists a distance $\varepsilon > 0$ such that

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) \tag{2.5}$$

if

$$\mathbf{x} \in \mathcal{R} \quad \text{and} \quad \|\mathbf{x} - \mathbf{x}^*\| < \varepsilon$$

■

Definition 2.2 A point $\mathbf{x}^* \in \mathcal{R}$ is said to be a *weak global minimizer* of $f(\mathbf{x})$ if

$$f(\mathbf{x}) \geq f(\mathbf{x}^*) \tag{2.6}$$

for all $\mathbf{x} \in \mathcal{R}$.

■

If Def. 2.2 is satisfied at $\mathbf{x}^*$, then Def. 2.1 is also satisfied at $\mathbf{x}^*$, and so a global minimizer is also a local minimizer.

Definition 2.3
If Eq. (2.5) in Def. 2.1 or Eq. (2.6) in Def. 2.2 is replaced by

$$f(\mathbf{x}) > f(\mathbf{x}^*) \tag{2.7}$$

$\mathbf{x}^*$ is said to be a *strong local* (or *global*) *minimizer*.

■

The minimum at a weak local, weak global, etc. minimizer is called a weak local, weak global, etc. minimum.

A strong global minimum in E^2 is depicted in Fig. 2.1.

Weak or strong and local or global maximizers can similarly be defined by reversing the inequalities in Eqs. (2.5) – (2.7).

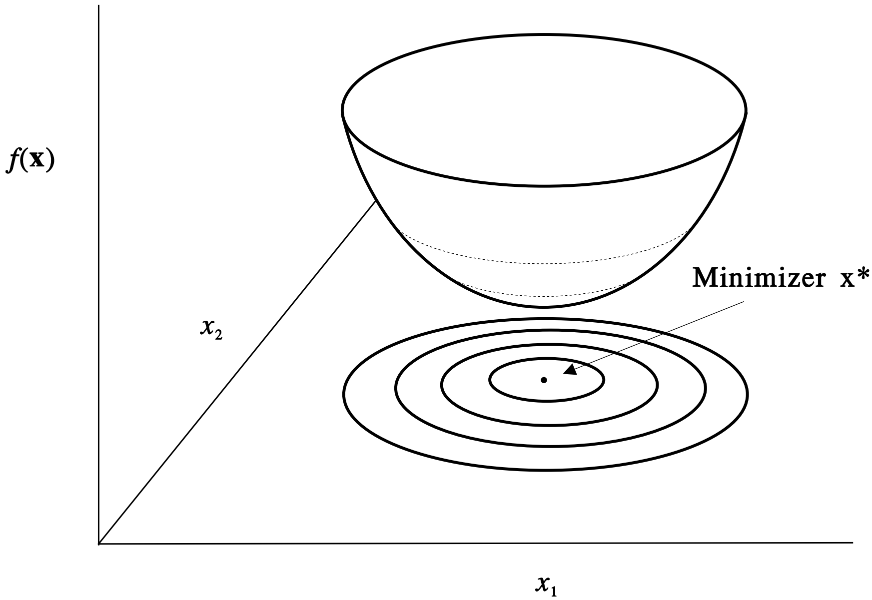

Figure 2.1. A strong global minimizer.

Example 2.1 The function of Fig. 2.2 has a feasible region defined by the set

$$\mathcal{R} = \{x : \ x_1 \leq x \leq x_2\}$$

Classify its minimizers.

Solution The function has a weak local minimum at point B, strong local minima at points A, C, and D, and a strong global minimum at point C.

■

In the general optimization problem, we are in principle seeking the global minimum (or maximum) of $f(\mathbf{x})$. In practice, an optimization problem may have two or more local minima. Since optimization algorithms in general are iterative procedures which start with an initial estimate of the solution and converge to a single solution, one or more local minima may be missed. If the global minimum is missed, a suboptimal solution will be achieved, which may or may not be acceptable. This problem can to some extent be overcome by performing the optimization several times using a different initial estimate for the solution in each case in the hope that several distinct local minima will be located. If this approach is successful, the best minimizer, namely, the one yielding the lowest value for the objective function can be selected. Although such a solution could be acceptable from a practical point of view, usually

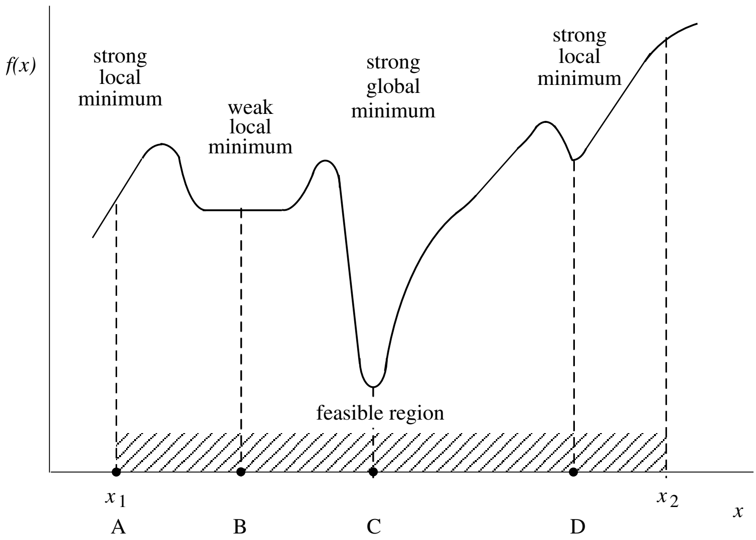

Figure 2.2. Types of minima. (Example 2.1)

there is no guarantee that the global minimum will be achieved. Therefore, for the sake of convenience, the term 'minimize $f(\mathbf{x})$' in the general optimization problem will be interpreted as 'find a local minimum of $f(\mathbf{x})$'.

In a specific class of problems where function $f(\mathbf{x})$ and set $\mathcal{R}$ satisfy certain convexity properties, any local minimum of $f(\mathbf{x})$ is also a global minimum of $f(\mathbf{x})$. In this class of problems an optimal solution can be assured. These problems will be examined in Sec. 2.7.

2.5 Necessary and Sufficient Conditions for Local Minima and Maxima

The gradient $\mathbf{g}(\mathbf{x})$ and the Hessian $\mathbf{H}(\mathbf{x})$ must satisfy certain conditions at a local minimizer $\mathbf{x}^*$, (see [2, Chap. 6]). Two sets of conditions will be discussed, as follows:

1. Conditions which are satisfied at a local minimizer $\mathbf{x}^*$. These are the necessary conditions.
2. Conditions which guarantee that $\mathbf{x}^*$ is a local minimizer. These are the sufficient conditions.

The necessary and sufficient conditions can be described in terms of a number of theorems. A concept that is used extensively in these theorems is the concept of a feasible direction.

Definition 2.4 Let $\boldsymbol{\delta} = \alpha\mathbf{d}$ be a change in $\mathbf{x}$ where α is a positive constant and $\mathbf{d}$ is a direction vector. If $\mathcal{R}$ is the feasible region and a constant $\hat{\alpha} > 0$ exists such that

$$\mathbf{x} + \alpha\mathbf{d} \in \mathcal{R}$$

for all α in the range $0 \leq \alpha \leq \hat{\alpha}$, then $\mathbf{d}$ is said to be a *feasible direction* at point $\mathbf{x}$.

■

In effect, if a point $\mathbf{x}$ remains in $\mathcal{R}$ after it is moved a finite distance in a direction $\mathbf{d}$, then $\mathbf{d}$ is a feasible direction vector at $\mathbf{x}$.

Example 2.2 The feasible region in an optimization problem is given by

$$\mathcal{R} = \{\mathbf{x} : \ x_1 \geq 2, \ x_2 \geq 0\}$$

as depicted in Fig. 2.3. Which of the vectors $\mathbf{d}_1 = [-2 \ 2]^T$, $\mathbf{d}_2 = [0 \ 2]^T$, $\mathbf{d}_3 = [2 \ 0]^T$ are feasible directions at points $\mathbf{x}_1 = [4 \ 1]^T$, $\mathbf{x}_2 = [2 \ 3]^T$, and $\mathbf{x}_3 = [1 \ 4]^T$?

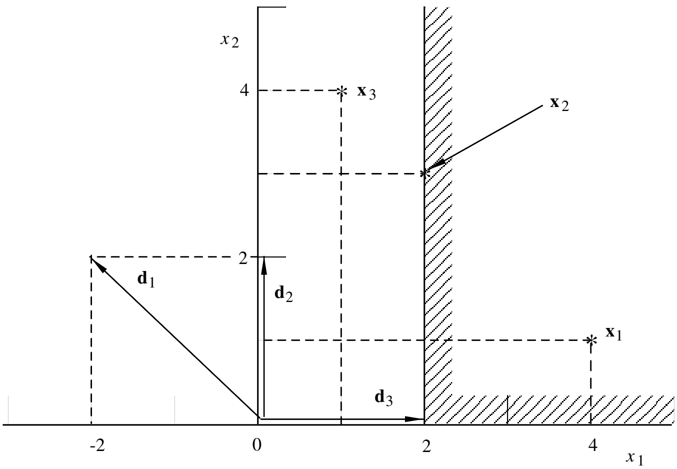

Figure 2.3. Graphical construction for Example 2.2.

Solution Since

$$\mathbf{x}_1 + \alpha\mathbf{d}_1 \in \mathcal{R}$$

for all α in the range $0 \leq \alpha \leq \hat{\alpha}$ for $\hat{\alpha} = 1$, $\mathbf{d}_1$ is a feasible direction at point $\mathbf{x}_1$; for any range $0 \leq \alpha \leq \hat{\alpha}$

$$\mathbf{x}_1 + \alpha\mathbf{d}_2 \in \mathcal{R} \quad \text{and} \quad \mathbf{x}_1 + \alpha\mathbf{d}_3 \in \mathcal{R}$$

Hence $\mathbf{d}_2$ and $\mathbf{d}_3$ are feasible directions at $\mathbf{x}_1$.

Since no constant $\hat{\alpha} > 0$ can be found such that

$$\mathbf{x}_2 + \alpha\mathbf{d}_1 \in \mathcal{R} \qquad \text{for } \ 0 \leq \alpha \leq \hat{\alpha}$$

$\mathbf{d}_1$ is not a feasible direction at $\mathbf{x}_2$. On the other hand, a positive constant $\hat{\alpha}$ exists such that

$$\mathbf{x}_2 + \alpha\mathbf{d}_2 \in \mathcal{R} \quad \text{and} \quad \mathbf{x}_2 + \alpha\mathbf{d}_3 \in \mathcal{R}$$

for $0 \le \alpha \le \hat{\alpha}$, and so $\mathbf{d}_2$ and $\mathbf{d}_3$ are feasible directions at $\mathbf{x}_2$.

Since $\mathbf{x}_3$ is not in $\mathcal{R}$, no $\hat{\alpha} > 0$ exists such that

$$\mathbf{x}_3 + \alpha\mathbf{d} \in \mathcal{R} \qquad \text{for } \ 0 \le \alpha \le \hat{\alpha}$$

for any $\mathbf{d}$. Hence $\mathbf{d}_1, \mathbf{d}_2$, and $\mathbf{d}_3$ are not feasible directions at $\mathbf{x}_3$.

■

2.5.1 First-order necessary conditions

The objective function must satisfy two sets of conditions in order to have a minimum, namely, first- and second-order conditions. The first-order conditions are in terms of the first derivatives, i.e., the gradient.

Theorem 2.1 ***First-order necessary conditions for a minimum***

(a) If $f(\mathbf{x}) \in C^1$ and $\mathbf{x}^*$ is a local minimizer, then

$$\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} \ge 0$$

for every feasible direction $\mathbf{d}$ at $\mathbf{x}^*$.

(b) If $\mathbf{x}^*$ is located in the interior of $\mathcal{R}$ then

$$\mathbf{g}(\mathbf{x}^*) = 0$$

Proof (a) If $\mathbf{d}$ is a feasible direction at $\mathbf{x}^*$, then from Def. 2.4

$$\mathbf{x} = \mathbf{x}^* + \alpha\mathbf{d} \in \mathcal{R} \qquad \text{for } \ 0 \le \alpha \le \hat{\alpha}$$

From the Taylor series

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \alpha\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} + o(\alpha\|\mathbf{d}\|)$$

If

$$\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} < 0$$

then as $\alpha \to 0$

$$\alpha\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} + o(\alpha\|\mathbf{d}\|) < 0$$

and so

$$f(\mathbf{x}) < f(\mathbf{x}^*)$$

This contradicts the assumption that $\mathbf{x}^*$ is a minimizer. Therefore, a necessary condition for $\mathbf{x}^*$ to be a minimizer is

$$\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} \geq 0$$

(b) If $\mathbf{x}^*$ is in the interior of $\mathcal{R}$, vectors exist in all directions which are feasible. Thus from part (a), a direction $\mathbf{d} = \mathbf{d}_1$ yields

$$\mathbf{g}(\mathbf{x}^*)^T\mathbf{d}_1 \geq 0$$

Similarly, for a direction $\mathbf{d} = -\mathbf{d}_1$

$$-\mathbf{g}(\mathbf{x}^*)^T\mathbf{d}_1 \geq 0$$

Therefore, in this case, a necessary condition for $\mathbf{x}^*$ to be a local minimizer is

$$\mathbf{g}(\mathbf{x}^*) = 0$$

■

2.5.2 Second-order necessary conditions

The second-order necessary conditions involve the first as well as the second derivatives or, equivalently, the gradient and the Hessian.

Definition 2.5

(a) Let $\mathbf{d}$ be an arbitrary direction vector at point $\mathbf{x}$. The quadratic form $\mathbf{d}^T\mathbf{H}(\mathbf{x})\mathbf{d}$ is said to be *positive definite, positive semidefinite, negative semidefinite, negative definite* if $\mathbf{d}^T\mathbf{H}(\mathbf{x})\mathbf{d} > 0,\ \geq 0,\ \leq 0,\ < 0$, respectively, for all $\mathbf{d} \neq \mathbf{0}$ at $\mathbf{x}$. If $\mathbf{d}^T\mathbf{H}(\mathbf{x})\mathbf{d}$ can assume positive as well as negative values, it is said to be *indefinite*.

(b) If $\mathbf{d}^T\mathbf{H}(\mathbf{x})\mathbf{d}$ is positive definite, positive semidefinite, etc., then matrix $\mathbf{H}(\mathbf{x})$ is said to be positive definite, positive semidefinite, etc.

■

Theorem 2.2 ***Second-order necessary conditions for a minimum***

(a) If $f(\mathbf{x}) \in C^2$ and $\mathbf{x}^*$ is a local minimizer, then for every feasible direction $\mathbf{d}$ at $\mathbf{x}^*$

(i) $\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} \geq 0$

(ii) If $\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} = 0$, *then* $\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} \geq 0$

(b) If $\mathbf{x}^*$ is a local minimizer in the interior of $\mathcal{R}$, then

(i) $\mathbf{g}(\mathbf{x}^*) = \mathbf{0}$

(ii) $\mathbf{d}^T\mathbf{H}(\mathbf{x})^*\mathbf{d} \geq 0$ *for all* $\mathbf{d} \neq \mathbf{0}$

Proof Conditions (i) in parts (a) and (b) are the same as in Theorem 2.1(a) and (b).

Condition (ii) of part (a) can be proved by letting $\mathbf{x} = \mathbf{x}^* + \alpha\mathbf{d}$, where $\mathbf{d}$ is a feasible direction. The Taylor series gives

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \alpha\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} + \tfrac{1}{2}\alpha^2\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} + o(\alpha^2\|\mathbf{d}\|^2)$$

Now if condition (i) is satisfied with the equal sign, then

$$f(\mathbf{x}) = f(\mathbf{x}^*) + \tfrac{1}{2}\alpha^2\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} + o(\alpha^2\|\mathbf{d}\|^2)$$

If

$$\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} < 0$$

then as $\alpha \to 0$

$$\tfrac{1}{2}\alpha^2\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} + o(\alpha^2\|\mathbf{d}\|^2) < 0$$

and so

$$f(\mathbf{x}) < f(\mathbf{x}^*)$$

This contradicts the assumption that $\mathbf{x}^*$ is a minimizer. Therefore, if $\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} = 0$, then

$$\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} \geq 0$$

If $\mathbf{x}^*$ is a local minimizer in the interior of $\mathcal{R}$, then all vectors $\mathbf{d}$ are feasible directions and, therefore, condition (ii) of part (b) holds. This condition is equivalent to stating that $\mathbf{H}(\mathbf{x}^*)$ is positive semidefinite, according to Def. 2.5.

■

Example 2.3 Point $\mathbf{x}^* = [\frac{1}{2}\ 0]^T$ is a local minimizer of the problem

$$\text{minimize } f(x_1,\ x_2) = x_1^2 - x_1 + x_2 + x_1x_2$$

$$\text{subject to :} \quad x_1 \geq 0,\ x_2 \geq 0$$

Show that the necessary conditions for $\mathbf{x}^*$ to be a local minimizer are satisfied.

Solution The partial derivatives of $f(x_1,\ x_2)$ are

$$\frac{\partial f}{\partial x_1} = 2x_1 - 1 + x_2, \quad \frac{\partial f}{\partial x_2} = 1 + x_1$$

Hence if $\mathbf{d} = [d_1\ d_2]^T$ is a feasible direction, we obtain

$$\mathbf{g}(\mathbf{x})^T\mathbf{d} = (2x_1 - 1 + x_2)d_1 + (1 + x_1)d_2$$

At $\mathbf{x} = \mathbf{x}^*$

$$\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} = \tfrac{3}{2}d_2$$

and since $d_2 \geq 0$ for $\mathbf{d}$ to be a feasible direction, we have

$$\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} \geq 0$$

Therefore, the first-order necessary conditions for a minimum are satisfied.

Now

$$\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} = 0$$

if $d_2 = 0$. The Hessian is

$$\mathbf{H}(\mathbf{x}^*) = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix}$$

and so

$$\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} = 2d_1^2 + 2d_1d_2$$

For $d_2 = 0$, we obtain

$$\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} = 2d_1^2 \geq 0$$

for every feasible value of d_1. Therefore, the second-order necessary conditions for a minimum are satisfied.

■

Example 2.4 Points $\mathbf{p}_1 = [0\ 0]^T$ and $\mathbf{p}_2 = [6\ 9]^T$ are probable minimizers for the problem

$$\text{minimize } f(x_1,\ x_2) = x_1^3 - x_1^2x_2 + 2x_2^2$$
$$\text{subject to :} \quad x_1 \geq 0,\ x_2 \geq 0$$

Check whether the necessary conditions of Theorems 2.1 and 2.2 are satisfied.

Solution The partial derivatives of $f(x_1,\ x_2)$ are

$$\frac{\partial f}{\partial x_1} = 3x_1^2 - 2x_1x_2, \qquad \frac{\partial f}{\partial x_2} = -x_1^2 + 4x_2$$

Hence if $\mathbf{d} = [d_1\ d_2]^T$, we obtain

$$\mathbf{g}(\mathbf{x})^T\mathbf{d} = (3x_1^2 - 2x_1x_2)d_1 + (-x_1^2 + 4x_2)d_2$$

At points $\mathbf{p}_1$ and $\mathbf{p}_2$

$$\mathbf{g}(\mathbf{x})^T\mathbf{d} = 0$$

i.e., the first-order necessary conditions are satisfied. The Hessian is

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} 6x_1 - 2x_2 & -2x_1 \\ -2x_1 & 4 \end{bmatrix}$$

and if $\mathbf{x} = \mathbf{p}_1$, then

$$\mathbf{H}(\mathbf{p}_1) = \begin{bmatrix} 0 & 0 \\ 0 & 4 \end{bmatrix}$$

and so

$$\mathbf{d}^T\mathbf{H}(\mathbf{p}_1)\mathbf{d} = 4d_2^2 \geq 0$$

Hence the second-order necessary conditions are satisfied at $\mathbf{x} = \mathbf{p}_1$, and $\mathbf{p}_1$ can be a local minimizer.

If $\mathbf{x} = \mathbf{p}_2$, then

$$\mathbf{H}(\mathbf{p}_2) = \begin{bmatrix} 18 & -12 \\ -12 & 4 \end{bmatrix}$$

and

$$\mathbf{d}^T\mathbf{H}(\mathbf{p}_2)\mathbf{d} = 18d_1^2 - 24d_1d_2 + 4d_2^2$$

Since $\mathbf{d}^T\mathbf{H}(\mathbf{p}_2)\mathbf{d}$ is indefinite, the second-order necessary conditions are violated, that is, $\mathbf{p}_2$ cannot be a local minimizer.

■

Analogous conditions hold for the case of a local maximizer as stated in the following theorem:

Theorem 2.3 ***Second-order necessary conditions for a maximum***

(a) If $f(\mathbf{x}) \in C^2$, and $\mathbf{x}^*$ is a local maximizer, then for every feasible direction $\mathbf{d}$ at $\mathbf{x}^*$
(i) $\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} \leq 0$
(ii) If $\mathbf{g}(\mathbf{x}^*)^T\,\mathbf{d} = 0$*, then* $\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} \leq 0$

(b) If $\mathbf{x}^*$ is a local maximizer in the interior of $\mathcal{R}$ then
(i) $\mathbf{g}(\mathbf{x}^*) = \mathbf{0}$
(ii) $\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} \leq 0$ *for all* $\mathbf{d} \neq \mathbf{0}$

Condition (ii) of part (b) is equivalent to stating that $\mathbf{H}(\mathbf{x}^*)$ is negative semidefinite.

The conditions considered are necessary but not sufficient for a point to be a local extremum point, that is, a point may satisfy these conditions without being a local extremum point. We now focus our attention on a set of stronger conditions that are *sufficient* for a point to be a local extremum. We consider conditions that are applicable in the case where $\mathbf{x}^*$ is located in the interior of the feasible region. Sufficient conditions that are applicable to the case where $\mathbf{x}^*$ is located on a boundary of the feasible region are somewhat more difficult to deduce and will be considered in Chap. 10.

Theorem 2.4 ***Second-order sufficient conditions for a minimum*** *If $f(\mathbf{x}) \in C^2$ and $\mathbf{x}^*$ is located in the interior of $\mathcal{R}$, then the conditions*
(a) $\mathbf{g}(\mathbf{x}^*) = \mathbf{0}$
(b) $\mathbf{H}(\mathbf{x}^*)$ *is positive definite*
are sufficient for $\mathbf{x}^*$ to be a strong local minimizer.

Proof For any direction $\mathbf{d}$, the Taylor series yields

$$f(\mathbf{x}^* + \mathbf{d}) = f(\mathbf{x}^*) + \mathbf{g}(\mathbf{x}^*)^T\mathbf{d} + \tfrac{1}{2}\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} + o(\|\mathbf{d}\|^2)$$

and if condition (a) is satisfied, we have

$$f(\mathbf{x}^* + \mathbf{d}) = f(\mathbf{x}^*) + \tfrac{1}{2}\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} + o(\|\mathbf{d}\|^2)$$

Now if condition (b) is satisfied, then

$$\tfrac{1}{2}\mathbf{d}^T\mathbf{H}(\mathbf{x}^*)\mathbf{d} + o(\|\mathbf{d}\|^2) > 0 \qquad \text{as } \|\mathbf{d}\| \to 0$$

Therefore,

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

that is, $\mathbf{x}^*$ is a strong local minimizer.

■

Analogous conditions hold for a maximizer as stated in Theorem 2.5 below.

Theorem 2.5 ***Second-order sufficient conditions for a maximum*** *If* $f(\mathbf{x}^*) \in C^2$ *and* $\mathbf{x}^*$ *is located in the interior of* $\mathcal{R}$*, then the conditions*
(a) $\mathbf{g}(\mathbf{x}) = \mathbf{0}$
(b) $\mathbf{H}(\mathbf{x}^*)$ *is negative definite*
are sufficient for $\mathbf{x}^*$ *to be a strong local maximizer.*

2.6 Classification of Stationary Points

If the extremum points of the type considered so far, namely, minimizers and maximizers, are located in the interior of the feasible region, they are called *stationary points* since $\mathbf{g}(\mathbf{x}) = \mathbf{0}$ at these points. Another type of stationary point of interest is the saddle point.

Definition 2.6 A point $\bar{\mathbf{x}} \in \mathcal{R}$, where $\mathcal{R}$ is the feasible region, is said to be a *saddle point* if
(a) $\mathbf{g}(\bar{\mathbf{x}}) = \mathbf{0}$
(b) point $\bar{\mathbf{x}}$ is neither a maximizer nor a minimizer.

■

A saddle point in E^2 is illustrated in Fig. 2.4.

At a point $\mathbf{x} = \bar{\mathbf{x}} + \alpha\mathbf{d} \in \mathcal{R}$ in the neighborhood of a saddle point $\bar{\mathbf{x}}$, the Taylor series gives

$$f(\mathbf{x}) = f(\bar{\mathbf{x}}) + \tfrac{1}{2}\alpha^2\mathbf{d}^T\mathbf{H}(\bar{\mathbf{x}})\mathbf{d} + o(\alpha^2\|\mathbf{d}\|^2)$$

since $\mathbf{g}(\bar{\mathbf{x}}) = \mathbf{0}$. From the definition of a saddle point, directions $\mathbf{d}_1$ and $\mathbf{d}_2$ must exist such that

$$f(\bar{\mathbf{x}} + \alpha\mathbf{d}_1) < f(\bar{\mathbf{x}}) \quad \text{and} \quad f(\bar{\mathbf{x}} + \alpha\mathbf{d}_2) > f(\bar{\mathbf{x}})$$

Since $\bar{\mathbf{x}}$ is neither a minimizer nor a maximizer, then as $\alpha \to 0$ we have

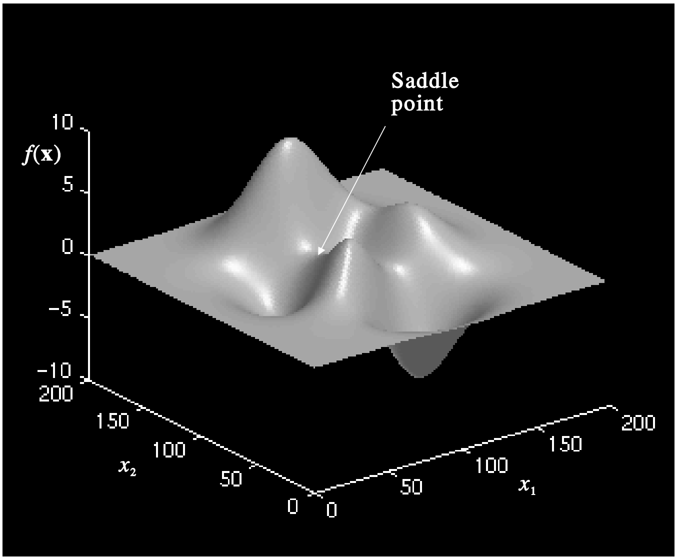

Figure 2.4. A saddle point in E^2.

$$\mathbf{d}_1^T \mathbf{H}(\bar{\mathbf{x}})\mathbf{d}_1 < 0 \quad \text{and} \quad \mathbf{d}_2^T \mathbf{H}(\bar{\mathbf{x}})\mathbf{d}_2 > 0$$

Therefore, matrix $\mathbf{H}(\bar{\mathbf{x}})$ must be indefinite.

Stationary points can be located and classified as follows:

1. Find the points $\mathbf{x}_i$ at which $\mathbf{g}(\mathbf{x}_i) = \mathbf{0}$.
2. Obtain the Hessian $\mathbf{H}(\mathbf{x}_i)$.
3. Determine the character of $\mathbf{H}(\mathbf{x}_i)$ for each point $\mathbf{x}_i$.

If $\mathbf{H}(\mathbf{x}_i)$ is positive (or negative) definite, $\mathbf{x}_i$ is a minimizer (or maximizer); if $\mathbf{H}(\mathbf{x}_i)$ is indefinite, $\mathbf{x}_i$ is a saddle point. If $\mathbf{H}(\mathbf{x}_i)$ is positive (or negative) semidefinite, $\mathbf{x}_i$ can be a minimizer (or maximizer); in the special case where $\mathbf{H}(\mathbf{x}_i) = \mathbf{0}$, $\mathbf{x}_i$ can be a minimizer or maximizer since the necessary conditions are satisfied in both cases. Evidently, if $\mathbf{H}(\mathbf{x}_i)$ is semidefinite, insufficient information is available for the complete characterization of a stationary point and further work is, therefore, necessary in such a case. A possible approach would be to deduce the third partial derivatives of $f(\mathbf{x})$ and then calculate the fourth term in the Taylor series, namely, term $D^3 f(\mathbf{x})/3!$ in Eq. (2.4i). If the fourth term is zero, then the fifth term needs to be calculated and so on. An alternative and more practical approach would be to compute $f(\mathbf{x}_i + \mathbf{e}_j)$ and $f(\mathbf{x}_i - \mathbf{e}_j)$ for $j = 1, 2, \ldots, n$ where $\mathbf{e}_j$ is a vector with elements

$$e_{jk} = \begin{cases} 0 & \text{for } k \neq j \\ \varepsilon & \text{for } k = j \end{cases}$$

for some small positive value of ε and then check whether the definition of a minimizer or maximizer is satisfied.

Example 2.5 Find and classify the stationary points of

$$f(\mathbf{x}) = (x_1 - 2)^3 + (x_2 - 3)^3$$

Solution The first-order partial derivatives of $f(\mathbf{x})$ are

$$\frac{\partial f}{\partial x_1} = 3(x_1 - 2)^2$$

$$\frac{\partial f}{\partial x_2} = 3(x_2 - 3)^2$$

If $\mathbf{g} = \mathbf{0}$, then

$$3(x_1 - 2)^2 = 0 \quad \text{and} \quad 3(x_2 - 3)^2 = 0$$

and so there is a stationary point at

$$\mathbf{x} = \mathbf{x}_1 = [2\ 3]^T$$

The Hessian is given by

$$\mathbf{H} = \begin{bmatrix} 6(x_1 - 2) & 0 \\ 0 & 6(x_2 - 3) \end{bmatrix}$$

and at $\mathbf{x} = \mathbf{x}_1$

$$\mathbf{H} = \mathbf{0}$$

Since $\mathbf{H}$ is semidefinite, more work is necessary in order to determine the type of stationary point.

The third derivatives are all zero except for $\partial^3 f/\partial x_1^3$ and $\partial^3 f/\partial x_2^3$ which are both equal to 6. For point $\mathbf{x}_1 + \boldsymbol{\delta}$, the fourth term in the Taylor series is given by

$$\frac{1}{3!}\left(\delta_1^3 \frac{\partial^3 f}{\partial x_1^3} + \delta_2^3 \frac{\partial^3 f}{\partial x_2^3}\right) = \delta_1^3 + \delta_2^3$$

and is positive for $\delta_1,\ \delta_2 > 0$ and negative for $\delta_1,\ \delta_2 < 0$. Hence

$$f(\mathbf{x}_1 + \boldsymbol{\delta}) > f(\mathbf{x}_1) \qquad \text{for } \delta_1,\ \delta_2 > 0$$

and

$$f(\mathbf{x}_1 + \boldsymbol{\delta}) < f(\mathbf{x}_1) \qquad \text{for } \delta_1,\ \delta_2 < 0$$

that is, $\mathbf{x}_1$ is neither a minimizer nor a maximizer. Therefore, $\mathbf{x}_1$ is a saddle point.

■

From the preceding discussion, it follows that the problem of classifying the stationary points of function $f(\mathbf{x})$ reduces to the problem of characterizing the Hessian. This problem can be solved by using the following theorems.

Theorem 2.6 ***Characterization of symmetric matrices*** *A real symmetric $n \times n$ matrix $\mathbf{H}$ is positive definite, positive semidefinite, etc., if for every nonsingular matrix $\mathbf{B}$ of the same order, the $n \times n$ matrix $\hat{\mathbf{H}}$ given by*

$$\hat{\mathbf{H}} = \mathbf{B}^T \mathbf{H} \mathbf{B}$$

is positive definite, positive semidefinite, etc.

Proof If $\mathbf{H}$ is positive definite, positive semidefinite etc., then for all $\mathbf{d} \neq 0$

$$\begin{aligned}\mathbf{d}^T \hat{\mathbf{H}} \mathbf{d} &= \mathbf{d}^T (\mathbf{B}^T \mathbf{H} \mathbf{B}) \mathbf{d} \\ &= (\mathbf{d}^T \mathbf{B}^T) \mathbf{H} (\mathbf{B} \mathbf{d}) \\ &= (\mathbf{B} \mathbf{d})^T \mathbf{H} (\mathbf{B} \mathbf{d})\end{aligned}$$

Since $\mathbf{B}$ is nonsingular, $\mathbf{B}\mathbf{d} = \hat{\mathbf{d}}$ is a nonzero vector and thus

$$\mathbf{d}^T \hat{\mathbf{H}} \mathbf{d} = \hat{\mathbf{d}}^T \mathbf{H} \hat{\mathbf{d}} > 0, \ \geq 0, \ \text{etc.}$$

for all $\mathbf{d} \neq \mathbf{0}$. Therefore,

$$\hat{\mathbf{H}} = \mathbf{B}^T \mathbf{H} \mathbf{B}$$

is positive definite, positive semidefinite, etc.

■

Theorem 2.7 ***Characterization of symmetric matrices via diagonalization***

(*a*) *If the $n \times n$ matrix $\mathbf{B}$ is nonsingular and*

$$\hat{\mathbf{H}} = \mathbf{B}^T \mathbf{H} \mathbf{B}$$

is a diagonal matrix with diagonal elements $\hat{h}_1, \ \hat{h}_2, \ \ldots, \ \hat{h}_n$ then $\mathbf{H}$ is positive definite, positive semidefinite, negative semidefinite, negative definite, if $\hat{h}_i > 0, \ \geq 0, \ \leq 0, \ < 0$ for $i = 1, 2, \ldots, n$. Otherwise, if some $\hat{h}_i$ are positive and some are negative, $\mathbf{H}$ is indefinite.

(*b*) *The converse of part (a) is also true, that is, if $\mathbf{H}$ is positive definite, positive semidefinite, etc., then $\hat{h}_i > 0, \ \geq 0$, etc., and if $\mathbf{H}$ is indefinite, then some $\hat{h}_i$ are positive and some are negative.*

Proof (a) For all $\mathbf{d} \neq \mathbf{0}$

$$\mathbf{d} \hat{\mathbf{H}} \mathbf{d} = d_1^2 \hat{h}_1 + d_2^2 \hat{h}_2 + \cdots + d_n^2 \hat{h}_n$$

Therefore, if $\hat{h}_i > 0,\ \geq 0$, etc. for $i = 1,\ 2,\ \ldots,\ n$, then

$$\mathbf{d}^T\hat{\mathbf{H}}\mathbf{d} > 0,\ \geq 0,\ \text{etc.}$$

that is, $\hat{\mathbf{H}}$ is positive definite, positive semidefinite etc. If some $\hat{h}_i$ are positive and some are negative, a vector $\mathbf{d}$ can be found which will yield a positive or negative $\mathbf{d}^T\hat{\mathbf{H}}\mathbf{d}$ and then $\hat{\mathbf{H}}$ is indefinite. Now since $\hat{\mathbf{H}} = \mathbf{B}^T\mathbf{H}\mathbf{B}$, it follows from Theorem 2.6 that if $\hat{h}_i > 0,\ \geq 0$, etc. for $i = 1,\ 2,\ \ldots,\ n$, then $\mathbf{H}$ is positive definite, positive semidefinite, etc.

(b) Suppose that $\mathbf{H}$ is positive definite, positive semidefinite, etc. Since $\hat{\mathbf{H}} = \mathbf{B}^T\mathbf{H}\mathbf{B}$, it follows from Theorem 2.6 that $\hat{\mathbf{H}}$ is positive definite, positive semidefinite, etc. If $\mathbf{d}$ is a vector with element d_k given by

$$d_k = \begin{cases} 0 & \text{for } k \neq i \\ 1 & \text{for } k = i \end{cases}$$

then

$$\mathbf{d}^T\hat{\mathbf{H}}\mathbf{d} = \hat{h}_i\ > 0,\ \geq 0,\ \text{etc.} \qquad \text{for } i = 1,\ 2,\ \ldots,\ n$$

If $\mathbf{H}$ is indefinite, then from Theorem 2.6 it follows that $\hat{\mathbf{H}}$ is indefinite, and, therefore, some $\hat{h}_i$ must be positive and some must be negative.

■

A diagonal matrix $\hat{\mathbf{H}}$ can be obtained by performing row and column operations on $\mathbf{H}$, like adding k times a given row to another row or adding m times a given column to another column. For a symmetric matrix, these operations can be carried out by applying *elementary transformations*, that is, $\hat{\mathbf{H}}$ can be formed as

$$\hat{\mathbf{H}} = \cdots \mathbf{E}_3\mathbf{E}_2\mathbf{E}_1\mathbf{H}\mathbf{E}_1^T\mathbf{E}_2^T\mathbf{E}_3^T \cdots \tag{2.8}$$

where $\mathbf{E}_1, \mathbf{E}_2, \cdots$ are elementary matrices. Typical elementary matrices are

$$\mathbf{E}_a = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & k & 1 \end{bmatrix}$$

and

$$\mathbf{E}_b = \begin{bmatrix} 1 & m & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

If E_a premultiplies a 3×3 matrix, it will cause k times the second row to be added to the third row, and if $\mathbf{E}_b$ postmultiplies a 4×4 matrix it will cause m times the first column to be added to the second column. If

$$\mathbf{B} = \mathbf{E}_1^T\mathbf{E}_2^T\mathbf{E}_3^T \cdots$$

then

$$\mathbf{B}^T = \cdots \mathbf{E}_3\mathbf{E}_2\mathbf{E}_1$$

and so Eq. (2.8) can be expressed as

$$\hat{\mathbf{H}} = \mathbf{B}^T\mathbf{H}\mathbf{B}$$

Since elementary matrices are nonsingular, B is nonsingular, and hence $\hat{\mathbf{H}}$ is positive definite, positive semidefinite, etc., if $\mathbf{H}$ is positive definite, positive semidefinite, etc.

Therefore, the characterization of $\mathbf{H}$ can be achieved by diagonalizing $\mathbf{H}$, through the use of appropriate elementary matrices, and then using Theorem 2.7.

Example 2.6 Diagonalize the matrix

$$\mathbf{H} = \begin{bmatrix} 1 & -2 & 4 \\ -2 & 2 & 0 \\ 4 & 0 & -7 \end{bmatrix}$$

and then characterize it.

Solution Add 2 times the first row to the second row as

$$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & -2 & 4 \\ -2 & 2 & 0 \\ 4 & 0 & -7 \end{bmatrix} \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 4 \\ 0 & -2 & 8 \\ 4 & 8 & -7 \end{bmatrix}$$

Add -4 times the first row to the third row as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -4 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 4 \\ 0 & -2 & 8 \\ 4 & 8 & -7 \end{bmatrix} \begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 8 \\ 0 & 8 & -23 \end{bmatrix}$$

Now add 4 times the second row to the third row as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 4 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 8 \\ 0 & 8 & -23 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 4 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

Since $\hat{h}_1 = 1, \hat{h}_2 = -2, \hat{h}_3 = 9$, $\mathbf{H}$ is indefinite.

■

Example 2.7 Diagonalize the matrix

$$\mathbf{H} = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 50 \end{bmatrix}$$

and determine its characterization.

Solution Add 0.5 times the first row to the second row as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 50 \end{bmatrix} \begin{bmatrix} 1 & 0.5 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 50 \end{bmatrix}$$

Hence **H** is positive definite.

■

Another theorem that can be used to characterize the Hessian is as follows:

Theorem 2.8 ***Eigendecomposition of symmetric matrices***

(a) If **H** *is a real symmetric matrix, then there exists a real unitary (or orthogonal) matrix* **U** *such that*

$$\mathbf{\Lambda} = \mathbf{U}^T \mathbf{H} \mathbf{U}$$

is a diagonal matrix whose diagonal elements are the eigenvalues of **H**.

(b) The eigenvalues of **H** *are real.*

(See Chap. 4 of Horn and Johnson [3] for proofs.)

For a real unitary matrix, we have $\mathbf{U}^T\mathbf{U} = \mathbf{I}_n$ where

$$\mathbf{I}_n = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

is the $n \times n$ identity matrix, and hence $\det \mathbf{U} = \pm 1$, that is, **U** is nonsingular. From Theorem 2.6, $\mathbf{\Lambda}$ is positive definite, positive semidefinite, etc. if **H** is positive definite, positive semidefinite, etc. Therefore, **H** can be characterized by deducing its eigenvalues and then checking their signs as in Theorem 2.7.

Another approach for the characterization of a square matrix **H** is based on the evaluation of the so-called *principal minors* and *leading principal minors* of **H**, which are described in Sec. A.6. The details of this approach are summarized in terms of the following theorem.

Theorem 2.9 ***Properties of matrices***

(a) If **H** *is positive semidefinite or positive definite, then*

$$\det \mathbf{H} \geq 0 \text{ or } > 0$$

(b) **H** *is positive definite if and only if all its leading principal minors are positive, i.e.,* $\det \mathbf{H}_i > 0$ *for* $i = 1, 2, \ldots, n$.

(*c*) $\mathbf{H}$ *is positive semidefinite if and only if all its principal minors are nonnegative, i.e.,* $\det\,(\mathbf{H}_i^{(l)}) \geq 0$ *for all possible selections of* $\{l_1, l_2, \ldots, l_i\}$ *for* $i = 1, 2, \ldots, n$.

(*d*) $\mathbf{H}$ *is negative definite if and only if all the leading principal minors of* $-\mathbf{H}$ *are positive, i.e.,* $\det\,(-\mathbf{H}_i) > 0$ *for* $i = 1, 2, \ldots, n$.

(*e*) $\mathbf{H}$ *is negative semidefinite if and only if all the principal minors of* $-\mathbf{H}$ *are nonnegative, i.e.,* $\det\,(-\mathbf{H}_i^{(l)}) \geq 0$ *for all possible selections of* $\{l_1, l_2, \ldots, l_i\}$ *for* $i = 1, 2, \ldots, n$.

(*f*) $\mathbf{H}$ *is indefinite if neither (c) nor (e) holds.*

Proof (a) Elementary transformations do not change the determinant of a matrix and hence

$$\det \mathbf{H} = \det \hat{\mathbf{H}} = \prod_{i=1}^{n} \hat{h}_i$$

where $\hat{\mathbf{H}}$ is a diagonalized version of $\mathbf{H}$ with diagonal elements $\hat{h}_i$. If $\mathbf{H}$ is positive semidefinite or positive definite, then $\hat{h}_i \geq 0$ or > 0 from Theorem 2.7 and, therefore,

$$\det \mathbf{H} \geq 0 \text{ or } > 0$$

(b) If

$$\mathbf{d} = [d_1\; d_2\; \cdots\; d_i\; 0\; 0 \cdots\; 0]^T$$

and $\mathbf{H}$ is positive definite, then

$$\mathbf{d}^T\mathbf{H}\mathbf{d} = \mathbf{d}_0^T\mathbf{H}_i\mathbf{d}_0 > 0$$

for all $\mathbf{d}_0 \neq \mathbf{0}$ where

$$\mathbf{d}_0 = [d_1\; d_2\; \cdots\; d_i]^T$$

and $\mathbf{H}_i$ is the ith leading principal submatrix of $\mathbf{H}$. The preceding inequality holds for $i = 1, 2, \ldots, n$ and, hence $\mathbf{H}_i$ is positive definite for $i = 1, 2, \ldots, n$. From part (a)

$$\det \mathbf{H}_i > 0 \qquad \text{for } i = 1, 2, \ldots, n$$

Now we prove the sufficiency of the theorem by induction. If $n = 1$, then $\mathbf{H} = a_{11}$, and $\det\,(\mathbf{H}_1) = a_{11} > 0$ implies that $\mathbf{H}$ is positive definite. We assume that the sufficiency is valid for matrix $\mathbf{H}$ of size $(n-1)$ by $(n-1)$ and we shall show that the sufficiency is also valid for matrix $\mathbf{H}$ of size n by n. First, we write $\mathbf{H}$ as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{n-1} & \mathbf{h} \\ \mathbf{h}^T & h_{nn} \end{bmatrix}$$

where

$$\mathbf{H}_{n-1} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1,n-1} \\ h_{21} & h_{22} & \cdots & h_{2,n-1} \\ \vdots & \vdots & & \vdots \\ h_{n-1,1} & h_{n-1,2} & \cdots & h_{n-1,n-1} \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} h_{1n} \\ h_{2n} \\ \vdots \\ h_{n-1,n} \end{bmatrix}$$

By assumption $\mathbf{H}_{n-1}$ is positive definite; hence there exists an $\mathbf{R}$ such that

$$\mathbf{R}^T\mathbf{H}_{n-1}\mathbf{R} = \mathbf{I}_{n-1}$$

where $\mathbf{I}_{n-1}$ is the $(n-1)\times(n-1)$ identity matrix. If we let

$$\mathbf{S} = \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix}$$

we obtain

$$\mathbf{S}^T\mathbf{H}\mathbf{S} = \begin{bmatrix} \mathbf{R}^T & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{H}_{n-1} & \mathbf{h} \\ \mathbf{h}^T & h_{nn} \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{0} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I}_{n-1} & \mathbf{R}^T\mathbf{h} \\ \mathbf{h}^T\mathbf{R} & h_{nn} \end{bmatrix}$$

If we define

$$\mathbf{T} = \begin{bmatrix} \mathbf{I}_{n-1} & -\mathbf{R}^T\mathbf{h} \\ \mathbf{0} & 1 \end{bmatrix}$$

then

$$\begin{aligned} \mathbf{T}^T\mathbf{S}^T\mathbf{H}\mathbf{S}\mathbf{T} &= \begin{bmatrix} \mathbf{I}_{n-1} & \mathbf{0} \\ -\mathbf{h}^T\mathbf{R} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{I}_{n-1} & \mathbf{R}^T\mathbf{h} \\ \mathbf{h}^T\mathbf{R} & h_{nn} \end{bmatrix} \begin{bmatrix} \mathbf{I}_{n-1} & -\mathbf{R}^T\mathbf{h} \\ \mathbf{0} & 1 \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{I}_{n-1} & \mathbf{0} \\ \mathbf{0} & h_{nn} - \mathbf{h}^T\mathbf{R}\mathbf{R}^T\mathbf{h} \end{bmatrix} \end{aligned}$$

So if we let $\mathbf{U} = \mathbf{S}\mathbf{T}$ and $\alpha = h_{nn} - \mathbf{h}^T\mathbf{R}\mathbf{R}^T\mathbf{h}$, then

$$\mathbf{U}^T\mathbf{H}\mathbf{U} = \begin{bmatrix} 1 & & & \\ & \ddots & & \\ & & 1 & \\ & & & \alpha \end{bmatrix}$$

which implies that

$$(\det \mathbf{U})^2 \det \mathbf{H} = \alpha$$

As $\det \mathbf{H} > 0$, we obtain $\alpha > 0$ and, therefore, $\mathbf{U}^T\mathbf{H}\mathbf{U}$ is positive definite which implies the positive definiteness of $\mathbf{H}$.

(c) The proof of the necessity is similar to the proof of part (b). If

$$\mathbf{d} = [0 \ \cdots \ 0 \ d_{l_1} \ 0 \ \cdots \ 0 \ d_{l_2} \ 0 \ \cdots \ 0 \ d_{l_i} \ 0 \ \cdots \ 0]^T$$

and $\mathbf{H}$ is positive semidefinite, then

$$\mathbf{d}^T\mathbf{H}\mathbf{d} = \mathbf{d}_0^T\mathbf{H}_i^{(l)}\mathbf{d}_0 \geq 0$$

for all $\mathbf{d}_0 \neq \mathbf{0}$ where

$$\mathbf{d}_0 = [d_{l_1}\ d_{l_2}\ \cdots\ d_{l_i}]^T$$

and $\mathbf{H}_i^{(l)}$ is an $i \times i$ principal submatrix. Hence $\mathbf{H}_i^{(l)}$ is positive semidefinite for all possible selections of rows (and columns) from the set $l = \{l_1,\ l_2,\ \ldots,\ l_i,\}$ with $1 \leq l_1 \leq l_2 < \ldots\ < l_i \leq n\}$ and $i = 1,\ 2,\ \ldots,\ n$. Now from part (a)

$$\det\,(\mathbf{H}_i^l) \geq 0 \qquad \text{for}\ \ 1,\ 2,\ \ldots,\ n.$$

The proof of sufficiency is rather lengthy and is omitted. The interested reader is referred to Chap. 7 of [3].

(d) If $\mathbf{H}_i$ is negative definite, then $-\mathbf{H}_i$ is positive definite by definition and hence the proof of part (b) applies to part (d).

(e) If $\mathbf{H}_i^{(l)}$ is negative semidefinite, then $-\mathbf{H}_i^{(l)}$ is positive semidefinite by definition and hence the proof of part (c) applies to part (e).

(f) If neither part (c) nor part (e) holds, then $\mathbf{d}^T\mathbf{H}\mathbf{d}$ can be positive or negative and hence $\mathbf{H}$ is indefinite.

■

Example 2.8 Characterize the Hessian matrices in Examples 2.6 and 2.7 by using the determinant method.

Solution Let

$$\Delta_i = \det\,(\mathbf{H}_i)$$

be the leading principal minors of $\mathbf{H}$. From Example 2.6, we have

$$\Delta_1 = 1, \quad \Delta_2 = -2, \quad \Delta_3 = -18$$

and if $\Delta_i^{'} = \det\,(-\mathbf{H}_i)$, then

$$\Delta_1^{'} = -1, \quad \Delta_2^{'} = -2, \quad \Delta_3^{'} = 18$$

since

$$\det\,(-\mathbf{H}_i) = (-1)^i \det\,(\mathbf{H}_i)$$

Hence $\mathbf{H}$ is indefinite.

From Example 2.7, we get

$$\Delta_1 = 4, \quad \Delta_2 = 8, \quad \Delta_3 = 400$$

Hence $\mathbf{H}$ is positive definite.

■

Example 2.9 Find and classify the stationary points of

$$f(\mathbf{x}) = x_1^2 + 2x_1x_2 + 2x_2^2 + 2x_1 + x_2$$

Solution The first partial derivatives of $f(\mathbf{x})$ are

$$\frac{\partial f}{\partial x_1} = 2x_1 + 2x_2 + 2$$

$$\frac{\partial f}{\partial x_2} = 2x_1 + 4x_2 + 1$$

If $\mathbf{g} = \mathbf{0}$, then

$$2x_1 + 2x_2 + 2 = 0$$

$$2x_1 + 4x_2 + 1 = 0$$

and so there is a stationary point at

$$\mathbf{x} = \mathbf{x}_1 = [-\tfrac{3}{2}\ \tfrac{1}{2}]^T$$

The Hessian is deduced as

$$\mathbf{H} = \begin{bmatrix} 2 & 2 \\ 2 & 4 \end{bmatrix}$$

and since $\Delta_1 = 2$ and $\Delta_2 = 4$, $\mathbf{H}$ is positive definite. Therefore, $\mathbf{x}_1$ is a minimizer.

■

Example 2.10 Find and classify the stationary points of function

$$f(\mathbf{x}) = x_1^2 - x_2^2 + x_3^2 - 2x_1x_3 - x_2x_3 + 4x_1 + 12$$

Solution The first-order partial derivatives of $f(\mathbf{x})$ are

$$\frac{\partial f}{\partial x_1} = 2x_1 - 2x_3 + 4$$
$$\frac{\partial f}{\partial x_2} = -2x_2 - x_3$$
$$\frac{\partial f}{\partial x_3} = -2x_1 - x_2 + 2x_3$$

On equating the gradient to zero and then solving the simultaneous equations obtained, the stationary point $\mathbf{x}_1 = [-10\ 4\ -8]^T$ can be deduced. The Hessian is

$$\mathbf{H} = \begin{bmatrix} 2 & 0 & -2 \\ 0 & -2 & -1 \\ -2 & -1 & 2 \end{bmatrix}$$

and since $\Delta_1 = 2,\ \Delta_2 = -4,\ \Delta_3 = -2$, and $\Delta_1^{'} = -2,\ \Delta_2^{'} = -4,\ \Delta_3^{'} = 2$, $\mathbf{H}$ is indefinite. Therefore, point $\mathbf{x}_1 = [-10\ 4\ \ -8]^T$ is a saddle point. The solution can be readily checked by diagonalizing $\mathbf{H}$ as

$$\hat{\mathbf{H}} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 2\frac{1}{2} \end{bmatrix}$$

■

2.7 Convex and Concave Functions

Usually, in practice, the function to be minimized has several extremum points and, consequently, the uncertainty arises as to whether the extremum point located by an optimization algorithm is the global one. In a specific class of functions referred to as convex and concave functions, any local extremum point is also a global extremum point. Therefore, if the objective function is convex or concave, optimality can be assured. The basic principles relating to convex and concave functions entail a collection of definitions and theorems.

Definition 2.7

A set $\mathcal{R}_c \subset E^n$ is said to be *convex* if for every pair of points $\mathbf{x}_1,\ \mathbf{x}_2 \subset \mathcal{R}_c$ and for every real number α in the range $0 < \alpha < 1$, the point

$$\mathbf{x} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$$

is located in $\mathcal{R}_c$, i.e., $\mathbf{x} \in \mathcal{R}_c$.

■

In effect, if any two points $\mathbf{x}_1,\ \mathbf{x}_2 \in \mathcal{R}_c$ are connected by a straight line, then $\mathcal{R}_c$ is convex if every point on the line segment between $\mathbf{x}_1$ and $\mathbf{x}_2$ is a member of $\mathcal{R}_c$. If some points on the line segment between $\mathbf{x}_1$ and $\mathbf{x}_2$ are not in $\mathcal{R}_c$, the set is said to be *nonconvex*. Convexity in sets is illustrated in Fig. 2.5.

The concept of convexity can also be applied to functions.

Definition 2.8

(a) A function $f(\mathbf{x})$ defined over a convex set $\mathcal{R}_c$ is said to be convex if for every pair of points $\mathbf{x}_1,\ \mathbf{x}_2 \in \mathcal{R}_c$ and every real number α in the range $0 < \alpha < 1$, the inequality

$$f[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] \le \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \tag{2.9}$$

holds. If $\mathbf{x}_1 \neq \mathbf{x}_2$ and

$$f[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] < \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2)$$

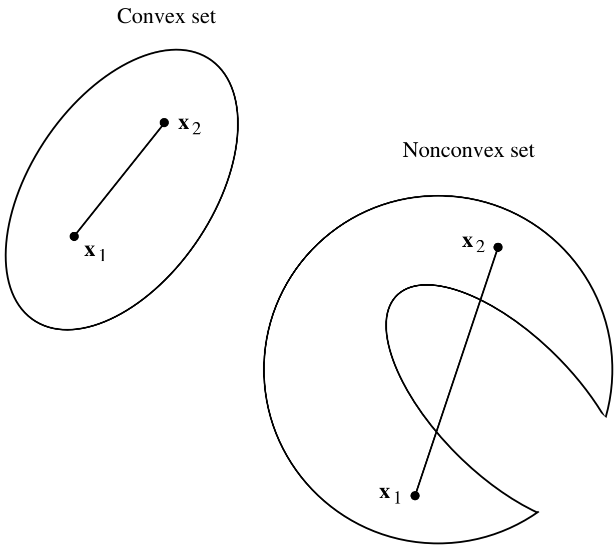

Figure 2.5. Convexity in sets.

then $f(\mathbf{x})$ is said to be *strictly convex*.

(b) If $\phi(\mathbf{x})$ is defined over a convex set $\mathcal{R}_c$ and $f(\mathbf{x}) = -\phi(\mathbf{x})$ is convex, then $\phi(\mathbf{x})$ is said to be *concave*. If $f(\mathbf{x})$ is strictly convex, $\phi(\mathbf{x})$ is *strictly concave*.

■

In the left-hand side of Eq. (2.9), function $f(\mathbf{x})$ is evaluated on the line segment joining points $\mathbf{x}_1$ and $\mathbf{x}_2$ whereas in the right-hand side of Eq. (2.9) an approximate value is obtained for $f(\mathbf{x})$ based on linear interpolation. Thus a function is convex if linear interpolation between any two points overestimates the value of the function. The functions shown in Fig. 2.6a and b are convex whereas that in Fig. 2.6c is nonconvex.

Theorem 2.10 ***Convexity of linear combination of convex functions*** *If*

$$f(\mathbf{x}) = af_1(\mathbf{x}) + bf_2(\mathbf{x})$$

where $a,\ b \geq 0$ *and* $f_1(\mathbf{x}), f_2(\mathbf{x})$ *are convex functions on the convex set* $\mathcal{R}_c$*, then* $f(\mathbf{x})$ *is convex on the set* $\mathcal{R}_c$*.*

Proof Since $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are convex, and $a,\ b \geq 0$, then for $\mathbf{x} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$ we have

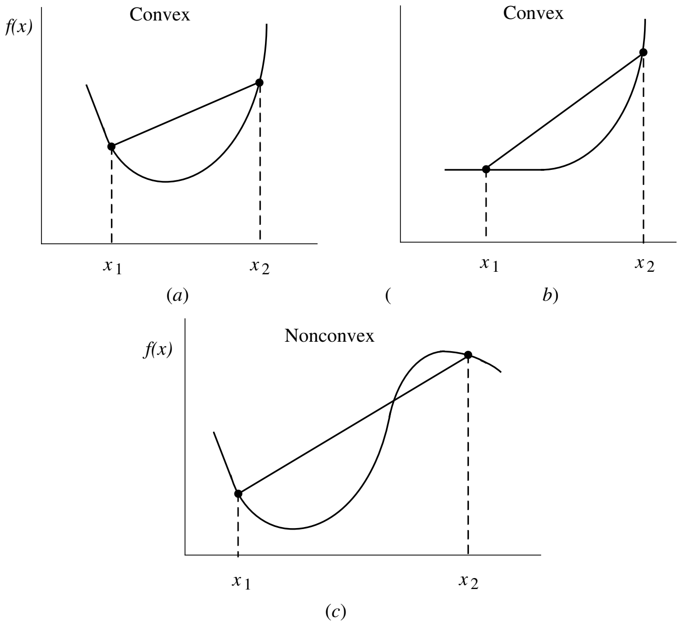

Figure 2.6. Convexity in functions.

$$af_1[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] \leq a[\alpha f_1(\mathbf{x}_1) + (1-\alpha)f_1(\mathbf{x}_2)]$$

$$bf_2[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] \leq b[\alpha f_2(\mathbf{x}_1) + (1-\alpha)f_2(\mathbf{x}_2)]$$

where $0 < \alpha < 1$. Hence

$$\begin{aligned} f(\mathbf{x}) &= af_1(\mathbf{x}) + bf_2(\mathbf{x}) \\ f[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] &= af_1[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] + bf_2[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] \\ &\leq \alpha[af_1(\mathbf{x}_1) + bf_2(\mathbf{x}_1)] + (1-\alpha)[af_1(\mathbf{x}_2) \\ &\quad +bf_2(\mathbf{x}_2)] \end{aligned}$$

Since

$$\begin{aligned} af_1(\mathbf{x}_1) + bf_2(\mathbf{x}_1) &= f(\mathbf{x}_1) \\ af_1(\mathbf{x}_2) + bf_2(\mathbf{x}_2) &= f(\mathbf{x}_2) \end{aligned}$$

the above inequality can be expressed as

$$f[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] \leq \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2)$$

that is, $f(\mathbf{x})$ is convex.

■

Theorem 2.11 ***Relation between convex functions and convex sets*** *If $f(\mathbf{x})$ is a convex function on a convex set $\mathcal{R}_c$, then the set*

$$\mathcal{S}_c = \{\mathbf{x} :\ \mathbf{x} \in \mathcal{R}_c,\ f(\mathbf{x}) \le K\}$$

is convex for every real number K.

Proof If $\mathbf{x}_1,\ \mathbf{x}_2 \in \mathcal{S}_c$, then $f(\mathbf{x}_1) \le K$ and $f(\mathbf{x}_2) \le K$ from the definition of $\mathcal{S}_c$. Since $f(\mathbf{x})$ is convex

$$\begin{aligned} f[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] &\le \alpha f(\mathbf{x}_1) + (1-\alpha) f(\mathbf{x}_2) \\ &\le \alpha K + (1-\alpha) K \end{aligned}$$

or

$$f(\mathbf{x}) \le K \qquad \text{for } \mathbf{x} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 \quad \text{and} \quad 0 < \alpha < 1$$

Therefore

$$\mathbf{x} \in \mathcal{S}_c$$

that is, $\mathcal{S}_c$ is convex by virtue of Def. 2.7.

■

Theorem 2.11 is illustrated in Fig. 2.7, where set $\mathcal{S}_C$ is convex if $f(\mathbf{x})$ is a convex function on convex set $\mathcal{R}_c$.

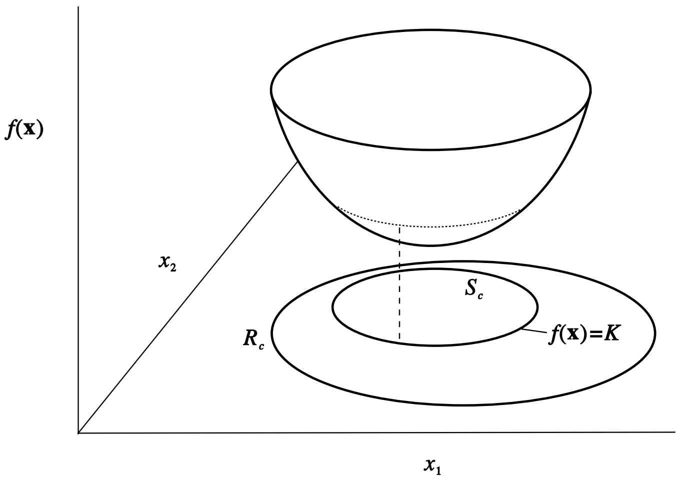

Figure 2.7. Graphical construction for Theorem 2.11.

An alternative view of convexity can be generated by examining some theorems which involve the gradient and Hessian of $f(\mathbf{x})$.

Theorem 2.12 ***Property of convex functions relating to gradient*** *If $f(\mathbf{x}) \in C^1$, then $f(\mathbf{x})$ is convex over a convex set $\mathcal{R}_c$ if and only if*

$$f(\mathbf{x}_1) \geq f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T(\mathbf{x}_1 - \mathbf{x})$$

for all $\mathbf{x}$ and $\mathbf{x}_1 \in \mathcal{R}_c$, where $\mathbf{g}(\mathbf{x})$ is the gradient of $f(\mathbf{x})$.

Proof The proof of this theorem consists of two parts. First we prove that if $f(\mathbf{x})$ is convex, the inequality holds. Then we prove that if the inequality holds, $f(\mathbf{x})$ is convex. The two parts constitute the necessary and sufficient conditions of the theorem. If $f(\mathbf{x})$ is convex, then for all α in the range $0 < \alpha < 1$

$$f[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}] \leq \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x})$$

or

$$f[\mathbf{x} + \alpha(\mathbf{x}_1 - \mathbf{x})] - f(\mathbf{x}) \leq \alpha[f(\mathbf{x}_1) - f(\mathbf{x})]$$

As $\alpha \to 0$, the Taylor series of $f[\mathbf{x} + \alpha(\mathbf{x}_1 - \mathbf{x})]$ yields

$$f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T\alpha(\mathbf{x}_1 - \mathbf{x}) - f(\mathbf{x}) \leq \alpha[f(\mathbf{x}_1) - f(\mathbf{x})]$$

and so

$$f(\mathbf{x}_1) \geq f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T(\mathbf{x}_1 - \mathbf{x}) \tag{2.10}$$

Now if this inequality holds at points $\mathbf{x}$ and $\mathbf{x}_2 \in \mathcal{R}_c$, then

$$f(\mathbf{x}_2) \geq f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T(\mathbf{x}_2 - \mathbf{x}) \tag{2.11}$$

Hence Eqs. (2.10) and (2.11) yield

$$\begin{aligned}\alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \geq{}& \alpha f(\mathbf{x}) + \alpha\mathbf{g}(\mathbf{x})^T(\mathbf{x}_1 - \mathbf{x}) + (1-\alpha)f(\mathbf{x}) \\ &+ (1-\alpha)\mathbf{g}(\mathbf{x})^T(\mathbf{x}_2 - \mathbf{x})\end{aligned}$$

or

$$\alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \geq f(\mathbf{x}) + \mathbf{g}^T(\mathbf{x})[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2 - \mathbf{x}]$$

With the substitution

$$\mathbf{x} = \alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2$$

we obtain

$$f[\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] \leq \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2)$$

for $0 < \alpha < 1$. Therefore, from Def. 2.8 $f(\mathbf{x})$ is convex.

■

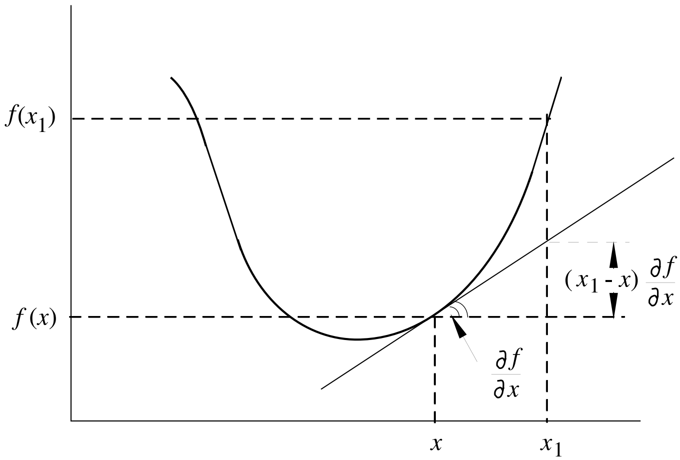

Figure 2.8. Graphical construction for Theorem 2.12.

Theorem 2.12 states that a linear approximation of $f(\mathbf{x})$ at point $\mathbf{x}_1$ based on the derivatives of $f(\mathbf{x})$ at $\mathbf{x}$ underestimates the value of the function. This fact is illustrated in Fig. 2.8.

Theorem 2.13 ***Property of convex functions relating to the Hessian*** *A function $f(\mathbf{x}) \in C^2$ is convex over a convex set $\mathcal{R}_c$ if and only if the Hessian $H(\mathbf{x})$ of $f(\mathbf{x})$ is positive semidefinite for $\mathbf{x} \in \mathcal{R}_c$.*

Proof If $\mathbf{x}_1 = \mathbf{x} + \mathbf{d}$ where $\mathbf{x}_1$ and $\mathbf{x}$ are arbitrary points in $\mathcal{R}_c$, then the Taylor series yields

$$f(\mathbf{x}_1) = f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T(\mathbf{x}_1 - \mathbf{x}) + \tfrac{1}{2}\mathbf{d}^T\mathbf{H}(\mathbf{x} + \alpha\mathbf{d})\mathbf{d} \tag{2.12}$$

where $0 \leq \alpha \leq 1$ (see Eq. (2.4h)). Now if $\mathbf{H}(\mathbf{x})$ is positive semidefinite everywhere in $\mathcal{R}_c$, then

$$\tfrac{1}{2}\mathbf{d}^T\mathbf{H}(\mathbf{x} + \alpha\mathbf{d})\mathbf{d} \geq 0$$

and so

$$f(\mathbf{x}_1) \geq f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T(\mathbf{x}_1 - \mathbf{x})$$

Therefore, from Theorem 2.12, $f(\mathbf{x})$ is convex.

If $\mathbf{H}(\mathbf{x})$ is not positive semidefinite everywhere in $\mathcal{R}_c$, then a point $\mathbf{x}$ and at least a $\mathbf{d}$ exist such that

$$\mathbf{d}^T\mathbf{H}(\mathbf{x} + \alpha\mathbf{d})\mathbf{d} < 0$$

and so Eq. (2.12) yields

$$f(\mathbf{x}_1) < f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T(\mathbf{x}_1 - \mathbf{x})$$

and $f(\mathbf{x})$ is nonconvex from Theorem 2.12. Therefore, $f(\mathbf{x})$ is convex if and only if $H(\mathbf{x})$ is positive semidefinite everywhere in $\mathcal{R}_c$.

■

For a strictly convex function, Theorems 2.11–2.13 are modified as follows.

Theorem 2.14 ***Properties of strictly convex functions***

(a) If $f(\mathbf{x})$ is a strictly convex function on a convex set $\mathcal{R}_c$, then the set

$$\mathcal{S}_c = \{\mathbf{x} :\ \mathbf{x} \in \mathcal{R}_c \text{ for } f(\mathbf{x}) < K\}$$

is convex for every real number K.

(b) If $f(\mathbf{x}) \in C^1$, then $f(\mathbf{x})$ is strictly convex over a convex set if and only if

$$f(\mathbf{x}_1) > f(\mathbf{x}) + \mathbf{g}(\mathbf{x})^T(\mathbf{x}_1 - \mathbf{x})$$

for all $\mathbf{x}$ and $\mathbf{x}_1 \in \mathcal{R}_c$ where $\mathbf{g}(\mathbf{x})$ is the gradient of $f(\mathbf{x})$.

(c) A function $f(\mathbf{x}) \in C^2$ is strictly convex over a convex set $\mathcal{R}_c$ if and only if the Hessian $\mathbf{H}(\mathbf{x})$ is positive definite for $\mathbf{x} \in \mathcal{R}_c$.

If the second-order sufficient conditions for a minimum hold at $\mathbf{x}^*$ as in Theorem 2.4, in which case $\mathbf{x}^*$ is a strong local minimizer, then from Theorem 2.14(c), $f(\mathbf{x})$ must be strictly convex in the neighborhood of $\mathbf{x}^*$. Consequently, convexity assumes considerable importance even though the class of convex functions is quite restrictive.

If $\phi(\mathbf{x})$ is defined over a convex set $\mathcal{R}_c$ and $f(\mathbf{x}) = -\phi(\mathbf{x})$ is strictly convex, then $\phi(\mathbf{x})$ is strictly concave and the Hessian of $\phi(\mathbf{x})$ is negative definite. Conversely, if the Hessian of $\phi(\mathbf{x})$ is negative definite, then $\phi(\mathbf{x})$ is strictly concave.

Example 2.11 Check the following functions for convexity:

(a) $f(\mathbf{x}) = e^{x_1} + x_2^2 + 5$

(b) $f(\mathbf{x}) = 3x_1^2 - 5x_1x_2 + x_2^2$

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

(d) $f(\mathbf{x}) = 50 + 10x_1 + x_2 - 6x_1^2 - 3x_2^2$

Solution In each case the problem reduces to the derivation and characterization of the Hessian $\mathbf{H}$.

(a) The Hessian can be obtained as

$$\mathbf{H} = \begin{bmatrix} e^{x_1} & 0 \\ 0 & 2 \end{bmatrix}$$

For $-\infty < x_1 < \infty$, $\mathbf{H}$ is positive definite and $f(\mathbf{x})$ is strictly convex.

(b) In this case, we have

$$\mathbf{H} = \begin{bmatrix} 6 & -5 \\ -5 & 2 \end{bmatrix}$$

Since $\Delta_1 = 6,\ \Delta_2 = -13$ and $\Delta_1^{'} = -6,\ \Delta_2^{'} = -13$, where $\Delta_i = \det\,(\mathbf{H}_i)$ and $\Delta_i^{'} = \det\,(-\mathbf{H}_i)$, $\mathbf{H}$ is indefinite. Thus $f(\mathbf{x})$ is neither convex nor concave.

(c) For this example, we get

$$\mathbf{H} = \begin{bmatrix} 3x_1^2 - 2 & 0 \\ 0 & 2 \end{bmatrix}$$

For $x_1 \leq -\sqrt{2/3}$ and $x_1 \geq \sqrt{2/3}$, $\mathbf{H}$ is positive semidefinite and $f(\mathbf{x})$ is convex; for $x_1 < -\sqrt{2/3}$ and $x_1 > \sqrt{2/3}$, $\mathbf{H}$ is positive definite and $f(\mathbf{x})$ is strictly convex; for $-\sqrt{2/3} < x_1 < \sqrt{2/3}$, $\mathbf{H}$ is indefinite, and $f(\mathbf{x})$ is neither convex nor concave.

(d) As before

$$\mathbf{H} = \begin{bmatrix} -12 & 0 \\ 0 & -6 \end{bmatrix}$$

In this case $\mathbf{H}$ is negative definite, and $f(\mathbf{x})$ is strictly concave.

■

2.8 Optimization of Convex Functions

The above theorems and results can now be used to deduce the following three important theorems.

Theorem 2.15 ***Relation between local and global minimizers in convex functions*** *If $f(\mathbf{x})$ is a convex function defined on a convex set $\mathcal{R}_c$, then*

(a) *the set of points $\mathcal{S}_c$ where $f(\mathbf{x})$ is minimum is convex;*

(b) *any local minimizer of $f(\mathbf{x})$ is a global minimizer.*

Proof (a) If F^* is a minimum of $f(\mathbf{x})$, then $\mathcal{S}_c = \{\mathbf{x} :\ f(\mathbf{x}) \leq F^*,\ \mathbf{x} \in \mathcal{R}_c\}$ is convex by virtue of Theorem 2.11.

(b) If $\mathbf{x}^* \in \mathcal{R}_c$ is a local minimizer but there is another point $\mathbf{x}^{**} \in \mathcal{R}_c$ which is a global minimizer such that

$$f(\mathbf{x}^{**}) < f(\mathbf{x}^*)$$

then on line $\mathbf{x} = \alpha\mathbf{x}^{**} + (1-\alpha)\mathbf{x}^*$

$$\begin{aligned} f[\alpha\mathbf{x}^{**} + (1-\alpha)\mathbf{x}^*] &\leq \alpha f(\mathbf{x}^{**}) + (1-\alpha)f(\mathbf{x}^*) \\ &< \alpha f(\mathbf{x}^*) + (1-\alpha)f(\mathbf{x}^*) \end{aligned}$$

or

$$f(\mathbf{x}) < f(\mathbf{x}^*) \qquad \text{for} \quad \text{all } \alpha$$

This contradicts the fact that $\mathbf{x}^*$ is a local minimizer and so

$$f(\mathbf{x}) \geq f(\mathbf{x}^*)$$

for all $\mathbf{x} \in \mathcal{R}_c$. Therefore, any local minimizers are located in a convex set, and all are global minimizers.

■

Theorem 2.16 ***Existence of a global minimizer in convex functions*** *If $f(\mathbf{x}) \in C^1$ is a convex function on a convex set $\mathcal{R}_c$ and there is a point $\mathbf{x}^*$ such that*

$$\mathbf{g}(x^*)^T\mathbf{d} \geq 0 \qquad \text{where } \mathbf{d} = \mathbf{x}_1 - \mathbf{x}^*$$

for all $\mathbf{x}_1 \in \mathcal{R}_c$, then $\mathbf{x}^*$ is a global minimizer of $f(\mathbf{x})$.

Proof From Theorem 2.12

$$f(\mathbf{x}_1) \geq f(\mathbf{x}^*) + \mathbf{g}(\mathbf{x}^*)^T(\mathbf{x}_1 - \mathbf{x}^*)$$

where $\mathbf{g}(\mathbf{x}^*)$ is the gradient of $f(\mathbf{x})$ at $\mathbf{x} = \mathbf{x}^*$. Since

$$\mathbf{g}(\mathbf{x}^*)^T(\mathbf{x}_1 - \mathbf{x}^*) \geq 0$$

we have

$$f(\mathbf{x}_1) \geq f(\mathbf{x}^*)$$

and so $\mathbf{x}^*$ is a local minimizer. By virtue of Theorem 2.15, $\mathbf{x}^*$ is also a global minimizer.

Similarly, if $f(\mathbf{x})$ is a strictly convex function and

$$\mathbf{g}(\mathbf{x}^*)^T\mathbf{d} > 0$$

then $\mathbf{x}^*$ is a strong global minimizer.

■

The above theorem states, in effect, that if $f(\mathbf{x})$ is convex, then the first-order necessary conditions become sufficient for $\mathbf{x}^*$ to be a global minimizer.

Since a convex function of one variable is in the form of the letter U whereas a convex function of two variables is in the form of a bowl, there are no theorems analogous to Theorems 2.15 and 2.16 pertaining to the maximization of a convex function. However, the following theorem, which is intuitively plausible, is sometimes useful.

Theorem 2.17 ***Location of maximum of a convex function*** *If $f(\mathbf{x})$ is a convex function defined on a bounded, closed, convex set $\mathcal{R}_c$, then if $f(\mathbf{x})$ has a maximum over $\mathcal{R}_c$, it occurs at the boundary of $\mathcal{R}_c$.*

Proof If point $\mathbf{x}$ is in the interior of $\mathcal{R}_c$, a line can be drawn through $\mathbf{x}$ which intersects the boundary at two points, say, $\mathbf{x}_1$ and $\mathbf{x}_2$, since $\mathcal{R}_c$ is bounded and closed. Since $f(\mathbf{x})$ is convex, some α exists in the range $0 < \alpha < 1$ such that

$$\mathbf{x} = \alpha\mathbf{x}_1 + (1 - \alpha)\mathbf{x}_2$$

and

$$f(\mathbf{x}) \leq \alpha f(\mathbf{x}_1) + (1 - \alpha) f(\mathbf{x}_2)$$

If $f(\mathbf{x}_1) > f(\mathbf{x}_2)$, we have

$$\begin{aligned} f(\mathbf{x}) &< \alpha f(\mathbf{x}_1) + (1 - \alpha) f(\mathbf{x}_1) \\ &= f(\mathbf{x}_1) \end{aligned}$$

If

$$f(\mathbf{x}_1) < f(\mathbf{x}_2)$$

we obtain

$$\begin{aligned} f(\mathbf{x}) &< \alpha f(\mathbf{x}_2) + (1 - \alpha) f(\mathbf{x}_2) \\ &= f(\mathbf{x}_2) \end{aligned}$$

Now if

$$f(\mathbf{x}_1) = f(\mathbf{x}_2)$$

the result

$$f(\mathbf{x}) \leq f(\mathbf{x}_1) \quad \text{and} \quad f(\mathbf{x}) \leq f(\mathbf{x}_2)$$

is obtained. Evidently, in all possibilities the maximizers occur on the boundary of $\mathcal{R}_c$.

■

This theorem is illustrated in Fig. 2.9.

References

1 M. H. Protter and C. B. Morrey, Jr., *Modern Mathematical Analysis*, Addison-Wesley, Reading, MA, 1964.
2 D. G. Luenberger, *Linear and Nonlinear Programming*, 2nd ed., Addison-Wesley, Reading, MA, 1984.
3 R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, Cambridge University Press, UK, 1990.

Problems

2.1 (a) Obtain a quadratic approximation for the function

$$f(\mathbf{x}) = 2x_1^3 + x_2^2 + x_1^2 x_2^2 + 4x_1 x_2 + 3$$

at point $\mathbf{x} + \boldsymbol{\delta}$ if $\mathbf{x}^T = [1\ 1]$.

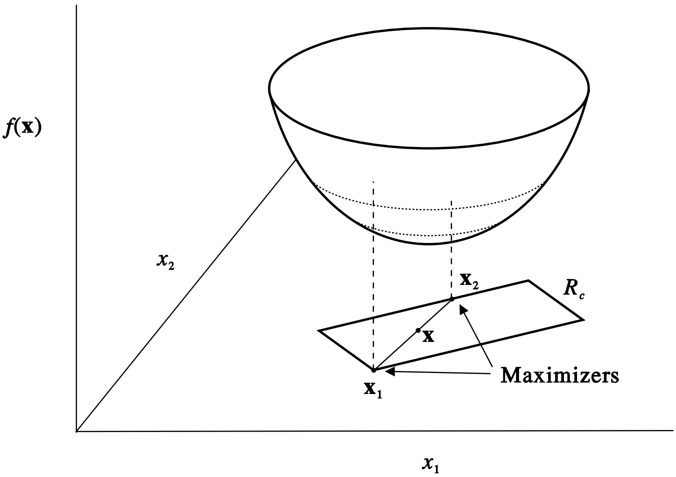

Figure 2.9. Graphical construction for Theorem 2.17.

(b) Now obtain a linear approximation.

2.2 An n-variable quadratic function is given by

$$f(\mathbf{x}) = a + \mathbf{b}^T\mathbf{x} + \tfrac{1}{2}\mathbf{x}^T\mathbf{Q}\mathbf{x}$$

where $\mathbf{Q}$ is an $n \times n$ symmetric matrix. Show that the gradient and Hessian of $f(\mathbf{x})$ are given by

$$\mathbf{g} = \mathbf{b} + \mathbf{Q}\mathbf{x} \quad \text{and} \quad \nabla^2 f(\mathbf{x}) = \mathbf{Q}$$

respectively.

2.3 Point $\mathbf{x}_a = [2\ 4]^T$ is a possible minimizer of the problem

$$\text{minimize } f(\mathbf{x}) = \tfrac{1}{4}[x_1^2 + 4x_2^2 - 4(3x_1 + 8x_2) + 100]$$

$$\text{subject to:} \quad x_1 = 2,\ x_2 \geq 0$$

(a) Find the feasible directions.

(b) Check if the second-order necessary conditions are satisfied.

2.4 Points $\mathbf{x}_a = [0\ 3]^T$, $\mathbf{x}_b = [4\ 0]^T$, $\mathbf{x}_c = [4\ 3]^T$ are possible maximizers of the problem

$$\text{maximize } f(\mathbf{x}) = 2(4x_1 + 3x_2) - (x_1^2 + x_2^2 + 25)$$

$$\text{subject to:} \quad x_1 \geq 0,\ x_2 \geq 0$$

(a) Find the feasible directions.

(b) Check if the second-order necessary conditions are satisfied.

2.5 Point $\mathbf{x}_a = [4 \ \ -1]^T$ is a possible minimizer of the problem

$$\text{minimize } f(\mathbf{x}) = \tfrac{16}{x_1} - x_2$$

$$\text{subject to:} \quad x_1 + x_2 = 3, \ x_1 \geq 0$$

(a) Find the feasible directions.

(b) Check if the second-order necessary conditions are satisfied.

2.6 Classify the following matrices as positive definite, positive semidefinite, etc. by using LDL^T factorization:

$$(a)\ \mathbf{H} = \begin{bmatrix} 5 & 3 & 1 \\ 3 & 4 & 2 \\ 1 & 2 & 6 \end{bmatrix}, \quad (b)\ \mathbf{H} = \begin{bmatrix} -5 & 1 & 1 \\ 1 & -2 & 2 \\ 1 & 2 & -4 \end{bmatrix}$$

$$(c)\ \mathbf{H} = \begin{bmatrix} -1 & 2 & -3 \\ 2 & 4 & 5 \\ -3 & 5 & -20 \end{bmatrix}$$

2.7 Check the results in Prob. 2.6 by using the determinant method.

2.8 Classify the following matrices by using the eigenvalue method:

$$(a)\ \mathbf{H} = \begin{bmatrix} 2 & 3 \\ 3 & 4 \end{bmatrix}, \quad (b)\ \mathbf{H} = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 2 & 0 \\ 4 & 0 & 18 \end{bmatrix}$$

2.9 One of the points $\mathbf{x}_a = [1 \ \ -1]^T$, $\mathbf{x}_b = [0 \ 0]^T$, $\mathbf{x}_c = [1 \ 1]^T$ minimizes the function

$$f(\mathbf{x}) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2$$

By using appropriate tests, identify the minimizer.

2.10 An optimization algorithm has given a solution $\mathbf{x}_a = [0.6959 \ -11.3479]^T$ for the problem

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

(a) Classify the general Hessian of $f(\mathbf{x})$ (i.e., positive definite, . . ., etc.).

(b) Determine whether $\mathbf{x}_a$ is a minimizer, maximizer, or saddle point.

2.11 Find and classify the stationary points for the function

$$f(\mathbf{x}) = x_1^2 - x_2^2 + x_3^2 - 2x_1x_3 - x_2x_3 + 4x_1 + 12$$

2.12 Find and classify the stationary points for the following functions:

$$(a)\ f(\mathbf{x}) = 2x_1^2 + x_2^2 - 2x_1x_2 + 2x_1^3 + x_1^4$$
$$(b)\ f(\mathbf{x}) = x_1^2x_2^2 - 4x_1^2x_2 + 4x_1^2 + 2x_1x_2^2 + x_2^2 - 8x_1x_2 + 8x_1 - 4x_2$$

2.13 Show that

$$f(\mathbf{x}) = (x_2 - x_1^2)^2 + x_1^5$$

has only one stationary point which is neither a minimizer or a maximizer.

2.14 Investigate the following functions and determine whether they are convex or concave:

$$\begin{aligned}
(a)\ \ f(\mathbf{x}) &= x_1^2 + \cosh x_2 \\
(b)\ \ f(\mathbf{x}) &= x_1^2 + 2x_2^2 + 2x_3^2 + x_4^2 - x_1x_2 + x_1x_3 - 2x_2x_4 + x_1x_4 \\
(c)\ \ f(\mathbf{x}) &= x_1^2 - 2x_2^2 - 2x_3^2 + x_4^2 - x_1x_2 + x_1x_3 - 2x_2x_4 + x_1x_4
\end{aligned}$$

2.15 A given quadratic function $f(\mathbf{x})$ is known to be convex for $\|\mathbf{x}\| < \varepsilon$. Show that it is convex for all $\mathbf{x} \in E^n$.

2.16 Two functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are convex over a convex set $\mathcal{R}_c$. Show that

$$f(\mathbf{x}) = \alpha f_1(\mathbf{x}) + \beta f_2(\mathbf{x})$$

where α and β are nonnegative scalars is convex over $\mathcal{R}_c$.

2.17 Assume that functions $f_1(\mathbf{x})$ and $f_2(\mathbf{x})$ are convex and let

$$f(\mathbf{x}) = \max\{f_1(\mathbf{x}),\ f_2(\mathbf{x})\}$$

Show that $f(\mathbf{x})$ is a convex function.

2.18 Let $\gamma(t)$ be a single-variable convex function which is monotonic non-decreasing, i.e., $\gamma(t_1) \geq \gamma(t_2)$ for $t_1 > t_2$. Show that the compound function $\gamma[f(\mathbf{x})]$ is convex if $f(\mathbf{x})$ is convex [2].

http://www.springer.com/978-0-387-71106-5

Practical Optimization
Algorithms and Engineering Applications
Antoniou, A.; Lu, W.-S.
2007, XX, 670 p., Hardcover
ISBN: 978-0-387-71106-5