

## CHAPTER 7

# Multidimensional Calculus

In the previous chapters we have mainly dealt with real-valued functions of a single variable  $x$ . In this chapter we extend the notions of limits, continuity, differentiation, and integration to multivariable functions, that is, functions of several variables. These functions can be real-valued or possibly vector-valued. More specifically, if  $R^n$  denotes the  $n$ -dimensional Euclidean space,  $n \geq 1$ , then we shall in general consider functions defined on a set  $D \subset R^n$  and have values in  $R^m$ ,  $m \geq 1$ . Such functions are represented symbolically as  $\mathbf{f}: D \rightarrow R^m$ , where for  $\mathbf{x} = (x_1, x_2, \dots, x_n)' \in D$ ,

$$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_m(\mathbf{x})]'$$

and  $f_i(\mathbf{x})$  is a real-valued function of  $x_1, x_2, \dots, x_n$  ( $i = 1, 2, \dots, m$ ).

Even though the basic framework of the methodology in this chapter is general and applies in any number of dimensions, most of the examples are associated with two- or three-dimensional spaces. At this stage, it would be helpful to review the basic concepts given in Chapters 1 and 2. This can facilitate the understanding of the methodology and its development in a multidimensional environment.

### 7.1. SOME BASIC DEFINITIONS

Some of the concepts described in Chapter 1 pertained to one-dimensional Euclidean spaces. In this section we extend these concepts to higher-dimensional Euclidean spaces.

Any point  $\mathbf{x}$  in  $R^n$  can be represented as a column vector of the form  $(x_1, x_2, \dots, x_n)'$ , where  $x_i$  is the  $i$ th element of  $\mathbf{x}$  ( $i = 1, 2, \dots, n$ ). The Euclidean norm of  $\mathbf{x}$  was defined in Chapter 2 (see Definition 2.1.4) as  $\|\mathbf{x}\|_2 = (\sum_{i=1}^n x_i^2)^{1/2}$ . For simplicity we shall drop the subindex 2 and denote this norm by  $\|\mathbf{x}\|$ .

Let  $\mathbf{x}_0 \in R^n$ . A neighborhood  $N_r(\mathbf{x}_0)$  of  $\mathbf{x}_0$  is a set of points in  $R^n$  that lie within some distance, say  $r$ , from  $\mathbf{x}_0$ , that is,

$$N_r(\mathbf{x}_0) = \{\mathbf{x} \in R^n \mid \|\mathbf{x} - \mathbf{x}_0\| < r\}.$$

If  $\mathbf{x}_0$  is deleted from  $N_r(\mathbf{x}_0)$ , we obtain the so-called deleted neighborhood of  $\mathbf{x}_0$ , which we denote by  $N_r^d(\mathbf{x}_0)$ .

A point  $\mathbf{x}_0$  in  $R^n$  is a limit point of a set  $A \subset R^n$  if every neighborhood of  $\mathbf{x}_0$  contains an element  $\mathbf{x}$  of  $A$  such that  $\mathbf{x} \neq \mathbf{x}_0$ , that is, every deleted neighborhood of  $\mathbf{x}_0$  contains points of  $A$ .

A set  $A \subset R^n$  is closed if every limit point of  $A$  belongs to  $A$ .

A point  $\mathbf{x}_0$  in  $R^n$  is an interior of a set  $A \subset R^n$  if there exists an  $r > 0$  such that  $N_r(\mathbf{x}_0) \subset A$ .

A set  $A \subset R^n$  is open if for every point  $\mathbf{x}$  in  $A$  there exists a neighborhood  $N_r(\mathbf{x})$  that is contained in  $A$ . Thus  $A$  is open if it consists entirely of interior points.

A point  $p \in R^n$  is a boundary point of a set  $A \subset R^n$  if every neighborhood of  $p$  contains points of  $A$  as well as points of  $\bar{A}$ , the complement of  $A$  with respect to  $R^n$ . The set of all boundary points of  $A$  is called its boundary and is denoted by  $Br(A)$ .

A set  $A \subset R^n$  is bounded if there exists an  $r > 0$  such that  $\|\mathbf{x}\| \leq r$  for all  $\mathbf{x}$  in  $A$ .

Let  $\mathbf{g}: J^+ \rightarrow R^n$  be a vector-valued function defined on the set of all positive integers. Let  $\mathbf{g}(i) = \mathbf{a}_i$ ,  $i \geq 1$ . Then  $\{\mathbf{a}_i\}_{i=1}^\infty$  represents a sequence of points in  $R^n$ . By a subsequence of  $\{\mathbf{a}_i\}_{i=1}^\infty$  we mean a sequence  $\{\mathbf{a}_{k_i}\}_{i=1}^\infty$  such that  $k_1 < k_2 < \cdots < k_i < \cdots$  and  $k_i \geq i$  for  $i \geq 1$  (see Definition 5.1.1).

A sequence  $\{\mathbf{a}_i\}_{i=1}^\infty$  converges to a point  $\mathbf{c} \in R^n$  if for a given  $\epsilon > 0$  there exists an integer  $N$  such that  $\|\mathbf{a}_i - \mathbf{c}\| < \epsilon$  whenever  $i > N$ . This is written symbolically as  $\lim_{i \rightarrow \infty} \mathbf{a}_i = \mathbf{c}$ , or  $\mathbf{a}_i \rightarrow \mathbf{c}$  as  $i \rightarrow \infty$ .

A sequence  $\{\mathbf{a}_i\}_{i=1}^\infty$  is bounded if there exists a number  $K > 0$  such that  $\|\mathbf{a}_i\| \leq K$  for all  $i$ .

## 7.2. LIMITS OF A MULTIVARIABLE FUNCTION

We recall from Chapter 3 that for a function of a single variable  $x$ , its limit at a point is considered when  $x$  approaches the point from two directions, left and right. Here, for a function of several variables, say  $x_1, x_2, \dots, x_n$ , its limit at a point  $\mathbf{a} = (a_1, a_2, \dots, a_n)'$  is considered when  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$  approaches  $\mathbf{a}$  in any possible way. Thus when  $n > 1$  there are infinitely many ways in which  $\mathbf{x}$  can approach  $\mathbf{a}$ .

**Definition 7.2.1.** Let  $\mathbf{f}: D \rightarrow R^m$ , where  $D \subset R^n$ . Then  $\mathbf{f}(\mathbf{x})$  is said to have a limit  $\mathbf{L} = (L_1, L_2, \dots, L_m)'$  as  $\mathbf{x}$  approaches  $\mathbf{a}$ , written symbolically as  $\mathbf{x} \rightarrow \mathbf{a}$ , where  $\mathbf{a}$  is a limit point of  $D$ , if for a given  $\epsilon > 0$  there exists a  $\delta > 0$  such

that  $\|\mathbf{f}(\mathbf{x}) - \mathbf{L}\| < \epsilon$  for all  $\mathbf{x}$  in  $D \cap N_\delta^d(\mathbf{a})$ , where  $N_\delta^d(\mathbf{a})$  is a deleted neighborhood of  $\mathbf{a}$  of radius  $\delta$ . If it exists, this limit is written symbolically as  $\lim_{\mathbf{x} \rightarrow \mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{L}$ .  $\square$

Note that whenever a limit of  $\mathbf{f}(\mathbf{x})$  exists, its value must be the same no matter how  $\mathbf{x}$  approaches  $\mathbf{a}$ . It is important here to understand the meaning of “ $\mathbf{x}$  approaches  $\mathbf{a}$ .” By this we do not necessarily mean that  $\mathbf{x}$  moves along a straight line leading into  $\mathbf{a}$ . Rather, we mean that  $\mathbf{x}$  moves closer and closer to  $\mathbf{a}$  along any curve that goes through  $\mathbf{a}$ .

EXAMPLE 7.2.1. Consider the behavior of the function

$$f(x_1, x_2) = \frac{x_1^3 - x_2^3}{x_1^2 + x_2^2}$$

as  $\mathbf{x} = (x_1, x_2)' \rightarrow \mathbf{0}$ , where  $\mathbf{0} = (0, 0)'$ . This function is defined everywhere in  $R^2$  except at  $\mathbf{0}$ . It is convenient here to represent the point  $\mathbf{x}$  using polar coordinates,  $r$  and  $\theta$ , such that  $x_1 = r \cos \theta$ ,  $x_2 = r \sin \theta$ ,  $r > 0$ ,  $0 \leq \theta \leq 2\pi$ . We then have

$$\begin{aligned} f(x_1, x_2) &= \frac{r^3 \cos^3 \theta - r^3 \sin^3 \theta}{r^2 \cos^2 \theta + r^2 \sin^2 \theta} \\ &= r(\cos^3 \theta - \sin^3 \theta). \end{aligned}$$

Since  $\mathbf{x} \rightarrow \mathbf{0}$  if and only if  $r \rightarrow 0$ ,  $\lim_{\mathbf{x} \rightarrow \mathbf{0}} f(x_1, x_2) = 0$  no matter how  $\mathbf{x}$  approaches  $\mathbf{0}$ .

EXAMPLE 7.2.2. Consider the function

$$f(x_1, x_2) = \frac{x_1 x_2}{x_1^2 + x_2^2}.$$

Using polar coordinates again, we obtain

$$f(x_1, x_2) = \cos \theta \sin \theta,$$

which depends on  $\theta$ , but not on  $r$ . Since  $\theta$  can have infinitely many values,  $f(x_1, x_2)$  cannot be made close to any one constant  $L$  no matter how small  $r$  is. Thus the limit of this function does not exist as  $\mathbf{x} \rightarrow \mathbf{0}$ .

EXAMPLE 7.2.3. Let  $f(x_1, x_2)$  be defined as

$$f(x_1, x_2) = \frac{x_2(x_1^2 + x_2^2)}{x_2^2 + (x_1^2 + x_2^2)^2}.$$

This function is defined everywhere in  $R^2$  except at  $(0, 0)$ '. On the line  $x_1 = 0$ ,  $f(0, x_2) = x_2^3/(x_2^2 + x_2^4)$ , which goes to zero as  $x_2 \rightarrow 0$ . When  $x_2 = 0$ ,  $f(x_1, 0) = 0$  for  $x_1 \neq 0$ ; hence,  $f(x_1, 0) \rightarrow 0$  as  $x_1 \rightarrow 0$ . Furthermore, for any other straight line  $x_2 = tx_1$  ( $t \neq 0$ ) through the origin we have

$$\begin{aligned} f(x_1, tx_1) &= \frac{tx_1(x_1^2 + t^2x_1^2)}{t^2x_1^2 + (x_1^2 + t^2x_1^2)^2}, \\ &= \frac{tx_1(1 + t^2)}{t^2 + x_1^2(1 + t^2)^2}, \quad x_1 \neq 0, \end{aligned}$$

which has a limit equal to zero as  $x_1 \rightarrow 0$ . We conclude that the limit of  $f(x_1, x_2)$  is zero as  $\mathbf{x} \rightarrow \mathbf{0}$  along any straight line through the origin. However,  $f(x_1, x_2)$  does not have a limit as  $\mathbf{x} \rightarrow \mathbf{0}$ . For example, along the circle  $x_2 = x_1^2 + x_2^2$  that passes through the origin,

$$f(x_1, x_2) = \frac{(x_1^2 + x_2^2)^2}{(x_1^2 + x_2^2)^2 + (x_1^2 + x_2^2)^2} = \frac{1}{2}, \quad x_1^2 + x_2^2 \neq 0.$$

Hence,  $f(x_1, x_2) \rightarrow \frac{1}{2} \neq 0$ .

This example demonstrates that a function may not have a limit as  $\mathbf{x} \rightarrow \mathbf{a}$  even though its limit exists for approaches toward  $\mathbf{a}$  along straight lines.

### 7.3. CONTINUITY OF A MULTIVARIABLE FUNCTION

The notion of continuity for a function of several variables is much the same as that for a function of a single variable.

**Definition 7.3.1.** Let  $\mathbf{f}: D \rightarrow R^m$ , where  $D \subset R^n$ , and let  $\mathbf{a} \in D$ . Then  $\mathbf{f}(\mathbf{x})$  is continuous at  $\mathbf{a}$  if

$$\lim_{\mathbf{x} \rightarrow \mathbf{a}} \mathbf{f}(\mathbf{x}) = \mathbf{f}(\mathbf{a}),$$

where  $\mathbf{x}$  remains in  $D$  as it approaches  $\mathbf{a}$ . This is equivalent to stating that for a given  $\epsilon > 0$  there exists a  $\delta > 0$  such that

$$\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a})\| < \epsilon$$

for all  $\mathbf{x} \in D \cap N_\delta(\mathbf{a})$ .

If  $\mathbf{f}(\mathbf{x})$  is continuous at every point  $\mathbf{x}$  in  $D$ , then it is said to be continuous in  $D$ . In particular, if  $\mathbf{f}(\mathbf{x})$  is continuous in  $D$  and if  $\delta$  (in the definition of continuity) depends only on  $\epsilon$  (that is,  $\delta$  is the same for all points in  $D$  for the given  $\epsilon$ ), then  $\mathbf{f}(\mathbf{x})$  is said to be uniformly continuous in  $D$ .  $\square$

We now present several theorems that provide some important properties of multivariable continuous functions. These theorems are analogous to those given in Chapter 3. Let us first consider the following lemmas (the proofs are left to the reader):

**Lemma 7.3.1.** Every bounded sequence in  $R^n$  has a convergent subsequence.

This lemma is analogous to Theorem 5.1.4.

**Lemma 7.3.2.** Suppose that  $f, g: D \rightarrow R$  are real-valued continuous functions, where  $D \subset R^n$ . Then we have the following:

1.  $f + g$ ,  $f - g$ , and  $fg$  are continuous in  $D$ .
2.  $|f|$  is continuous in  $D$ .
3.  $1/f$  is continuous in  $D$  provided that  $f(\mathbf{x}) \neq 0$  for all  $\mathbf{x}$  in  $D$ .

This lemma is analogous to Theorem 3.4.1.

**Lemma 7.3.3.** Suppose that  $\mathbf{f}: D \rightarrow R^m$  is continuous, where  $D \subset R^n$ , and that  $\mathbf{g}: G \rightarrow R^p$  is also continuous, where  $G \subset R^m$  is the image of  $D$  under  $\mathbf{f}$ . Then the composite function  $\mathbf{g} \circ \mathbf{f}: D \rightarrow R^p$ , defined as  $\mathbf{g} \circ \mathbf{f}(\mathbf{x}) = \mathbf{g}[\mathbf{f}(\mathbf{x})]$ , is continuous in  $D$ .

This lemma is analogous to Theorem 3.4.2.

**Theorem 7.3.1.** Let  $f: D \rightarrow R$  be a real-valued continuous function defined on a closed and bounded set  $D \subset R^n$ . Then there exist points  $\mathbf{p}$  and  $\mathbf{q}$  in  $D$  for which

$$f(\mathbf{p}) = \sup_{\mathbf{x} \in D} f(\mathbf{x}), \quad (7.1)$$

$$f(\mathbf{q}) = \inf_{\mathbf{x} \in D} f(\mathbf{x}). \quad (7.2)$$

Thus  $f(\mathbf{x})$  attains each of its infimum and supremum at least once in  $D$ .

*Proof.* Let us first show that  $f(\mathbf{x})$  is bounded in  $D$ . We shall prove this by contradiction. Suppose that  $f(\mathbf{x})$  is not bounded in  $D$ . Then we can find a sequence of points  $\{\mathbf{p}_i\}_{i=1}^{\infty}$  in  $D$  such that  $|f(\mathbf{p}_i)| \geq i$  for  $i \geq 1$  and hence

$|f(\mathbf{p}_i)| \rightarrow \infty$  as  $i \rightarrow \infty$ . Since the terms of this sequence are elements in a bounded set,  $\{\mathbf{p}_i\}_{i=1}^\infty$  must be a bounded sequence. By Lemma 7.3.1, this sequence has a convergent subsequence  $\{\mathbf{p}_{k_i}\}_{i=1}^\infty$ . Let  $\mathbf{p}_0$  be the limit of this subsequence, which is also a limit point of  $D$ ; hence, it belongs to  $D$ , since  $D$  is closed. Now, on one hand,  $|f(\mathbf{p}_{k_i})| \rightarrow |f(\mathbf{p}_0)|$  as  $i \rightarrow \infty$ , by the continuity of  $f(\mathbf{x})$  and hence of  $|f(\mathbf{x})|$  [see Lemma 7.3.2(2)]. On the other hand,  $|f(\mathbf{p}_{k_i})| \rightarrow \infty$ . This contradiction shows that  $f(\mathbf{x})$  must be bounded in  $D$ . Consequently, the infimum and supremum of  $f(\mathbf{x})$  in  $D$  are finite.

Suppose now equality (7.1) does not hold for any  $\mathbf{p} \in D$ . Then,  $M - f(\mathbf{x}) > 0$  for all  $\mathbf{x} \in D$ , where  $M = \sup_{\mathbf{x} \in D} f(\mathbf{x})$ . Consequently,  $[M - f(\mathbf{x})]^{-1}$  is positive and continuous in  $D$  by Lemma 7.3.2(3) and is therefore bounded by the first half of this proof. However, if  $\delta > 0$  is any given positive number, then, by the definition of  $M$ , we can find a point  $\mathbf{x}_\delta$  in  $D$  for which  $f(\mathbf{x}_\delta) > M - \delta$ , or

$$\frac{1}{M - f(\mathbf{x}_\delta)} > \frac{1}{\delta}.$$

This implies that  $[M - f(\mathbf{x})]^{-1}$  is not bounded, a contradiction, which proves equality (7.1). The proof of equality (7.2) is similar.  $\square$

**Theorem 7.3.2.** Suppose that  $D$  is a closed and bounded set in  $R^n$ . If  $\mathbf{f}: D \rightarrow R^m$  is continuous, then it is uniformly continuous in  $D$ .

*Proof.* We shall prove this theorem by contradiction. Suppose that  $\mathbf{f}$  is not uniformly continuous in  $D$ . Then there exists an  $\epsilon > 0$  such that for every  $\delta > 0$  we can find  $\mathbf{a}$  and  $\mathbf{b}$  in  $D$  such that  $\|\mathbf{a} - \mathbf{b}\| < \delta$ , but  $\|\mathbf{f}(\mathbf{a}) - \mathbf{f}(\mathbf{b})\| \geq \epsilon$ . Let us choose  $\delta = 1/i$ ,  $i \geq 1$ . We can therefore find two sequences  $\{\mathbf{a}_i\}_{i=1}^\infty, \{\mathbf{b}_i\}_{i=1}^\infty$  with  $\mathbf{a}_i, \mathbf{b}_i \in D$  such that  $\|\mathbf{a}_i - \mathbf{b}_i\| < 1/i$ , and

$$\|\mathbf{f}(\mathbf{a}_i) - \mathbf{f}(\mathbf{b}_i)\| \geq \epsilon \quad (7.3)$$

for  $i \geq 1$ . Now, the sequence  $\{\mathbf{a}_i\}_{i=1}^\infty$  is bounded. Hence, by Lemma 7.3.1, it has a convergent subsequence  $\{\mathbf{a}_{k_i}\}_{i=1}^\infty$  whose limit, denoted by  $\mathbf{a}_0$ , is in  $D$ , since  $D$  is closed. Also, since  $\mathbf{f}$  is continuous at  $\mathbf{a}_0$ , we can find a  $\lambda > 0$  such that  $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{a}_0)\| < \epsilon/2$  if  $\|\mathbf{x} - \mathbf{a}_0\| < \lambda$ , where  $\mathbf{x} \in D$ . By the convergence of  $\{\mathbf{a}_{k_i}\}_{i=1}^\infty$  to  $\mathbf{a}_0$ , we can choose  $k_i$  large enough so that

$$\frac{1}{k_i} < \frac{\lambda}{2} \quad (7.4)$$

and

$$\|\mathbf{a}_{k_i} - \mathbf{a}_0\| < \frac{\lambda}{2}. \quad (7.5)$$

From (7.5) it follows that

$$\|\mathbf{f}(\mathbf{a}_{k_i}) - \mathbf{f}(\mathbf{a}_0)\| < \frac{\epsilon}{2}. \quad (7.6)$$

Furthermore, since  $\|\mathbf{a}_{k_i} - \mathbf{b}_{k_i}\| < 1/k_i$ , we can write

$$\begin{aligned} \|\mathbf{b}_{k_i} - \mathbf{a}_0\| &\leq \|\mathbf{a}_{k_i} - \mathbf{a}_0\| + \|\mathbf{a}_{k_i} - \mathbf{b}_{k_i}\| \\ &< \frac{\lambda}{2} + \frac{1}{k_i} < \lambda. \end{aligned}$$

Hence, by the continuity of  $\mathbf{f}$  at  $\mathbf{a}_0$ ,

$$\|\mathbf{f}(\mathbf{b}_{k_i}) - \mathbf{f}(\mathbf{a}_0)\| < \frac{\epsilon}{2}. \quad (7.7)$$

From inequalities (7.6) and (7.7) we conclude that whenever  $k_i$  satisfies inequalities (7.4) and (7.5),

$$\|\mathbf{f}(\mathbf{a}_{k_i}) - \mathbf{f}(\mathbf{b}_{k_i})\| \leq \|\mathbf{f}(\mathbf{a}_{k_i}) - \mathbf{f}(\mathbf{a}_0)\| + \|\mathbf{f}(\mathbf{b}_{k_i}) - \mathbf{f}(\mathbf{a}_0)\| < \epsilon,$$

which contradicts inequality (7.3). This leads us to assert that  $\mathbf{f}$  is uniformly continuous in  $D$ .  $\square$

## 7.4. DERIVATIVES OF A MULTIVARIABLE FUNCTION

In this section we generalize the concept of differentiation given in Chapter 4 to a multivariable function  $\mathbf{f}: D \rightarrow R^m$ , where  $D \subset R^n$ .

Let  $\mathbf{a} = (a_1, a_2, \dots, a_n)'$  be an interior point of  $D$ . Suppose that the limit

$$\lim_{h_i \rightarrow 0} \frac{\mathbf{f}(a_1, a_2, \dots, a_i + h_i, \dots, a_n) - \mathbf{f}(a_1, a_2, \dots, a_i, \dots, a_n)}{h_i}$$

exists; then  $\mathbf{f}$  is said to have a partial derivative with respect to  $x_i$  at  $\mathbf{a}$ . This derivative is denoted by  $\partial \mathbf{f}(\mathbf{a}) / \partial x_i$ , or just  $\mathbf{f}_{x_i}(\mathbf{a})$ ,  $i = 1, 2, \dots, n$ . Hence, partial differentiation with respect to  $x_i$  is done in the usual fashion while treating all the remaining variables as constants. For example, if  $f: R^3 \rightarrow R$  is defined

as  $f(x_1, x_2, x_3) = x_1 x_2^2 + x_2 x_3^3$ , then at any point  $\mathbf{x} \in R^3$  we have

$$\begin{aligned}\frac{\partial f(\mathbf{x})}{\partial x_1} &= x_2^2, \\ \frac{\partial f(\mathbf{x})}{\partial x_2} &= 2x_1 x_2 + x_3^3, \\ \frac{\partial f(\mathbf{x})}{\partial x_3} &= 3x_2 x_3^2.\end{aligned}$$

In general, if  $f_j$  is the  $j$ th element of  $\mathbf{f}$  ( $j = 1, 2, \dots, m$ ), then the terms  $\partial f_j(\mathbf{x}) / \partial x_i$ , for  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, m$ , constitute an  $m \times n$  matrix called the Jacobian matrix (after Carl Gustav Jacobi, 1804–1851) of  $\mathbf{f}$  at  $\mathbf{x}$  and is denoted by  $\mathbf{J}_f(\mathbf{x})$ . If  $m = n$ , the determinant of  $\mathbf{J}_f(\mathbf{x})$  is called the Jacobian determinant; it is sometimes represented as

$$\det[\mathbf{J}_f(\mathbf{x})] = \frac{\partial(f_1, f_2, \dots, f_n)}{\partial(x_1, x_2, \dots, x_n)}. \quad (7.8)$$

For example, if  $\mathbf{f}: R^3 \rightarrow R^2$  is such that

$$\mathbf{f}(x_1, x_2, x_3) = (x_1^2 \cos x_2, x_2^2 + x_3^2 e^{x_1})',$$

then

$$\mathbf{J}_f(x_1, x_2, x_3) = \begin{bmatrix} 2x_1 \cos x_2 & -x_1^2 \sin x_2 & 0 \\ x_3^2 e^{x_1} & 2x_2 & 2x_3 e^{x_1} \end{bmatrix}.$$

Higher-order partial derivatives of  $\mathbf{f}$  are defined similarly. For example, the second-order partial derivative of  $\mathbf{f}$  with respect to  $x_i$  at  $\mathbf{a}$  is defined as

$$\lim_{h_i \rightarrow 0} \frac{\mathbf{f}_{x_i}(a_1, a_2, \dots, a_i + h_i, \dots, a_n) - \mathbf{f}_{x_i}(a_1, a_2, \dots, a_i, \dots, a_n)}{h_i}$$

and is denoted by  $\partial^2 \mathbf{f}(\mathbf{a}) / \partial x_i^2$ , or  $\mathbf{f}_{x_i x_i}(\mathbf{a})$ . Also, the second-order partial derivative of  $\mathbf{f}$  with respect to  $x_i$  and  $x_j$ ,  $i \neq j$ , at  $\mathbf{a}$  is given by

$$\lim_{h_j \rightarrow 0} \frac{\mathbf{f}_{x_i}(a_1, a_2, \dots, a_j + h_j, \dots, a_n) - \mathbf{f}_{x_i}(a_1, a_2, \dots, a_j, \dots, a_n)}{h_j}$$

and is denoted by  $\partial^2 \mathbf{f}(\mathbf{a}) / \partial x_j \partial x_i$ , or  $\mathbf{f}_{x_j x_i}(\mathbf{a})$ ,  $i \neq j$ .



Under certain conditions, the order in which differentiation with respect to  $x_i$  and  $x_j$  takes place is irrelevant, that is,  $\mathbf{f}_{x_i x_j}(\mathbf{a})$  is identical to  $\mathbf{f}_{x_j x_i}(\mathbf{a})$ ,  $i \neq j$ . This property is known as the commutative property of partial differentiation and is proved in the next theorem.

**Theorem 7.4.1.** Let  $\mathbf{f}: D \rightarrow R^m$ , where  $D \subset R^n$ , and let  $\mathbf{a}$  be an interior point of  $D$ . Suppose that in a neighborhood of  $\mathbf{a}$  the following conditions are satisfied:

1.  $\partial \mathbf{f}(\mathbf{x})/\partial x_i$  and  $\partial \mathbf{f}(\mathbf{x})/\partial x_j$  exist and are finite ( $i, j = 1, 2, \dots, n$ ,  $i \neq j$ ).
2. Of the derivatives  $\partial^2 \mathbf{f}(\mathbf{x})/\partial x_i \partial x_j$ ,  $\partial^2 \mathbf{f}(\mathbf{x})/\partial x_j \partial x_i$  one exists and is continuous.

Then

$$\frac{\partial^2 \mathbf{f}(\mathbf{a})}{\partial x_i \partial x_j} = \frac{\partial^2 \mathbf{f}(\mathbf{a})}{\partial x_j \partial x_i}.$$

*Proof.* Let us suppose that  $\partial^2 \mathbf{f}(\mathbf{x})/\partial x_j \partial x_i$  exists and is continuous in a neighborhood of  $\mathbf{a}$ . Without loss of generality we assume that  $i < j$ . If  $\partial^2 \mathbf{f}(\mathbf{a})/\partial x_i \partial x_j$  exists, then it must be equal to the limit

$$\lim_{h_i \rightarrow 0} \frac{\mathbf{f}_{x_j}(a_1, a_2, \dots, a_i + h_i, \dots, a_n) - \mathbf{f}_{x_j}(a_1, a_2, \dots, a_i, \dots, a_n)}{h_i},$$

that is,

$$\begin{aligned} & \lim_{h_i \rightarrow 0} \frac{1}{h_i} \left[ \lim_{h_j \rightarrow 0} \frac{1}{h_j} \{ \mathbf{f}(a_1, a_2, \dots, a_i + h_i, \dots, a_j + h_j, \dots, a_n) \right. \\ & \quad \left. - \mathbf{f}(a_1, a_2, \dots, a_i + h_i, \dots, a_j, \dots, a_n) \} \right. \\ & \quad \left. - \lim_{h_j \rightarrow 0} \frac{1}{h_j} \{ \mathbf{f}(a_1, a_2, \dots, a_i, \dots, a_j + h_j, \dots, a_n) \right. \\ & \quad \left. - \mathbf{f}(a_1, a_2, \dots, a_i, \dots, a_j, \dots, a_n) \} \right]. \quad (7.9) \end{aligned}$$

Let us denote  $\mathbf{f}(x_1, x_2, \dots, x_j + h_j, \dots, x_n) - \mathbf{f}(x_1, x_2, \dots, x_j, \dots, x_n)$  by  $\Psi(x_1, x_2, \dots, x_n)$ . Then the double limit in (7.9) can be written as

$$\begin{aligned} & \lim_{h_i \rightarrow 0} \lim_{h_j \rightarrow 0} \frac{1}{h_i h_j} \left[ \Psi(a_1, a_2, \dots, a_i + h_i, \dots, a_j, \dots, a_n) \right. \\ & \quad \left. - \Psi(a_1, a_2, \dots, a_i, \dots, a_j, \dots, a_n) \right] \\ & = \lim_{h_i \rightarrow 0} \lim_{h_j \rightarrow 0} \frac{1}{h_j} \left[ \frac{\partial \Psi(a_1, a_2, \dots, a_i + \theta_i h_i, \dots, a_j, \dots, a_n)}{\partial x_i} \right], \quad (7.10) \end{aligned}$$

where  $0 < \theta_i < 1$ . In formula (7.10) we have applied the mean value theorem (Theorem 4.2.2) to  $\Psi$  as if it were a function of the single variable  $x_i$  (since  $\partial \mathbf{f}/\partial x_i$ , and hence  $\partial \Psi/\partial x_i$  exists in a neighborhood of  $\mathbf{a}$ ). The right-hand side of (7.10) can then be written as

$$\begin{aligned} & \lim_{h_i \rightarrow 0} \lim_{h_j \rightarrow 0} \frac{1}{h_j} \left[ \frac{\partial \mathbf{f}(a_1, a_2, \dots, a_i + \theta_i h_i, \dots, a_j + h_j, \dots, a_n)}{\partial x_i} \right. \\ & \quad \left. - \frac{\partial \mathbf{f}(a_1, a_2, \dots, a_i + \theta_i h_i, \dots, a_j, \dots, a_n)}{\partial x_i} \right] \\ &= \lim_{h_i \rightarrow 0} \lim_{h_j \rightarrow 0} \frac{\partial}{\partial x_j} \left[ \frac{\partial \mathbf{f}(a_1, a_2, \dots, a_i + \theta_i h_i, \dots, a_j + \theta_j h_j, \dots, a_n)}{\partial x_i} \right], \end{aligned} \quad (7.11)$$

where  $0 < \theta_j < 1$ . In formula (7.11) we have again made use of the mean value theorem, since  $\partial^2 \mathbf{f}(\mathbf{x})/\partial x_j \partial x_i$  exists in the given neighborhood around  $\mathbf{a}$ . Furthermore, since  $\partial^2 \mathbf{f}(\mathbf{x})/\partial x_j \partial x_i$  is continuous in this neighborhood, the double limit in (7.11) is equal to  $\partial^2 \mathbf{f}(\mathbf{a})/\partial x_j \partial x_i$ . This establishes the assertion that the two second-order partial derivatives of  $\mathbf{f}$  are equal.  $\square$

**EXAMPLE 7.4.1.** Consider the function  $f: R^3 \rightarrow R$ , where  $f(x_1, x_2, x_3) = x_1 e^{x_2} + x_2 \cos x_1$ . Then

$$\begin{aligned} \frac{\partial f(x_1, x_2, x_3)}{\partial x_1} &= e^{x_2} - x_2 \sin x_1, \\ \frac{\partial f(x_1, x_2, x_3)}{\partial x_2} &= x_1 e^{x_2} + \cos x_1, \\ \frac{\partial^2 f(x_1, x_2, x_3)}{\partial x_2 \partial x_1} &= e^{x_2} - \sin x_1, \\ \frac{\partial^2 f(x_1, x_2, x_3)}{\partial x_1 \partial x_2} &= e^{x_2} - \sin x_1. \end{aligned}$$

### 7.4.1. The Total Derivative

Let  $f(\mathbf{x})$  be a real-valued function defined on a set  $D \subset R^n$ , where  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ . Suppose that  $x_1, x_2, \dots, x_n$  are functions of a single variable  $t$ . Then  $f$  is a function of  $t$ . The ordinary derivative of  $f$  with respect to  $t$ , namely  $df/dt$ , is called the total derivative of  $f$ .

Let us now assume that for the values of  $t$  under consideration  $dx_i/dt$  exists for  $i = 1, 2, \dots, n$  and that  $\partial f(\mathbf{x})/\partial x_i$  exists and is continuous in the interior of  $D$  for  $i = 1, 2, \dots, n$ . Under these considerations, the total derivative of  $f$  is given by

$$\frac{df}{dt} = \sum_{i=1}^n \frac{\partial f(\mathbf{x})}{\partial x_i} \frac{dx_i}{dt}. \quad (7.12)$$

To show this we proceed as follows: Let  $\Delta x_1, \Delta x_2, \dots, \Delta x_n$  be increments of  $x_1, x_2, \dots, x_n$  that correspond to an increment  $\Delta t$  of  $t$ . In turn,  $f$  will have the increment  $\Delta f$ . We then have

$$\Delta f = f(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n) - f(x_1, x_2, \dots, x_n).$$

This can be written as

$$\begin{aligned} \Delta f = & [f(x_1 + \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n) \\ & - f(x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n)] \\ & + [f(x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n) \\ & - f(x_1, x_2, x_3 + \Delta x_3, \dots, x_n + \Delta x_n)] \\ & + [f(x_1, x_2, x_3 + \Delta x_3, \dots, x_n + \Delta x_n) \\ & - f(x_1, x_2, x_3, x_4 + \Delta x_4, \dots, x_n + \Delta x_n)] \\ & + \dots + [f(x_1, x_2, \dots, x_{n-1}, x_n + \Delta x_n) - f(x_1, x_2, \dots, x_n)]. \end{aligned}$$

By applying the mean value theorem to the difference in each bracket we obtain

$$\begin{aligned} \Delta f = & \Delta x_1 \frac{\partial f(x_1 + \theta_1 \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n)}{\partial x_1} \\ & + \Delta x_2 \frac{\partial f(x_1, x_2 + \theta_2 \Delta x_2, x_3 + \Delta x_3, \dots, x_n + \Delta x_n)}{\partial x_2} \\ & + \Delta x_3 \frac{\partial f(x_1, x_2, x_3 + \theta_3 \Delta x_3, x_4 + \Delta x_4, \dots, x_n + \Delta x_n)}{\partial x_3} \\ & + \dots + \Delta x_n \frac{\partial f(x_1, x_2, \dots, x_{n-1}, x_n + \theta_n \Delta x_n)}{\partial x_n}, \end{aligned}$$

where  $0 < \theta_i < 1$  for  $i = 1, 2, \dots, n$ . Hence,

$$\begin{aligned} \frac{\Delta f}{\Delta t} &= \frac{\Delta x_1}{\Delta t} \frac{\partial f(x_1 + \theta_1 \Delta x_1, x_2 + \Delta x_2, \dots, x_n + \Delta x_n)}{\partial x_1} \\ &+ \frac{\Delta x_2}{\Delta t} \frac{\partial f(x_1, x_2 + \theta_2 \Delta x_2, x_3 + \Delta x_3, \dots, x_n + \Delta x_n)}{\partial x_2} \\ &+ \frac{\Delta x_3}{\Delta t} \frac{\partial f(x_1, x_2, x_3 + \theta_3 \Delta x_3, x_4 + \Delta x_4, \dots, x_n + \Delta x_n)}{\partial x_3} \\ &+ \dots + \frac{\Delta x_n}{\Delta t} \frac{\partial f(x_1, x_2, \dots, x_{n-1}, x_n + \theta_n \Delta x_n)}{\partial x_n}. \end{aligned} \quad (7.13)$$

As  $\Delta t \rightarrow 0$ ,  $\Delta x_i/\Delta t \rightarrow dx_i/dt$ , and the partial derivatives in (7.13), being continuous, tend to  $\partial f(\mathbf{x})/\partial x_i$  for  $i = 1, 2, \dots, n$ . Thus  $\Delta f/\Delta t$  tends to the right-hand side of formula (7.12).

For example, consider the function  $f(x_1, x_2) = x_1^2 - x_2^3$ , where  $x_1 = e^t \cos t$ ,  $x_2 = \cos t + \sin t$ . Then,

$$\begin{aligned} \frac{df}{dt} &= 2x_1(e^t \cos t - e^t \sin t) - 3x_2^2(-\sin t + \cos t) \\ &= 2e^t \cos t(e^t \cos t - e^t \sin t) - 3(\cos t + \sin t)^2(-\sin t + \cos t) \\ &= (\cos t - \sin t)(2e^{2t} \cos t - 6 \sin t \cos t - 3). \end{aligned}$$

Of course, the same result could have been obtained by expressing  $f$  directly as a function of  $t$  via  $x_1$  and  $x_2$  and then differentiating it with respect to  $t$ .

We can generalize formula (7.12) by assuming that each of  $x_1, x_2, \dots, x_n$  is a function of several variables including the variable  $t$ . In this case, we need to consider the partial derivative  $\partial f/\partial t$ , which can be similarly shown to have the value

$$\frac{\partial f}{\partial t} = \sum_{i=1}^n \frac{\partial f(\mathbf{x})}{\partial x_i} \frac{\partial x_i}{\partial t}. \quad (7.14)$$

In general, the expression

$$df = \sum_{i=1}^n \frac{\partial f(\mathbf{x})}{\partial x_i} dx_i \quad (7.15)$$

is called the total differential of  $f$  at  $\mathbf{x}$ .

EXAMPLE 7.4.2. Consider the equation  $f(x_1, x_2) = 0$ , which in general represents a relation between  $x_1$  and  $x_2$ . It may or may not define  $x_2$  as a function of  $x_1$ . In this case,  $x_2$  is said to be an implicit function of  $x_1$ . If  $x_2$  can be obtained as a function of  $x_1$ , then we write  $x_2 = g(x_1)$ . Consequently,  $f[x_1, g(x_1)]$  will be identically equal to zero. Hence,  $f[x_1, g(x_1)]$ , being a function of one variable  $x_1$ , will have a total derivative identically equal to zero. By applying formula (7.12) with  $t = x_1$  we obtain

$$\frac{df}{dx_1} = \frac{\partial f}{\partial x_1} + \frac{\partial f}{\partial x_2} \frac{dx_2}{dx_1} \equiv 0.$$

If  $\partial f / \partial x_2 \neq 0$ , then the derivative of  $x_2$  is given by

$$\frac{dx_2}{dx_1} = \frac{-\partial f / \partial x_1}{\partial f / \partial x_2}. \quad (7.16)$$

In particular, if  $f(x_1, x_2) = 0$  is of the form  $x_1 - h(x_2) = 0$ , and if this equation can be solved uniquely for  $x_2$  in terms of  $x_1$ , then  $x_2$  represents the inverse function of  $h$ , that is,  $x_2 = h^{-1}(x_1)$ . Thus according to formula (7.16),

$$\frac{dh^{-1}}{dx_1} = \frac{1}{dh/dx_2}.$$

This agrees with the formula for the derivative of the inverse function given in Theorem 4.2.4.

### 7.4.2. Directional Derivatives

Let  $\mathbf{f}: D \rightarrow R^m$ , where  $D \subset R^n$ , and let  $\mathbf{v}$  be a unit vector in  $R^n$  (that is, a vector whose length is equal to one), which represents a certain direction in the  $n$ -dimensional Euclidean space. By definition, the directional derivative of  $\mathbf{f}$  at a point  $\mathbf{x}$  in the direction of  $\mathbf{v}$  is given by the limit

$$\lim_{h \rightarrow 0} \frac{\mathbf{f}(\mathbf{x} + h\mathbf{v}) - \mathbf{f}(\mathbf{x})}{h},$$

if it exists. In particular, if  $\mathbf{v} = \mathbf{e}_i$ , the unit vector in the direction of the  $i$ th coordinate axis, then the directional derivative of  $\mathbf{f}$  in the direction of  $\mathbf{v}$  is just the partial derivative of  $\mathbf{f}$  with respect to  $x_i$  ( $i = 1, 2, \dots, n$ ).

**Lemma 7.4.1.** Let  $\mathbf{f}: D \rightarrow R^m$ , where  $D \subset R^n$ . If the partial derivatives  $\partial f_j / \partial x_i$  exist at a point  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$  in the interior of  $D$  for  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, m$ , where  $f_j$  is the  $j$ th element of  $\mathbf{f}$ , then the directional derivative of  $\mathbf{f}$  at  $\mathbf{x}$  in the direction of a unit vector  $\mathbf{v}$  exists and is equal to  $\mathbf{J}_f(\mathbf{x})\mathbf{v}$ , where  $\mathbf{J}_f(\mathbf{x})$  is the  $m \times n$  Jacobian of  $\mathbf{f}$  at  $\mathbf{x}$ .

*Proof.* Let us first consider the directional derivative of  $f_j$  in the direction of  $\mathbf{v}$ . To do so, we rotate the coordinate axes so that  $\mathbf{v}$  coincides with the direction of the  $\xi_1$ -axis, where  $\xi_1, \xi_2, \dots, \xi_n$  are the resulting new coordinates. By the well-known relations for rotation of axes in analytic geometry of  $n$  dimensions we have

$$x_i = \xi_1 v_i + \sum_{l=2}^n \xi_l \lambda_{li}, \quad i = 1, 2, \dots, n, \quad (7.17)$$

where  $v_i$  is the  $i$ th element of  $\mathbf{v}$  ( $i = 1, 2, \dots, n$ ) and  $\lambda_{li}$  is the  $i$ th element of  $\boldsymbol{\lambda}_l$ , the unit vector in the direction of the  $\xi_l$ -axis ( $l = 2, 3, \dots, n$ ).

Now, the directional derivative of  $f_j$  in the direction of  $\mathbf{v}$  can be obtained by first expressing  $f_j$  as a function of  $\xi_1, \xi_2, \dots, \xi_n$  using the relations (7.17) and then differentiating it with respect to  $\xi_1$ . By formula (7.14), this is equal to

$$\begin{aligned} \frac{\partial f_j}{\partial \xi_1} &= \sum_{i=1}^n \frac{\partial f_j}{\partial x_i} \frac{\partial x_i}{\partial \xi_1} \\ &= \sum_{i=1}^n \frac{\partial f_j}{\partial x_i} v_i, \quad j = 1, 2, \dots, m. \end{aligned} \quad (7.18)$$

From formula (7.18) we conclude that the directional derivative of  $\mathbf{f} = (f_1, f_2, \dots, f_m)'$  in the direction of  $\mathbf{v}$  is equal to  $\mathbf{J}_f(\mathbf{x})\mathbf{v}$ .  $\square$

EXAMPLE 7.4.3. Let  $\mathbf{f}: \mathbb{R}^3 \rightarrow \mathbb{R}^2$  be defined as

$$\mathbf{f}(x_1, x_2, x_3) = \begin{bmatrix} x_1^2 + x_2^2 + x_3^2 \\ x_1^2 - x_1 x_2 + x_3^2 \end{bmatrix}.$$

The directional derivative of  $\mathbf{f}$  at  $\mathbf{x} = (1, 2, 1)'$  in the direction of  $\mathbf{v} = (1/\sqrt{2}, -1/\sqrt{2}, 0)'$  is

$$\begin{aligned} \mathbf{J}_f(\mathbf{x})\mathbf{v} &= \begin{bmatrix} 2x_1 & 2x_2 & 2x_3 \\ 2x_1 - x_2 & -x_1 & 2x_3 \end{bmatrix}_{(1,2,1)} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 4 & 2 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{-2}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}. \end{aligned}$$

**Definition 7.4.1.** Let  $f: D \rightarrow R$ , where  $D \subset R^n$ . If the partial derivatives  $\partial f / \partial x_i$  ( $i = 1, 2, \dots, n$ ) exist at a point  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$  in the interior of  $D$ , then the vector  $(\partial f / \partial x_1, \partial f / \partial x_2, \dots, \partial f / \partial x_n)'$  is called the gradient of  $f$  at  $\mathbf{x}$  and is denoted by  $\nabla f(\mathbf{x})$ .  $\square$

Using Definition 7.4.1, the directional derivative of  $f$  at  $\mathbf{x}$  in the direction of a unit vector  $\mathbf{v}$  can be expressed as  $\nabla f(\mathbf{x})' \mathbf{v}$ , where  $\nabla f(\mathbf{x})'$  denotes the transpose of  $\nabla f(\mathbf{x})$ .

### *The Geometric Meaning of the Gradient*

Let  $f: D \rightarrow R$ , where  $D \subset R^n$ . Suppose that the partial derivatives of  $f$  exist at a point  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$  in the interior of  $D$ . Let  $C$  denote a smooth curve that lies on the surface of  $f(\mathbf{x}) = c_0$ , where  $c_0$  is a constant, and passes through the point  $\mathbf{x}$ . This curve can be represented by the equations  $x_1 = g_1(t), x_2 = g_2(t), \dots, x_n = g_n(t)$ , where  $a \leq t \leq b$ . By formula (7.12), the total derivative of  $f$  with respect to  $t$  at  $\mathbf{x}$  is

$$\frac{df}{dt} = \sum_{i=1}^n \frac{\partial f(\mathbf{x})}{\partial x_i} \frac{dg_i}{dt}. \quad (7.19)$$

The vector  $\boldsymbol{\lambda} = (dg_1/dt, dg_2/dt, \dots, dg_n/dt)'$  is tangent to  $C$  at  $\mathbf{x}$ . Thus from (7.19) we obtain

$$\frac{df}{dt} = \nabla f(\mathbf{x})' \boldsymbol{\lambda}. \quad (7.20)$$

Now, since  $f[g_1(t), g_2(t), \dots, g_n(t)] = c_0$  along  $C$ , then  $df/dt = 0$  and hence  $\nabla f(\mathbf{x})' \boldsymbol{\lambda} = 0$ . This indicates that the gradient vector is orthogonal to  $\boldsymbol{\lambda}$ , and hence to  $C$ , at  $\mathbf{x} \in D$ . Since this result is true for any smooth curve through  $\mathbf{x}$ , we conclude that the gradient vector  $\nabla f(\mathbf{x})$  is orthogonal to the surface of  $f(\mathbf{x}) = c_0$  at  $\mathbf{x}$ .

**Definition 7.4.2.** Let  $f: D \rightarrow R$ , where  $D \subset R^n$ . Then  $\nabla f: D \rightarrow R^n$ . The Jacobian matrix of  $\nabla f(\mathbf{x})$  is called the Hessian matrix of  $f$  and is denoted by  $\mathbf{H}_f(\mathbf{x})$ . Thus  $\mathbf{H}_f(\mathbf{x}) = \mathbf{J}_{\nabla f}(\mathbf{x})$ , that is,

$$\mathbf{H}_f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix}. \quad (7.21)$$

The determinant of  $\mathbf{H}_f(\mathbf{x})$  is called the *Hessian determinant*. If the conditions of Theorem 7.4.1 regarding the commutative property of partial differentiation are valid, then  $\mathbf{H}_f(\mathbf{x})$  is a symmetric matrix. As we shall see in Section 7.7, the Hessian matrix plays an important role in the identification of maxima and minima of a multivariable function.  $\square$

### 7.4.3. Differentiation of Composite Functions

Let  $\mathbf{f}: D_1 \rightarrow R^m$ , where  $D_1 \subset R^n$ , and let  $\mathbf{g}: D_2 \rightarrow R^p$ , where  $D_2 \subset R^m$ . Let  $\mathbf{x}_0$  be an interior point of  $D_1$  and  $\mathbf{f}(\mathbf{x}_0)$  be an interior point of  $D_2$ . If the  $m \times n$  Jacobian matrix  $\mathbf{J}_f(\mathbf{x}_0)$  and the  $p \times m$  Jacobian matrix  $\mathbf{J}_g[\mathbf{f}(\mathbf{x}_0)]$  both exist, then the  $p \times n$  Jacobian matrix  $\mathbf{J}_h(\mathbf{x}_0)$  for the composite function  $\mathbf{h} = \mathbf{g} \circ \mathbf{f}$  exists and is given by

$$\mathbf{J}_h(\mathbf{x}_0) = \mathbf{J}_g[\mathbf{f}(\mathbf{x}_0)]\mathbf{J}_f(\mathbf{x}_0). \quad (7.22)$$

To prove formula (7.22), let us consider the  $(k, i)$ th element of  $\mathbf{J}_h(\mathbf{x}_0)$ , namely  $\partial h_k(\mathbf{x}_0)/\partial x_i$ , where  $h_k(\mathbf{x}_0) = g_k[\mathbf{f}(\mathbf{x}_0)]$  is the  $k$ th element of  $\mathbf{h}(\mathbf{x}_0) = \mathbf{g}[\mathbf{f}(\mathbf{x}_0)]$ ,  $i = 1, 2, \dots, n$ ;  $k = 1, 2, \dots, p$ . By applying formula (7.14) we obtain

$$\frac{\partial h_k(\mathbf{x}_0)}{\partial x_i} = \sum_{j=1}^m \frac{\partial g_k[\mathbf{f}(\mathbf{x}_0)]}{\partial f_j} \frac{\partial f_j(\mathbf{x}_0)}{\partial x_i}, \quad i = 1, 2, \dots, n; k = 1, 2, \dots, p, \quad (7.23)$$

where  $f_j(\mathbf{x}_0)$  is the  $j$ th element of  $\mathbf{f}(\mathbf{x}_0)$ ,  $j = 1, 2, \dots, m$ . But  $\partial g_k[\mathbf{f}(\mathbf{x}_0)]/\partial f_j$  is the  $(k, j)$ th element of  $\mathbf{J}_g[\mathbf{f}(\mathbf{x}_0)]$ , and  $\partial f_j(\mathbf{x}_0)/\partial x_i$  is the  $(j, i)$ th element of  $\mathbf{J}_f(\mathbf{x}_0)$ ,  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, m$ ;  $k = 1, 2, \dots, p$ . Hence, formula (7.22) follows from formula (7.23) and the rule of matrix multiplication.

In particular, if  $m = n = p$ , then from formula (7.22), the determinant of  $\mathbf{J}_h(\mathbf{x}_0)$  is given by

$$\det[\mathbf{J}_h(\mathbf{x}_0)] = \det[\mathbf{J}_g(\mathbf{f}(\mathbf{x}_0))]\det[\mathbf{J}_f(\mathbf{x}_0)]. \quad (7.24)$$

Using the notation in formula (7.8), formula (7.24) can be expressed as

$$\frac{\partial(h_1, h_2, \dots, h_n)}{\partial(x_1, x_2, \dots, x_n)} = \frac{\partial(g_1, g_2, \dots, g_n)}{\partial(f_1, f_2, \dots, f_n)} \frac{\partial(f_1, f_2, \dots, f_n)}{\partial(x_1, x_2, \dots, x_n)}. \quad (7.25)$$

EXAMPLE 7.4.4. Let  $\mathbf{f}: R^2 \rightarrow R^3$  be given by

$$\mathbf{f}(x_1, x_2) = \begin{bmatrix} x_1^2 - x_2 \cos x_1 \\ x_1 x_2 \\ x_1^3 + x_2^3 \end{bmatrix}.$$



Let  $g: R^3 \rightarrow R$  be defined as

$$g(\xi_1, \xi_2, \xi_3) = \xi_1 - \xi_2^2 + \xi_3,$$

where

$$\xi_1 = x_1^2 - x_2 \cos x_1,$$

$$\xi_2 = x_1 x_2,$$

$$\xi_3 = x_1^3 + x_2^3.$$

In this case,

$$\mathbf{J}_f(\mathbf{x}) = \begin{bmatrix} 2x_1 + x_2 \sin x_1 & -\cos x_1 \\ x_2 & x_1 \\ 3x_1^2 & 3x_2^2 \end{bmatrix},$$

$$\mathbf{J}_g[\mathbf{f}(\mathbf{x})] = (1, -2\xi_2, 1).$$

Hence, by formula (7.22),

$$\begin{aligned} \mathbf{J}_h(\mathbf{x}) &= (1, -2\xi_2, 1) \begin{bmatrix} 2x_1 + x_2 \sin x_1 & -\cos x_1 \\ x_2 & x_1 \\ 3x_1^2 & 3x_2^2 \end{bmatrix} \\ &= (2x_1 + x_2 \sin x_1 - 2x_1 x_2^2 + 3x_1^2, -\cos x_1 - 2x_1^2 x_2 + 3x_2^2). \end{aligned}$$

## 7.5. TAYLOR'S THEOREM FOR A MULTIVARIABLE FUNCTION

We shall now consider a multidimensional analogue of Taylor's theorem, which was discussed in Section 4.3 for a single-variable function.

Let us first introduce the following notation: Let  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ . Then  $\mathbf{x}'\nabla$  denotes a first-order differential operator of the form

$$\mathbf{x}'\nabla = \sum_{i=1}^n x_i \frac{\partial}{\partial x_i}.$$

The symbol  $\nabla$ , called the del operator, was used earlier to define the gradient vector. If  $m$  is a positive integer, then  $(\mathbf{x}'\nabla)^m$  denotes an  $m$ th-order differential operator. For example, for  $m = n = 2$ ,

$$\begin{aligned} (\mathbf{x}'\nabla)^2 &= \left( x_1 \frac{\partial}{\partial x_1} + x_2 \frac{\partial}{\partial x_2} \right)^2 \\ &= x_1^2 \frac{\partial^2}{\partial x_1^2} + 2x_1 x_2 \frac{\partial^2}{\partial x_1 \partial x_2} + x_2^2 \frac{\partial^2}{\partial x_2^2}. \end{aligned}$$

Thus  $(\mathbf{x}'\nabla)^2$  is obtained by squaring  $x_1 \partial/\partial x_1 + x_2 \partial/\partial x_2$  in the usual fashion, except that the squares of  $\partial/\partial x_1$  and  $\partial/\partial x_2$  are replaced by  $\partial^2/\partial x_1^2$  and  $\partial^2/\partial x_2^2$ , respectively, and the product of  $\partial/\partial x_1$  and  $\partial/\partial x_2$  is replaced by  $\partial^2/\partial x_1 \partial x_2$  (here we are assuming that the commutative property of partial differentiation holds once these differential operators are applied to a real-valued function). In general,  $(\mathbf{x}'\nabla)^m$  is obtained by a multinomial expansion of degree  $m$  of the form

$$(\mathbf{x}'\nabla)^m = \sum_{k_1, k_2, \dots, k_n} \binom{m}{k_1, k_2, \dots, k_n} x_1^{k_1} x_2^{k_2} \cdots x_n^{k_n} \frac{\partial^m}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_n^{k_n}},$$

where the sum is taken over all  $n$ -tuples  $(k_1, k_2, \dots, k_n)$  for which  $\sum_{i=1}^n k_i = m$ , and

$$\binom{m}{k_1, k_2, \dots, k_n} = \frac{m!}{k_1! k_2! \cdots k_n!}.$$

If a real-valued function  $f(\mathbf{x})$  has partial derivatives through order  $m$ , then an application of the differential operator  $(\mathbf{x}'\nabla)^m$  to  $f(\mathbf{x})$  results in

$$\begin{aligned} (\mathbf{x}'\nabla)^m f(\mathbf{x}) &= \sum_{k_1, k_2, \dots, k_n} \binom{m}{k_1, k_2, \dots, k_n} x_1^{k_1} x_2^{k_2} \cdots \\ &\quad \times x_n^{k_n} \frac{\partial^m f(\mathbf{x})}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots \partial x_n^{k_n}}. \end{aligned} \quad (7.26)$$

The notation  $(\mathbf{x}'\nabla)^m f(\mathbf{x}_0)$  indicates that  $(\mathbf{x}'\nabla)^m f(\mathbf{x})$  is evaluated at  $\mathbf{x}_0$ .

**Theorem 7.5.1.** Let  $f: D \rightarrow R$ , where  $D \subset R^n$ , and let  $N_\delta(\mathbf{x}_0)$  be a neighborhood of  $\mathbf{x}_0 \in D$  such that  $N_\delta(\mathbf{x}_0) \subset D$ . If  $f$  and all its partial derivatives of order  $\leq r$  exist and are continuous in  $N_\delta(\mathbf{x}_0)$ , then for any  $\mathbf{x} \in N_\delta(\mathbf{x}_0)$ ,

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{r-1} \frac{[(\mathbf{x} - \mathbf{x}_0)'\nabla]^i f(\mathbf{x}_0)}{i!} + \frac{[(\mathbf{x} - \mathbf{x}_0)'\nabla]^r f(\mathbf{z}_0)}{r!}, \quad (7.27)$$

where  $\mathbf{z}_0$  is a point on the line segment from  $\mathbf{x}_0$  to  $\mathbf{x}$ .

*Proof.* Let  $\mathbf{h} = \mathbf{x} - \mathbf{x}_0$ . Let the function  $\phi(t)$  be defined as  $\phi(t) = f(\mathbf{x}_0 + t\mathbf{h})$ , where  $0 \leq t \leq 1$ . If  $t = 0$ , then  $\phi(0) = f(\mathbf{x}_0)$  and  $\phi(1) = f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x})$ , if

$t = 1$ . Now, by formula (7.12),

$$\begin{aligned}\frac{d\phi(t)}{dt} &= \sum_{i=1}^n h_i \frac{\partial f(\mathbf{x})}{\partial x_i} \bigg|_{\mathbf{x}=\mathbf{x}_0+t\mathbf{h}} \\ &= (\mathbf{h}'\nabla)f(\mathbf{x}_0+t\mathbf{h}),\end{aligned}$$

where  $h_i$  is the  $i$ th element of  $\mathbf{h}$  ( $i = 1, 2, \dots, n$ ). Furthermore, the derivative of order  $m$  of  $\phi(t)$  is

$$\frac{d^m\phi(t)}{dt^m} = (\mathbf{h}'\nabla)^m f(\mathbf{x}_0+t\mathbf{h}), \quad 1 \leq m \leq r.$$

Since the partial derivatives of  $f$  through order  $r$  are continuous, then the same order derivatives of  $\phi(t)$  are also continuous on  $[0, 1]$  and

$$\frac{d^m\phi(t)}{dt^m} \bigg|_{t=0} = (\mathbf{h}'\nabla)^m f(\mathbf{x}_0), \quad 1 \leq m \leq r.$$

If we now apply Taylor's theorem in Section 4.3 to the single-variable function  $\phi(t)$ , we obtain

$$\phi(t) = \phi(0) + \sum_{i=1}^{r-1} \frac{t^i}{i!} \frac{d^i\phi(t)}{dt^i} \bigg|_{t=0} + \frac{t^r}{r!} \frac{d^r\phi(t)}{dt^r} \bigg|_{t=\xi}, \quad (7.28)$$

where  $0 < \xi < t$ . By setting  $t = 1$  in formula (7.28), we obtain

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{r-1} \frac{[(\mathbf{x} - \mathbf{x}_0)'\nabla]^i f(\mathbf{x}_0)}{i!} + \frac{[(\mathbf{x} - \mathbf{x}_0)'\nabla]^r f(\mathbf{z}_0)}{r!},$$

where  $\mathbf{z}_0 = \mathbf{x}_0 + \xi\mathbf{h}$ . Since  $0 < \xi < 1$ , the point  $\mathbf{z}_0$  lies on the line segment between  $\mathbf{x}_0$  and  $\mathbf{x}$ .  $\square$

In particular, if  $f(\mathbf{x})$  has partial derivatives of all orders in  $N_\delta(\mathbf{x}_0)$ , then we have the series expansion

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{\infty} \frac{[(\mathbf{x} - \mathbf{x}_0)'\nabla]^i f(\mathbf{x}_0)}{i!}. \quad (7.29)$$

In this case, the last term in formula (7.27) serves as a remainder of Taylor's series.

EXAMPLE 7.5.1. Consider the function  $f: R^2 \rightarrow R$  defined as  $f(x_1, x_2) = x_1 x_2 + x_1^2 + e^{x_1} \cos x_2$ . This function has partial derivatives of all orders. Thus in a neighborhood of  $\mathbf{x}_0 = (0, 0)'$  we can write

$$f(x_1, x_2) = 1 + (\mathbf{x}' \nabla) f(0, 0) + \frac{1}{2!} (\mathbf{x}' \nabla)^2 f(0, 0) + \frac{1}{3!} (\mathbf{x}' \nabla)^3 f(\xi x_1, \xi x_2),$$

$$0 < \xi < 1.$$

It can be verified that

$$\begin{aligned} (\mathbf{x}' \nabla) f(0, 0) &= x_1, \\ (\mathbf{x}' \nabla)^2 f(0, 0) &= 3x_1^2 + 2x_1 x_2 - x_2^2, \\ (\mathbf{x}' \nabla)^3 f(\xi x_1, \xi x_2) &= x_1^3 e^{\xi x_1} \cos(\xi x_2) - 3x_1^2 x_2 e^{\xi x_1} \sin(\xi x_2) \\ &\quad - 3x_1 x_2^2 e^{\xi x_1} \cos(\xi x_2) + x_2^3 e^{\xi x_1} \sin(\xi x_2). \end{aligned}$$

Hence,

$$\begin{aligned} f(x_1, x_2) &= 1 + x_1 + \frac{1}{2!} (3x_1^2 + 2x_1 x_2 - x_2^2) \\ &\quad + \frac{1}{3!} \{ [x_1^3 - 3x_1 x_2^2] e^{\xi x_1} \cos(\xi x_2) + [x_2^3 - 3x_1^2 x_2] e^{\xi x_1} \sin(\xi x_2) \}. \end{aligned}$$

The first three terms serve as a second-order approximation of  $f(x_1, x_2)$ , while the last term serves as a remainder.

## 7.6. INVERSE AND IMPLICIT FUNCTION THEOREMS

Consider the function  $\mathbf{f}: D \rightarrow R^n$ , where  $D \subset R^n$ . Let  $\mathbf{y} = \mathbf{f}(\mathbf{x})$ . The purpose of this section is to present conditions for the existence of an inverse function  $\mathbf{f}^{-1}$  which expresses  $\mathbf{x}$  as a function of  $\mathbf{y}$ . These conditions are given in the next theorem, whose proof can be found in Sagan (1974, page 371). See also Fulks (1978, page 346).

**Theorem 7.6.1** (Inverse Function Theorem). Let  $\mathbf{f}: D \rightarrow R^n$ , where  $D$  is an open subset of  $R^n$  and  $\mathbf{f}$  has continuous first-order partial derivatives in  $D$ . If for some  $\mathbf{x}_0 \in D$ , the  $n \times n$  Jacobian matrix  $\mathbf{J}_{\mathbf{f}}(\mathbf{x}_0)$  is nonsingular,

that is,

$$\det[\mathbf{J}_f(\mathbf{x}_0)] = \frac{\partial(f_1, f_2, \dots, f_n)}{\partial(x_1, x_2, \dots, x_n)} \bigg|_{\mathbf{x}=\mathbf{x}_0} \neq 0,$$

where  $f_i$  is the  $i$ th element of  $\mathbf{f}$  ( $i = 1, 2, \dots, n$ ), then there exist an  $\epsilon > 0$  and a  $\delta > 0$  such that an inverse function  $\mathbf{f}^{-1}$  exists in the neighborhood  $N_\delta[\mathbf{f}(\mathbf{x}_0)]$  and takes values in the neighborhood  $N_\epsilon(\mathbf{x}_0)$ . Moreover,  $\mathbf{f}^{-1}$  has continuous first-order partial derivatives in  $N_\delta[\mathbf{f}(\mathbf{x}_0)]$ , and its Jacobian matrix at  $\mathbf{f}(\mathbf{x}_0)$  is the inverse of  $\mathbf{J}_f(\mathbf{x}_0)$ ; hence,

$$\det\{\mathbf{J}_{f^{-1}}[\mathbf{f}(\mathbf{x}_0)]\} = \frac{1}{\det[\mathbf{J}_f(\mathbf{x}_0)]}. \quad (7.30)$$

EXAMPLE 7.6.1. Let  $\mathbf{f}: R^3 \rightarrow R^3$  be given by

$$\mathbf{f}(x_1, x_2, x_3) = \begin{bmatrix} 2x_1x_2 - x_2 \\ x_1^2 + x_2 + 2x_3^2 \\ x_1x_2 + x_2 \end{bmatrix}.$$

Here,

$$\mathbf{J}_f(\mathbf{x}) = \begin{bmatrix} 2x_2 & 2x_1 - 1 & 0 \\ 2x_1 & 1 & 4x_3 \\ x_2 & x_1 + 1 & 0 \end{bmatrix},$$

and  $\det[\mathbf{J}_f(\mathbf{x})] = -12x_2x_3$ . Hence, all  $\mathbf{x} \in R^3$  at which  $x_2x_3 \neq 0$ ,  $\mathbf{f}$  has an inverse function  $\mathbf{f}^{-1}$ . For example, if  $D = \{(x_1, x_2, x_3) | x_2 > 0, x_3 > 0\}$ , then  $\mathbf{f}$  is invertible in  $D$ . From the equations

$$y_1 = 2x_1x_2 - x_2,$$

$$y_2 = x_1^2 + x_2 + 2x_3^2,$$

$$y_3 = x_1x_2 + x_2,$$

we obtain the inverse function  $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{y})$ , where

$$x_1 = \frac{y_1 + y_3}{2y_3 - y_1},$$

$$x_2 = \frac{-y_1 + 2y_3}{3},$$

$$x_3 = \frac{1}{\sqrt{2}} \left[ y_2 - \frac{(y_1 + y_3)^2}{(2y_3 - y_1)^2} - \frac{2y_3 - y_1}{3} \right]^{1/2}.$$

If, for example, we consider  $\mathbf{x}_0 = (1, 1, 1)'$ , then  $\mathbf{y}_0 = \mathbf{f}(\mathbf{x}_0) = (1, 4, 2)'$ , and  $\det[\mathbf{J}_f(\mathbf{x}_0)] = -12$ . The Jacobian matrix of  $\mathbf{f}^{-1}$  at  $\mathbf{y}_0$  is

$$\mathbf{J}_{f^{-1}}(\mathbf{y}_0) = \begin{bmatrix} \frac{2}{3} & 0 & -\frac{1}{3} \\ -\frac{1}{3} & 0 & \frac{2}{3} \\ -\frac{1}{4} & \frac{1}{4} & 0 \end{bmatrix}.$$

Its determinant is equal to

$$\det[\mathbf{J}_{f^{-1}}(\mathbf{y}_0)] = -\frac{1}{12}.$$

We note that this is the reciprocal of  $\det[\mathbf{J}_f(\mathbf{x}_0)]$ , as it should be according to formula (7.30).

The inverse function theorem can be viewed as providing a unique solution to a system of  $n$  equations given by  $\mathbf{y} = \mathbf{f}(\mathbf{x})$ . There are, however, situations in which  $\mathbf{y}$  is not explicitly expressed as a function of  $\mathbf{x}$ . In general, we may have two vectors,  $\mathbf{x}$  and  $\mathbf{y}$ , of orders  $n \times 1$  and  $m \times 1$ , respectively, that satisfy the relation

$$\mathbf{g}(\mathbf{x}, \mathbf{y}) = \mathbf{0}, \quad (7.31)$$

where  $\mathbf{g}: R^{m+n} \rightarrow R^n$ . In this more general case, we have  $n$  equations involving  $m+n$  variables, namely, the elements of  $\mathbf{x}$  and those of  $\mathbf{y}$ . The question now is what conditions will allow us to solve equations (7.31) uniquely for  $\mathbf{x}$  in terms of  $\mathbf{y}$ . The answer to this question is given in the next theorem, whose proof can be found in Fulks (1978, page 352).

**Theorem 7.6.2** (Implicit Function Theorem). Let  $\mathbf{g}: D \rightarrow R^n$ , where  $D$  is an open subset of  $R^{m+n}$ , and  $\mathbf{g}$  has continuous first-order partial derivatives in  $D$ . If there is a point  $\mathbf{z}_0 \in D$ , where  $\mathbf{z}_0 = (\mathbf{x}'_0, \mathbf{y}'_0)'$  with  $\mathbf{x}_0 \in R^n$ ,  $\mathbf{y}_0 \in R^m$  such that  $\mathbf{g}(\mathbf{z}_0) = \mathbf{0}$ , and if at  $\mathbf{z}_0$ ,

$$\frac{\partial(g_1, g_2, \dots, g_n)}{\partial(x_1, x_2, \dots, x_n)} \neq 0,$$

where  $g_i$  is the  $i$ th element of  $\mathbf{g}$  ( $i = 1, 2, \dots, n$ ), then there is a neighborhood  $N_\delta(\mathbf{y}_0)$  of  $\mathbf{y}_0$  in which the equation  $\mathbf{g}(\mathbf{x}, \mathbf{y}) = \mathbf{0}$  can be solved uniquely for  $\mathbf{x}$  as a continuously differentiable function of  $\mathbf{y}$ .

EXAMPLE 7.6.2. Let  $\mathbf{g}: R^3 \rightarrow R^2$  be given by

$$\mathbf{g}(x_1, x_2, y) = \begin{bmatrix} x_1 + x_2 + y^2 - 18 \\ x_1 - x_1x_2 + y - 4 \end{bmatrix}.$$

We have

$$\frac{\partial(g_1, g_2)}{\partial(x_1, x_2)} = \det \left( \begin{bmatrix} 1 & 1 \\ 1 - x_2 & -x_1 \end{bmatrix} \right) = x_2 - x_1 - 1.$$

Let  $\mathbf{z} = (x_1, x_2, y)'$ . At the point  $\mathbf{z}_0 = (1, 1, 4)'$ , for example,  $\mathbf{g}(\mathbf{z}_0) = \mathbf{0}$  and  $\partial(g_1, g_2)/\partial(x_1, x_2) = -1 \neq 0$ . Hence, by Theorem 7.6.2, we can solve the equations

$$x_1 + x_2 + y^2 - 18 = 0, \quad (7.32)$$

$$x_1 - x_1x_2 + y - 4 = 0 \quad (7.33)$$

uniquely in terms of  $y$  in some neighborhood of  $y_0 = 4$ . For example, if  $D$  in Theorem 7.6.2 is of the form

$$D = \{(x_1, x_2, y) | x_1 > 0, x_2 > 0, y < 4.06\},$$

then from equations (7.32) and (7.33) we obtain the solution

$$x_1 = \frac{1}{2} \left\{ -(y^2 - 17) + \left[ (y^2 - 17)^2 - 4y + 16 \right]^{1/2} \right\},$$

$$x_2 = \frac{1}{2} \left\{ 19 - y^2 - \left[ (y^2 - 17)^2 - 4y + 16 \right]^{1/2} \right\}.$$

We note that the sign preceding the square root in the formula for  $x_1$  was chosen as  $+$ , so that  $x_1 = x_2 = 1$  when  $y = 4$ . It can be verified that  $(y^2 - 17)^2 - 4y + 16$  is positive for  $y < 4.06$ .

## 7.7. OPTIMA OF A MULTIVARIABLE FUNCTION

Let  $f(\mathbf{x})$  be a real-valued function defined on a set  $D \subset R^n$ . A point  $\mathbf{x}_0 \in D$  is said to be a point of local maximum of  $f$  if there exists a neighborhood  $N_\delta(\mathbf{x}_0) \subset D$  such that  $f(\mathbf{x}) \leq f(\mathbf{x}_0)$  for all  $\mathbf{x} \in N_\delta(\mathbf{x}_0)$ . If  $f(\mathbf{x}) \geq f(\mathbf{x}_0)$  for all  $\mathbf{x} \in N_\delta(\mathbf{x}_0)$ , then  $\mathbf{x}_0$  is a point of local minimum. If one of these inequalities holds for all  $\mathbf{x}$  in  $D$ , then  $\mathbf{x}_0$  is called a point of absolute maximum, or a point of absolute minimum, respectively, of  $f$  in  $D$ . In either case,  $\mathbf{x}_0$  is referred to as a point of optimum (or extremum), and the value of  $f(\mathbf{x})$  at  $\mathbf{x} = \mathbf{x}_0$  is called an optimum value of  $f(\mathbf{x})$ .

In this section we shall discuss conditions under which  $f(\mathbf{x})$  attains local optima in  $D$ . Then, we shall investigate the determination of the optima of  $f(\mathbf{x})$  over a constrained region of  $D$ .

As in the case of a single-variable function, if  $f(\mathbf{x})$  has first-order partial derivatives at a point  $\mathbf{x}_0$  in the interior of  $D$ , and if  $\mathbf{x}_0$  is a point of local optimum, then  $\partial f / \partial x_i = 0$  for  $i = 1, 2, \dots, n$  at  $\mathbf{x}_0$ . The proof of this fact is similar to that of Theorem 4.4.1. Thus the vanishing of the first-order partial derivatives of  $f(\mathbf{x})$  at  $\mathbf{x}_0$  is a necessary condition for a local optimum at  $\mathbf{x}_0$ , but is obviously not sufficient. The first-order partial derivatives can be zero without necessarily having a local optimum at  $\mathbf{x}_0$ .

In general, any point at which  $\partial f / \partial x_i = 0$  for  $i = 1, 2, \dots, n$  is called a stationary point. It follows that any point of local optimum at which  $f$  has first-order partial derivatives is a stationary point, but not every stationary point is a point of local optimum. If no local optimum is attained at a stationary point  $\mathbf{x}_0$ , then  $\mathbf{x}_0$  is called a saddle point. The following theorem gives the conditions needed to have a local optimum at a stationary point.

**Theorem 7.7.1.** Let  $f: D \rightarrow R$ , where  $D \subset R^n$ . Suppose that  $f$  has continuous second-order partial derivatives in  $D$ . If  $\mathbf{x}_0$  is a stationary point of  $f$ , then at  $\mathbf{x}_0$   $f$  has the following:

- i. A local minimum if  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0) > 0$  for all  $\mathbf{h} = (h_1, h_2, \dots, h_n)'$  in a neighborhood of  $\mathbf{0}$ , where the elements of  $\mathbf{h}$  are not all equal to zero.
- ii. A local maximum if  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0) < 0$ , where  $\mathbf{h}$  is the same as in (i).
- iii. A saddle point if  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)$  changes sign for values of  $\mathbf{h}$  in a neighborhood of  $\mathbf{0}$ .

*Proof.* By applying Taylor's theorem to  $f(\mathbf{x})$  in a neighborhood of  $\mathbf{x}_0$  we obtain

$$f(\mathbf{x}_0 + \mathbf{h}) = f(\mathbf{x}_0) + (\mathbf{h}'\nabla)f(\mathbf{x}_0) + \frac{1}{2!}(\mathbf{h}'\nabla)^2 f(\mathbf{z}_0),$$

where  $\mathbf{h}$  is a nonzero vector in a neighborhood of  $\mathbf{0}$  and  $\mathbf{z}_0$  is a point on the line segment from  $\mathbf{x}_0$  to  $\mathbf{x}_0 + \mathbf{h}$ . Since  $\mathbf{x}_0$  is a stationary point, then  $(\mathbf{h}'\nabla)f(\mathbf{x}_0) = 0$ . Hence,

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = \frac{1}{2!}(\mathbf{h}'\nabla)^2 f(\mathbf{z}_0).$$

Also, since the second-order partial derivatives of  $f$  are continuous at  $\mathbf{x}_0$ , then we can write

$$f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0) = \frac{1}{2!}(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0) + o(\|\mathbf{h}\|),$$



where  $\|\mathbf{h}\| = (\mathbf{h}'\mathbf{h})^{1/2}$  and  $o(\|\mathbf{h}\|) \rightarrow 0$  as  $\mathbf{h} \rightarrow \mathbf{0}$ . We note that for small values of  $\|\mathbf{h}\|$ , the sign of  $f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)$  depends on the value of  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)$ . It follows that if

- i.  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0) > 0$ , then  $f(\mathbf{x}_0 + \mathbf{h}) > f(\mathbf{x}_0)$  for all nonzero values of  $\mathbf{h}$  in some neighborhood of  $\mathbf{0}$ . Thus  $\mathbf{x}_0$  is a point of local minimum of  $f$ .
- ii.  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0) < 0$ , then  $f(\mathbf{x}_0 + \mathbf{h}) < f(\mathbf{x}_0)$  for all nonzero values of  $\mathbf{h}$  in some neighborhood of  $\mathbf{0}$ . In this case,  $\mathbf{x}_0$  is a point of local maximum of  $f$ .
- iii.  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)$  changes sign inside a neighborhood of  $\mathbf{0}$ , then  $\mathbf{x}_0$  is neither a point of local maximum nor a point of local minimum. Therefore,  $\mathbf{x}_0$  must be a saddle point.  $\square$

We note that  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)$  can be written as a quadratic form of the form  $\mathbf{h}'\mathbf{A}\mathbf{h}$ , where  $\mathbf{A} = \mathbf{H}_f(\mathbf{x}_0)$  is the  $n \times n$  Hessian matrix of  $f$  evaluated at  $\mathbf{x}_0$ , that is,

$$\mathbf{A} = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \vdots & \vdots & & \vdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix}, \quad (7.34)$$

where for simplicity we have denoted  $\partial^2 f(\mathbf{x}_0) / \partial x_i \partial x_j$  by  $f_{ij}$ ,  $i, j = 1, 2, \dots, n$  [see formula (7.21)].

**Corollary 7.7.1.** Let  $f$  be the same function as in Theorem 7.7.1, and let  $\mathbf{A}$  be the matrix given by formula (7.34). If  $\mathbf{x}_0$  is a stationary point of  $f$ , then at  $\mathbf{x}_0$   $f$  has the following:

- i. A local minimum if  $\mathbf{A}$  is positive definite, that is, the leading principal minors of  $\mathbf{A}$  (see Definition 2.3.6) are all positive,

$$f_{11} > 0, \quad \det \left( \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \right) > 0, \dots, \quad \det(\mathbf{A}) > 0. \quad (7.35)$$

- ii. A local maximum if  $\mathbf{A}$  is negative definite, that is, the leading principal minors of  $\mathbf{A}$  have alternating signs as follows:

$$f_{11} < 0, \quad \det \left( \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix} \right) > 0, \dots, (-1)^n \det(\mathbf{A}) > 0. \quad (7.36)$$

- iii. A saddle point if  $\mathbf{A}$  is neither positive definite nor negative definite.

*Proof.*

- i. By Theorem 7.7.1,  $f$  has a local minimum at  $\mathbf{x}_0$  if  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0) = \mathbf{h}'\mathbf{A}\mathbf{h}$  is positive for all  $\mathbf{h} \neq \mathbf{0}$ , that is, if  $\mathbf{A}$  is positive definite. By Theorem 2.3.12(2),  $\mathbf{A}$  is positive definite if and only if its leading principal minors are all positive. The conditions stated in (7.35) are therefore sufficient for a local minimum at  $\mathbf{x}_0$ .
- ii.  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0) < 0$  if and only if  $\mathbf{A}$  is negative definite, or  $-\mathbf{A}$  is positive definite. Now, a leading principal minor of order  $m$  ( $= 1, 2, \dots, n$ ) of  $-\mathbf{A}$  is equal to  $(-1)^m$  multiplied by the corresponding leading principal minor of  $\mathbf{A}$ . This leads to conditions (7.36).
- iii. If  $\mathbf{A}$  is neither positive definite nor negative definite, then  $(\mathbf{h}'\nabla)^2 f(\mathbf{x}_0)$  must change sign inside a neighborhood of  $\mathbf{x}_0$ . This makes  $\mathbf{x}_0$  a saddle point.  $\square$

#### *A Special Case*

If  $f$  is a function of only  $n = 2$  variables,  $x_1$  and  $x_2$ , then conditions (7.35) and (7.36) can be written as:

- i.  $f_{11} > 0, f_{11}f_{22} - f_{12}^2 > 0$  for a local minimum at  $\mathbf{x}_0$ .
- ii.  $f_{11} < 0, f_{11}f_{22} - f_{12}^2 > 0$  for a local maximum at  $\mathbf{x}_0$ .

If  $f_{11}f_{22} - f_{12}^2 < 0$ , then  $\mathbf{x}_0$  is a saddle point, since in this case

$$\begin{aligned}\mathbf{h}'\mathbf{A}\mathbf{h} &= h_1^2 \frac{\partial^2 f(\mathbf{x}_0)}{\partial x_1^2} + 2h_1h_2 \frac{\partial^2 f(\mathbf{x}_0)}{\partial x_1 \partial x_2} + h_2^2 \frac{\partial^2 f(\mathbf{x}_0)}{\partial x_2^2} \\ &= \frac{\partial^2 f(\mathbf{x}_0)}{\partial x_1^2} (h_1 - ah_2)(h_1 - bh_2),\end{aligned}$$

where  $ah_2$  and  $bh_2$  are the real roots of the equation  $\mathbf{h}'\mathbf{A}\mathbf{h} = 0$  with respect to  $h_1$ . Hence,  $\mathbf{h}'\mathbf{A}\mathbf{h}$  changes sign in a neighborhood of  $\mathbf{0}$ .

If  $f_{11}f_{22} - f_{12}^2 = 0$ , then  $\mathbf{h}'\mathbf{A}\mathbf{h}$  can be written as

$$\mathbf{h}'\mathbf{A}\mathbf{h} = \frac{\partial^2 f(\mathbf{x}_0)}{\partial x_1^2} \left[ h_1 + h_2 \frac{\partial^2 f(\mathbf{x}_0)/\partial x_1 \partial x_2}{\partial^2 f(\mathbf{x}_0)/\partial x_1^2} \right]^2$$

provided that  $\partial^2 f(\mathbf{x}_0)/\partial x_1^2 \neq 0$ . Thus  $\mathbf{h}'\mathbf{A}\mathbf{h}$  has the same sign as that of  $\partial^2 f(\mathbf{x}_0)/\partial x_1^2$  except for those values of  $\mathbf{h} = (h_1, h_2)'$  for which

$$h_1 + h_2 \frac{\partial^2 f(\mathbf{x}_0)/\partial x_1 \partial x_2}{\partial^2 f(\mathbf{x}_0)/\partial x_1^2} = 0,$$

in which case it is zero. In the event  $\partial^2 f(\mathbf{x}_0)/\partial x_1^2 = 0$ , then  $\partial^2 f(\mathbf{x}_0)/\partial x_1 \partial x_2 = 0$ , and  $\mathbf{h}'\mathbf{A}\mathbf{h} = h_2^2 \partial^2 f(\mathbf{x}_0)/\partial x_2^2$ , which has the same sign as that of  $\partial^2 f(\mathbf{x}_0)/\partial x_2^2$ , if it is different from zero, except for those values of  $\mathbf{h} = (h_1, h_2)'$  for which  $h_2 = 0$ , where it is zero. It follows that when  $f_{11}f_{22} - f_{12}^2 = 0$ ,  $\mathbf{h}'\mathbf{A}\mathbf{h}$  has a constant sign for all  $h$  inside a neighborhood of  $\mathbf{0}$ . However, it can vanish for some nonzero values of  $\mathbf{h}$ . For such values of  $\mathbf{h}$ , the sign of  $f(\mathbf{x}_0 + \mathbf{h}) - f(\mathbf{x}_0)$  depends on the signs of higher-order partial derivatives (higher than second order) of  $f$  at  $\mathbf{x}_0$ . These can be obtained from Taylor's expansion. In this case, no decision can be made regarding the nature of the stationary point until these higher-order partial derivatives (if they exist) have been investigated.

EXAMPLE 7.7.1. Let  $f: R^2 \rightarrow R$  be the function  $f(x_1, x_2) = x_1^2 + 2x_2^2 - x_1$ . Consider the equations

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= 2x_1 - 1 = 0, \\ \frac{\partial f}{\partial x_2} &= 4x_2 = 0.\end{aligned}$$

The only solution is  $\mathbf{x}_0 = (0.5, 0)'$ . The Hessian matrix is

$$\mathbf{A} = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix},$$

which is positive definite, since  $2 > 0$  and  $\det(\mathbf{A}) = 8 > 0$ . The point  $\mathbf{x}_0$  is therefore a local minimum. Since it is the only one in  $R^2$ , it must also be the absolute minimum.

EXAMPLE 7.7.2. Consider the function  $f: R^3 \rightarrow R$ , where

$$\begin{aligned}f(x_1, x_2, x_3) &= \frac{1}{3}x_1^3 + 2x_2^2 + x_3^2 - 2x_1x_2 + 3x_1x_3 + x_2x_3 \\ &\quad - 10x_1 + 4x_2 - 6x_3 + 1.\end{aligned}$$

A stationary point must satisfy the equations

$$\frac{\partial f}{\partial x_1} = x_1^2 - 2x_2 + 3x_3 - 10 = 0, \quad (7.37)$$

$$\frac{\partial f}{\partial x_2} = -2x_1 + 4x_2 + x_3 + 4 = 0, \quad (7.38)$$

$$\frac{\partial f}{\partial x_3} = 3x_1 + x_2 + 2x_3 - 6 = 0. \quad (7.39)$$

From (7.38) and (7.39) we get

$$\begin{aligned}x_2 &= x_1 - 2, \\x_3 &= 4 - 2x_1.\end{aligned}$$

By substituting these expressions in equation (7.37) we obtain

$$x_1^2 - 8x_1 + 6 = 0.$$

This equation has two solutions,  $4 - \sqrt{10}$  and  $4 + \sqrt{10}$ . We therefore have two stationary points,

$$\begin{aligned}\mathbf{x}_0^{(1)} &= (4 + \sqrt{10}, 2 + \sqrt{10}, -4 - 2\sqrt{10})', \\ \mathbf{x}_0^{(2)} &= (4 - \sqrt{10}, 2 - \sqrt{10}, -4 + 2\sqrt{10})' .\end{aligned}$$

Now, the Hessian matrix is

$$\mathbf{A} = \begin{bmatrix} 2x_1 & -2 & 3 \\ -2 & 4 & 1 \\ 3 & 1 & 2 \end{bmatrix}.$$

Its leading principal minors are  $2x_1$ ,  $8x_1 - 4$ , and  $14x_1 - 56$ . The last one is the determinant of  $\mathbf{A}$ . At  $\mathbf{x}_0^{(1)}$  all three are positive. Therefore,  $\mathbf{x}_0^{(1)}$  is a point of local minimum. At  $\mathbf{x}_0^{(2)}$  the values of the leading principal minors are 1.675, 2.7018, and  $-44.272$ . In this case,  $\mathbf{A}$  is neither positive definite over negative definite. Thus  $\mathbf{x}_0^{(2)}$  is a saddle point.

## 7.8. THE METHOD OF LAGRANGE MULTIPLIERS

This method, which is due to Joseph Louis de Lagrange (1736–1813), is used to optimize a real-valued function  $f(x_1, x_2, \dots, x_n)$ , where  $x_1, x_2, \dots, x_n$  are subject to  $m$  ( $< n$ ) equality constraints of the form

$$\begin{aligned}g_1(x_1, x_2, \dots, x_n) &= 0, \\ g_2(x_1, x_2, \dots, x_n) &= 0, \\ &\vdots \\ g_m(x_1, x_2, \dots, x_n) &= 0,\end{aligned}\tag{7.40}$$

where  $g_1, g_2, \dots, g_m$  are differentiable functions.

The determination of the stationary points in this constrained optimization problem is done by first considering the function

$$F(\mathbf{x}) = f(\mathbf{x}) + \sum_{j=1}^m \lambda_j g_j(\mathbf{x}),\tag{7.41}$$

where  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$  and  $\lambda_1, \lambda_2, \dots, \lambda_m$  are scalars called *Lagrange multipliers*. By differentiating (7.41) with respect to  $x_1, x_2, \dots, x_n$  and equating the partial derivatives to zero we obtain

$$\frac{\partial F}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \quad i = 1, 2, \dots, n. \quad (7.42)$$

Equations (7.40) and (7.42) consist of  $m + n$  equations in  $m + n$  unknowns, namely,  $x_1, x_2, \dots, x_n$ ;  $\lambda_1, \lambda_2, \dots, \lambda_m$ . The solutions for  $x_1, x_2, \dots, x_n$  determine the locations of the stationary points. The following argument explains why this is the case:

Suppose that in equation (7.40) we can solve for  $m$   $x_i$ 's, for example,  $x_1, x_2, \dots, x_m$ , in terms of the remaining  $n - m$  variables. By Theorem 7.6.2, this is possible whenever

$$\frac{\partial(g_1, g_2, \dots, g_m)}{\partial(x_1, x_2, \dots, x_m)} \neq 0. \quad (7.43)$$

In this case, we can write

$$\begin{aligned} x_1 &= h_1(x_{m+1}, x_{m+2}, \dots, x_n), \\ x_2 &= h_2(x_{m+1}, x_{m+2}, \dots, x_n), \\ &\vdots \\ x_m &= h_m(x_{m+1}, x_{m+2}, \dots, x_n). \end{aligned} \quad (7.44)$$

Thus  $f(\mathbf{x})$  is a function of only  $n - m$  variables, namely,  $x_{m+1}, x_{m+2}, \dots, x_n$ . If the partial derivatives of  $f$  with respect to these variables exist and if  $f$  has a local optimum, then these partial derivatives must necessarily vanish, that is,

$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^m \frac{\partial f}{\partial h_j} \frac{\partial h_j}{\partial x_i} = 0, \quad i = m + 1, m + 2, \dots, n. \quad (7.45)$$

Now, if equations (7.44) are used to substitute  $h_1, h_2, \dots, h_m$  for  $x_1, x_2, \dots, x_m$ , respectively, in equation (7.40), then we obtain the identities

$$\begin{aligned} g_1(h_1, h_2, \dots, h_m, x_{m+1}, x_{m+2}, \dots, x_n) &\equiv 0, \\ g_2(h_1, h_2, \dots, h_m, x_{m+1}, x_{m+2}, \dots, x_n) &\equiv 0, \\ &\vdots \\ g_m(h_1, h_2, \dots, h_m, x_{m+1}, x_{m+2}, \dots, x_n) &\equiv 0. \end{aligned}$$

By differentiating these identities with respect to  $x_{m+1}, x_{m+2}, \dots, x_n$  we obtain

$$\frac{\partial g_k}{\partial x_i} + \sum_{j=1}^m \frac{\partial g_k}{\partial h_j} \frac{\partial h_j}{\partial x_i} = 0, \quad i = m+1, m+2, \dots, n; k = 1, 2, \dots, m. \quad (7.46)$$

Let us now define the vectors

$$\delta_k = \left( \frac{\partial g_k}{\partial x_{m+1}}, \frac{\partial g_k}{\partial x_{m+2}}, \dots, \frac{\partial g_k}{\partial x_n} \right)', \quad k = 1, 2, \dots, m,$$

$$\gamma_k = \left( \frac{\partial g_k}{\partial h_1}, \frac{\partial g_k}{\partial h_2}, \dots, \frac{\partial g_k}{\partial h_m} \right)', \quad k = 1, 2, \dots, m,$$

$$\eta_j = \left( \frac{\partial h_j}{\partial x_{m+1}}, \frac{\partial h_j}{\partial x_{m+2}}, \dots, \frac{\partial h_j}{\partial x_n} \right)', \quad j = 1, 2, \dots, m,$$

$$\psi = \left( \frac{\partial f}{\partial x_{m+1}}, \frac{\partial f}{\partial x_{m+2}}, \dots, \frac{\partial f}{\partial x_n} \right)',$$

$$\tau = \left( \frac{\partial f}{\partial h_1}, \frac{\partial f}{\partial h_2}, \dots, \frac{\partial f}{\partial h_m} \right)'.$$

Equations (7.45) and (7.46) can then be written as

$$[\delta_1: \delta_2: \dots: \delta_m] + [\eta_1: \eta_2: \dots: \eta_m] \Gamma = 0, \quad (7.47)$$

$$\psi + [\eta_1: \eta_2: \dots: \eta_m] \tau = 0, \quad (7.48)$$

where  $\Gamma = [\gamma_1: \gamma_2: \dots: \gamma_m]$ , which is a nonsingular  $m \times m$  matrix if condition (7.43) is valid. From equation (7.47) we have

$$[\eta_1: \eta_2: \dots: \eta_m] = -[\delta_1: \delta_2: \dots: \delta_m] \Gamma^{-1}.$$

By making the proper substitution in equation (7.48) we obtain

$$\psi + [\delta_1: \delta_2: \dots: \delta_m] \lambda = 0, \quad (7.49)$$

where

$$\lambda = -\Gamma^{-1} \tau. \quad (7.50)$$

Equations (7.49) can then be expressed as

$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \quad i = m+1, m+2, \dots, n. \quad (7.51)$$

From equation (7.50) we also have

$$\frac{\partial f}{\partial x_i} + \sum_{j=1}^m \lambda_j \frac{\partial g_j}{\partial x_i} = 0, \quad i = 1, 2, \dots, m. \quad (7.52)$$

Equations (7.51) and (7.52) can now be combined into a single vector equation of the form

$$\nabla f(\mathbf{x}) + \sum_{j=1}^m \lambda_j \nabla g_j = \mathbf{0},$$

which is the same as equation (7.42). We conclude that at a stationary point of  $f$ , the values of  $x_1, x_2, \dots, x_n$  and the corresponding values of  $\lambda_1, \lambda_2, \dots, \lambda_m$  must satisfy equations (7.40) and (7.42).

*Sufficient Conditions for a Local Optimum  
in the Method of Lagrange Multipliers*

Equations (7.42) are only necessary for a stationary point  $\mathbf{x}_0$  to be a point of local optimum of  $f$  subject to the constraints given by equations (7.40). Sufficient conditions for a local optimum are given in Gillespie (1954, pages 97–98). The following is a reproduction of these conditions:

Let  $\mathbf{x}_0$  be a stationary point of  $f$  whose coordinates satisfy equations (7.40) and (7.42), and let  $\lambda_1, \lambda_2, \dots, \lambda_m$  be the corresponding Lagrange multipliers. Let  $F_{ij}$  denote the second-order partial derivative of  $F$  in formula (7.41) with respect to  $x_i$ , and  $x_j$ ,  $i, j = 1, 2, \dots, n$ ;  $i \neq j$ . Consider the  $(m+n) \times (m+n)$  matrix

$$\mathbf{B}_1 = \begin{bmatrix} F_{11} & F_{12} & \cdots & F_{1n} & g_1^{(1)} & g_2^{(1)} & \cdots & g_m^{(1)} \\ F_{21} & F_{22} & \cdots & F_{2n} & g_1^{(2)} & g_2^{(2)} & \cdots & g_m^{(2)} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ F_{n1} & F_{n2} & \cdots & F_{nn} & g_1^{(n)} & g_2^{(n)} & \cdots & g_m^{(n)} \\ g_1^{(1)} & g_1^{(2)} & \cdots & g_1^{(n)} & 0 & 0 & \cdots & 0 \\ g_2^{(1)} & g_2^{(2)} & \cdots & g_2^{(n)} & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ g_m^{(1)} & g_m^{(2)} & \cdots & g_m^{(n)} & 0 & 0 & \cdots & 0 \end{bmatrix}, \quad (7.53)$$

where  $g_j^{(i)} = \partial g_j / \partial x_i$ ,  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, m$ . Let  $\Delta_1$  denote the determinant of  $\mathbf{B}_1$ . Furthermore, let  $\Delta_2, \Delta_3, \dots, \Delta_{n-m}$  denote a set of principal minors of  $\mathbf{B}_1$  (see Definition 2.3.6), namely, the determinants of the principal submatrices  $\mathbf{B}_2, \mathbf{B}_3, \dots, \mathbf{B}_{n-m}$ , where  $\mathbf{B}_i$  is obtained by deleting the first  $i - 1$  rows and the first  $i - 1$  columns of  $\mathbf{B}_1$  ( $i = 2, 3, \dots, n - m$ ). All the partial derivatives used in  $\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_{n-m}$  are evaluated at  $\mathbf{x}_0$ . Then sufficient conditions for  $\mathbf{x}_0$  to be a point of local minimum of  $f$  are the following:

i. If  $m$  is even,

$$\Delta_1 > 0, \quad \Delta_2 > 0, \dots, \quad \Delta_{n-m} > 0.$$

ii. If  $m$  is odd,

$$\Delta_1 < 0, \quad \Delta_2 < 0, \dots, \quad \Delta_{n-m} < 0.$$

However, sufficient conditions for  $\mathbf{x}_0$  to be a point of local maximum are the following:

i. If  $n$  is even,

$$\Delta_1 > 0, \quad \Delta_2 < 0, \dots, \quad (-1)^{n-m} \Delta_{n-m} < 0.$$

ii. If  $n$  is odd,

$$\Delta_1 < 0, \quad \Delta_2 > 0, \dots, \quad (-1)^{n-m} \Delta_{n-m} > 0.$$

EXAMPLE 7.8.1. Let us find the minimum and maximum distances from the origin to the curve determined by the intersection of the plane  $x_2 + x_3 = 0$  with the ellipsoid  $x_1^2 + 2x_2^2 + x_3^2 + 2x_2x_3 = 1$ . Let  $f(x_1, x_2, x_3)$  be the squared distance function from the origin, that is,

$$f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2.$$

The equality constraints are

$$g_1(x_1, x_2, x_3) \equiv x_2 + x_3 = 0,$$

$$g_2(x_1, x_2, x_3) \equiv x_1^2 + 2x_2^2 + x_3^2 + 2x_2x_3 - 1 = 0.$$

Then

$$\begin{aligned} F(x_1, x_2, x_3) &= x_1^2 + x_2^2 + x_3^2 + \lambda_1(x_2 + x_3) \\ &\quad + \lambda_2(x_1^2 + 2x_2^2 + x_3^2 + 2x_2x_3 - 1), \\ \frac{\partial F}{\partial x_1} &= 2x_1 + 2\lambda_2x_1 = 0, \end{aligned} \tag{7.54}$$



$$\frac{\partial F}{\partial x_2} = 2x_2 + \lambda_1 + 2\lambda_2(2x_2 + x_3) = 0, \quad (7.55)$$

$$\frac{\partial F}{\partial x_3} = 2x_3 + \lambda_1 + 2\lambda_2(x_2 + x_3) = 0. \quad (7.56)$$

Equations (7.54), (7.55), and (7.56) and the equality constraints are satisfied by the following sets of solutions:

- I.  $x_1 = 0, x_2 = 1, x_3 = -1, \lambda_1 = 2, \lambda_2 = -2.$
- II.  $x_1 = 0, x_2 = -1, x_3 = 1, \lambda_1 = -2, \lambda_2 = -2.$
- III.  $x_1 = 1, x_2 = 0, x_3 = 0, \lambda_1 = 0, \lambda_2 = -1.$
- IV.  $x_1 = -1, x_2 = 0, x_3 = 0, \lambda_1 = 0, \lambda_2 = -1.$

To determine if any of these four sets correspond to local maxima or minima, we need to examine the values of  $\Delta_1, \Delta_2, \dots, \Delta_{n-m}$ . Here, the matrix  $\mathbf{B}_1$  in formula (7.53) has the value

$$\mathbf{B}_1 = \begin{bmatrix} 2 + 2\lambda_2 & 0 & 0 & 0 & 2x_1 \\ 0 & 2 + 4\lambda_2 & 2\lambda_2 & 1 & 4x_2 + 2x_3 \\ 0 & 2\lambda_2 & 2 + 2\lambda_2 & 1 & 2x_2 + 2x_3 \\ 0 & 1 & 1 & 0 & 0 \\ 2x_1 & 4x_2 + 2x_3 & 2x_2 + 2x_3 & 0 & 0 \end{bmatrix}.$$

Since  $n = 3$  and  $m = 2$ , only one  $\Delta_i$ , namely,  $\Delta_1$ , the determinant of  $\mathbf{B}_1$ , is needed. Furthermore, since  $m$  is even and  $n$  is odd, a sufficient condition for a local minimum is  $\Delta_1 > 0$ , and for a local maximum the condition is  $\Delta_1 < 0$ . It can be verified that  $\Delta_1 = -8$  for solution sets I and II, and  $\Delta_1 = 8$  for solution sets III and IV. We therefore have local maxima at the points  $(0, 1, -1)$  and  $(0, -1, 1)$  with a common maximum value  $f_{\max} = 2$ . We also have local minima at the points  $(1, 0, 0)$  and  $(-1, 0, 0)$  with a common minimum value  $f_{\min} = 1$ . Since these are the only local optima on the curve of intersection, we conclude that the minimum distance from the origin to this curve is 1 and the maximum distance is  $\sqrt{2}$ .

## 7.9. THE RIEMANN INTEGRAL OF A MULTIVARIABLE FUNCTION

In Chapter 6 we discussed the Riemann integral of a real-valued function of a single variable  $x$ . In this section we extend the concept of Riemann integration to real-valued functions of  $n$  variables,  $x_1, x_2, \dots, x_n$ .

**Definition 7.9.1.** The set of points in  $R^n$  whose coordinates satisfy the inequalities

$$a_i \leq x_i \leq b_i, \quad i = 1, 2, \dots, n, \quad (7.57)$$

where  $a_i < b_i$ ,  $i = 1, 2, \dots, n$ , form an  $n$ -dimensional cell denoted by  $c_n(a, b)$ . The content (or volume) of this cell is  $\prod_{i=1}^n (b_i - a_i)$  and is denoted by  $\mu[c_n(a, b)]$ .

Suppose that  $P_i$  is a partition of the interval  $[a_i, b_i]$ ,  $i = 1, 2, \dots, n$ . The Cartesian product  $P = \times_{i=1}^n P_i$  is a partition of  $c_n(a, b)$  and consists of  $n$ -dimensional subcells of  $c_n(a, b)$ . We denote these subcells by  $S_1, S_2, \dots, S_\nu$ . The content of  $S_i$  is denoted by  $\mu(S_i)$ ,  $i = 1, 2, \dots, \nu$ , where  $\nu$  is the number of subcells.  $\square$

We shall first define the Riemann integral of a real-valued function  $f(\mathbf{x})$  on an  $n$ -dimensional cell; then we shall extend this definition to any bounded region in  $R^n$ .

### 7.9.1. The Riemann Integral on Cells

Let  $f: D \rightarrow R$ , where  $D \subset R^n$ . Suppose that  $c_n(a, b)$  is an  $n$ -dimensional cell contained in  $D$  and that  $f$  is bounded on  $c_n(a, b)$ . Let  $P$  be a partition of  $c_n(a, b)$  consisting of the subcells  $S_1, S_2, \dots, S_\nu$ . Let  $m_i$  and  $M_i$  be, respectively, the infimum and supremum of  $f$  on  $S_i$ ,  $i = 1, 2, \dots, \nu$ . Consider the sums

$$LS_P(f) = \sum_{i=1}^{\nu} m_i \mu(S_i), \quad (7.58)$$

$$US_P(f) = \sum_{i=1}^{\nu} M_i \mu(S_i). \quad (7.59)$$

We note the similarity of these sums to the ones defined in Section 6.2. As before, we refer to  $LS_P(f)$  and  $US_P(f)$  as the lower and upper sums, respectively, of  $f$  with respect to the partition  $P$ .

The following theorem is an  $n$ -dimensional analogue of Theorem 6.2.1. The proof is left to the reader.

**Theorem 7.9.1.** Let  $f: D \rightarrow R$ , where  $D \subset R^n$ . Suppose that  $f$  is bounded on  $c_n(a, b) \subset D$ . Then  $f$  is Riemann integrable on  $c_n(a, b)$  if and only if for every  $\epsilon > 0$  there exists a partition  $P$  of  $c_n(a, b)$  such that

$$US_P(f) - LS_P(f) < \epsilon.$$

**Definition 7.9.2.** Let  $P_1$  and  $P_2$  be two partitions of  $c_n(a, b)$ . Then  $P_2$  is a refinement of  $P_1$  if every point in  $P_1$  is also a point in  $P_2$ , that is,  $P_1 \subset P_2$ .  $\square$

Using this definition, it is possible to prove results similar to those of Lemmas 6.2.1 and 6.2.2. In particular, we have the following lemma:

**Lemma 7.9.1.** Let  $f: D \rightarrow R$ , where  $D \subset R^n$ . Suppose that  $f$  is bounded on  $c_n(a, b) \subset D$ . Then  $\sup_P LS_P(f)$  and  $\inf_P US_P(f)$  exist, and

$$\sup_P LS_P(f) \leq \inf_P US_P(f).$$

**Definition 7.9.3.** Let  $f: c_n(a, b) \rightarrow R$  be a bounded function. Then  $f$  is Riemann integrable on  $c_n(a, b)$  if and only if

$$\sup_P LS_P(f) = \inf_P US_P(f). \quad (7.60)$$

Their common value is called the Riemann integral of  $f$  on  $c_n(a, b)$  and is denoted by  $\int_{c_n(a, b)} f(\mathbf{x}) d\mathbf{x}$ . This is equivalent to the expression  $\int_{a_1}^{b_1} \int_{a_2}^{b_2} \cdots \int_{a_n}^{b_n} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \cdots dx_n$ . For example, for  $n = 2, 3$  we have

$$\int_{c_2(a, b)} f(\mathbf{x}) d\mathbf{x} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x_1, x_2) dx_1 dx_2, \quad (7.61)$$

$$\int_{c_3(a, b)} f(\mathbf{x}) d\mathbf{x} = \int_{a_1}^{b_1} \int_{a_2}^{b_2} \int_{a_3}^{b_3} f(x_1, x_2, x_3) dx_1 dx_2 dx_3. \quad (7.62)$$

The integral in formula (7.61) is called a double Riemann integral, and the one in formula (7.62) is called a triple Riemann integral. In general, for  $n \geq 2$ ,  $\int_{c_n(a, b)} f(\mathbf{x}) d\mathbf{x}$  is called an  $n$ -tuple Riemann integral.  $\square$

The integral  $\int_{c_n(a, b)} f(\mathbf{x}) d\mathbf{x}$  has properties similar to those of a single-variable Riemann integral in Section 6.4. The following theorem is an extension of Theorem 6.3.1.

**Theorem 7.9.2.** If  $f$  is continuous on an  $n$ -dimensional cell  $c_n(a, b)$ , then it is Riemann integrable there.

## 7.9.2. Iterated Riemann Integrals on Cells

The definition of the  $n$ -tuple Riemann integral in Section 7.9.1 does not provide a practicable way to evaluate it. We now show that the evaluation of this integral can be obtained by performing  $n$  Riemann integrals each of which is carried out with respect to one variable. Let us first consider the double integral as in formula (7.61).

**Lemma 7.9.2.** Suppose that  $f$  is real-valued and continuous on  $c_2(a, b)$ . Define the function  $g(x_2)$  as

$$g(x_2) = \int_{a_1}^{b_1} f(x_1, x_2) dx_1.$$

Then  $g(x_2)$  is continuous on  $[a_2, b_2]$ .

*Proof.* Let  $\epsilon > 0$  be given. Since  $f$  is continuous on  $c_2(a, b)$ , which is closed and bounded, then by Theorem 7.3.2,  $f$  is uniformly continuous on  $c_2(a, b)$ . We can therefore find a  $\delta > 0$  such that

$$|f(\xi) - f(\eta)| < \frac{\epsilon}{b_1 - a_1}$$

if  $\|\xi - \eta\| < \delta$ , where  $\xi = (x_1, x_2)'$ ,  $\eta = (y_1, y_2)'$ , and  $x_1, y_1 \in [a_1, b_1]$ ,  $x_2, y_2 \in [a_2, b_2]$ . It follows that if  $|y_2 - x_2| < \delta$ , then

$$\begin{aligned} |g(y_2) - g(x_2)| &= \left| \int_{a_1}^{b_1} [f(x_1, y_2) - f(x_1, x_2)] dx_1 \right| \\ &\leq \int_{a_1}^{b_1} |f(x_1, y_2) - f(x_1, x_2)| dx_1 \\ &< \int_{a_1}^{b_1} \frac{\epsilon}{b_1 - a_1} dx_1, \end{aligned} \quad (7.63)$$

since  $\|(x_1, y_2)' - (x_1, x_2)'\| = |y_2 - x_2| < \delta$ . From inequality (7.63) we conclude that

$$|g(y_2) - g(x_2)| < \epsilon$$

if  $|y_2 - x_2| < \delta$ . Hence,  $g(x_2)$  is continuous on  $[a_2, b_2]$ . Consequently, from Theorem 6.3.1,  $g(x_2)$  is Riemann integrable on  $[a_2, b_2]$ , that is,  $\int_{a_2}^{b_2} g(x_2) dx_2$  exists. We call the integral

$$\int_{a_2}^{b_2} g(x_2) dx_2 = \int_{a_2}^{b_2} \left[ \int_{a_1}^{b_1} f(x_1, x_2) dx_1 \right] dx_2 \quad (7.64)$$

an iterated integral of order 2.  $\square$

The next theorem states that the iterated integral (7.64) is equal to the double integral  $\int_{c_2(a, b)} f(\mathbf{x}) d\mathbf{x}$ .

**Theorem 7.9.3.** If  $f$  is continuous on  $c_2(a, b)$ , then

$$\int_{c_2(a, b)} f(\mathbf{x}) d\mathbf{x} = \int_{a_2}^{b_2} \left[ \int_{a_1}^{b_1} f(x_1, x_2) dx_1 \right] dx_2.$$

*Proof.* Exercise 7.22.  $\square$

We note that the iterated integral in (7.64) was obtained by integrating first with respect to  $x_1$ , then with respect to  $x_2$ . This order of integration

could have been reversed, that is, we could have integrated  $f$  with respect to  $x_2$  and then with respect to  $x_1$ . The result would be the same in both cases. This is based on the following theorem due to Guido Fubini (1879–1943).

**Theorem 7.9.4** (Fubini's Theorem). If  $f$  is continuous on  $c_2(a, b)$ , then

$$\int_{c_2(a, b)} f(\mathbf{x}) \, d\mathbf{x} = \int_{a_2}^{b_2} \left[ \int_{a_1}^{b_1} f(x_1, x_2) \, dx_1 \right] dx_2 = \int_{a_1}^{b_1} \left[ \int_{a_2}^{b_2} f(x_1, x_2) \, dx_2 \right] dx_1$$

*Proof.* See Corwin and Szczarba (1982, page 287).  $\square$

A generalization of this theorem to multiple integrals of order  $n$  is given by the next theorem [see Corwin and Szczarba (1982, Section 11.1)].

**Theorem 7.9.5** (Generalized Fubini's Theorem). If  $f$  is continuous on  $c_n(a, b) = \{\mathbf{x} \mid a_i \leq x_i \leq b_i, i = 1, 2, \dots, n\}$ , then

$$\int_{c_n(a, b)} f(\mathbf{x}) \, d\mathbf{x} = \int_{c_{n-1}^{(i)}(a, b)} \left[ \int_{a_i}^{b_i} f(\mathbf{x}) \, dx_i \right] d\mathbf{x}_{(i)}, \quad i = 1, 2, \dots, n,$$

where  $d\mathbf{x}_{(i)} = dx_1 dx_2 \cdots dx_{i-1} dx_{i+1} \cdots dx_n$  and  $c_{n-1}^{(i)}(a, b)$  is an  $(n-1)$ -dimensional cell such that  $a_1 \leq x_1 \leq b_1, a_2 \leq x_2 \leq b_2, \dots, a_{i-1} \leq x_{i-1} \leq b_{i-1}, a_{i+1} \leq x_{i+1} \leq b_{i+1}, \dots, a_n \leq x_n \leq b_n$ .

### 7.9.3. Integration over General Sets

We now consider  $n$ -tuple Riemann integration over regions in  $R^n$  that are not necessarily cell shaped as in Section 7.9.1.

Let  $f: D \rightarrow R$  be a bounded and continuous function, where  $D$  is a bounded region in  $R^n$ . There exists an  $n$ -dimensional cell  $c_n(a, b)$  such that  $D \subset c_n(a, b)$ . Let  $g: c_n(a, b) \rightarrow R$  be defined as

$$g(\mathbf{x}) = \begin{cases} f(\mathbf{x}), & \mathbf{x} \in D, \\ 0, & \mathbf{x} \notin D. \end{cases}$$

Then

$$\int_{c_n(a, b)} g(\mathbf{x}) \, d\mathbf{x} = \int_D f(\mathbf{x}) \, d\mathbf{x}. \quad (7.65)$$

The integral on the right-hand side of (7.65) is independent of the choice of  $c_n(a, b)$  provided that it contains  $D$ . It should be noted that the function  $g(\mathbf{x})$  may not be continuous on  $Br(D)$ , the boundary of  $D$ . This, however, should not affect the existence of the integral on the left-hand side of (7.65). The reason for this is given in Theorem 7.9.7. First, we need to define the so-called Jordan content of a set.

**Definition 7.9.4.** Let  $D \subset R^n$  be a bounded set such that  $D \subset c_n(a, b)$  for some  $n$ -dimensional cell. Let the function  $\lambda_D: R^n \rightarrow R$  be defined as

$$\lambda_D(\mathbf{x}) = \begin{cases} 1, & \mathbf{x} \in D, \\ 0, & \mathbf{x} \notin D. \end{cases}$$

This is called the characteristic function of  $D$ . Suppose that

$$\sup_P LS_P(\lambda_D) = \inf_P US_P(\lambda_D), \quad (7.66)$$

where  $LS_P(\lambda_D)$  and  $US_P(\lambda_D)$  are, respectively, the lower and upper sums of  $\lambda_D(\mathbf{x})$  with respect to a partition  $P$  of  $c_n(a, b)$ . Then,  $D$  is said to have an  $n$ -dimensional Jordan content denoted by  $\mu_j(D)$ , where  $\mu_j(D)$  is equal to the common value of the terms in equality (7.66). In this case,  $D$  is said to be *Jordan measurable*.  $\square$

The proofs of the next two theorems can be found in Sagan (1974, Chapter 11).

**Theorem 7.9.6.** A bounded set  $D \subset R^n$  is Jordan measurable if and only if its boundary  $Br(D)$  has a Jordan content equal to zero.

**Theorem 7.9.7.** Let  $f: D \rightarrow R$ , where  $D \subset R^n$  is bounded and Jordan measurable. If  $f$  is bounded and continuous in  $D$  except on a set that has a Jordan content equal to zero, then  $\int_D f(\mathbf{x}) d\mathbf{x}$  exists.

It follows from Theorems 7.9.6 and 7.9.7 that the integral in equality (7.75) must exist even though  $g(\mathbf{x})$  may not be continuous on the boundary  $Br(D)$  of  $D$ , since  $Br(D)$  has a Jordan content equal to zero.

**EXAMPLE 7.9.1.** Let  $f(x_1, x_2) = x_1 x_2$  and  $D$  be the region

$$D = \{(x_1, x_2) | x_1^2 + x_2^2 \leq 1, x_1 \geq 0, x_2 \geq 0\}.$$

It is easy to see that  $D$  is contained inside the two-dimensional cell

$$c_2(0, 1) = \{(x_1, x_2) | 0 \leq x_1 \leq 1, 0 \leq x_2 \leq 1\}.$$

Then

$$\int \int_D x_1 x_2 dx_1 dx_2 = \int_0^1 \left[ \int_0^{(1-x_1^2)^{1/2}} x_1 x_2 dx_2 \right] dx_1.$$

We note that for a fixed  $x_1$  in  $[0, 1]$ , the part of the line through  $(x_1, 0)$  that lies inside  $D$  and is parallel to the  $x_2$ -axis is in fact the interval  $0 \leq x_2 \leq (1 - x_1^2)^{1/2}$ . For this reason, the limits of  $x_2$  are 0 and  $(1 - x_1^2)^{1/2}$ . Consequently,

$$\begin{aligned} \int_0^1 \left[ \int_0^{(1-x_1^2)^{1/2}} x_1 x_2 \, dx_2 \right] dx_1 &= \int_0^1 x_1 \left[ \int_0^{(1-x_1^2)^{1/2}} x_2 \, dx_2 \right] dx_1 \\ &= \frac{1}{2} \int_0^1 x_1 (1 - x_1^2) \, dx_1 \\ &= \frac{1}{8}. \end{aligned}$$

In practice, it is not always necessary to make reference to  $c_n(a, b)$  that encloses  $D$  in order to evaluate the integral on  $D$ . Rather, we only need to recognize that the limits of integration in the iterated Riemann integral depend in general on variables that have not yet been integrated out, as was seen in Example 7.9.1. Care should therefore be exercised in correctly identifying the limits of integration. By changing the order of integration (according to Fubini's theorem), it is possible to facilitate the evaluation of the integral.

**EXAMPLE 7.9.2.** Consider  $\int \int_D e^{x_2^3} dx_1 dx_2$ , where  $D$  is the region in the first quadrant bounded by  $x_2 = 1$  and  $x_1 = x_2$ . In this example, it is easier to integrate first with respect to  $x_1$  and then with respect to  $x_2$ . Thus

$$\begin{aligned} \int \int_D e^{x_2^3} dx_1 dx_2 &= \int_0^1 \left[ \int_0^{x_2} e^{x_2^3} dx_1 \right] dx_2 \\ &= \int_0^1 x_2 e^{x_2^3} dx_2 \\ &= \frac{1}{2} (e - 1). \end{aligned}$$

**EXAMPLE 7.9.3.** Consider the integral  $\int \int_D (x_1^2 + x_2^3) dx_1 dx_2$ , where  $D$  is a region in the first quadrant bounded by  $x_2 = x_1^2$  and  $x_1 = x_2^4$ . Hence,

$$\begin{aligned} \int \int_D (x_1^2 + x_2^3) dx_1 dx_2 &= \int_0^1 \left[ \int_{x_2^4}^{\sqrt{x_2}} (x_1^2 + x_2^3) dx_1 \right] dx_2 \\ &= \int_0^1 \left[ \frac{1}{3} (x_2^{3/2} - x_2^{12}) + (\sqrt{x_2} - x_2^4) x_2^3 \right] dx_2 \\ &= \frac{959}{4680}. \end{aligned}$$

#### 7.9.4. Change of Variables in $n$ -Tuple Riemann Integrals

In this section we give an extension of the change of variables formula in Section 6.4.1 to  $n$ -tuple Riemann integrals.

**Theorem 7.9.8.** Suppose that  $D$  is a closed and bounded set in  $R^n$ . Let  $f: D \rightarrow R$  be continuous. Suppose that  $\mathbf{h}: D \rightarrow R^n$  is a one-to-one function with continuous first-order partial derivatives such that the Jacobian determinant,

$$\det[\mathbf{J}_{\mathbf{h}}(\mathbf{x})] = \frac{\partial(h_1, h_2, \dots, h_n)}{\partial(x_1, x_2, \dots, x_n)},$$

is different from zero for all  $\mathbf{x}$  in  $D$ , where  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$  and  $h_i$  is the  $i$ th element of  $\mathbf{h}(i = 1, 2, \dots, n)$ . Then

$$\int_D f(\mathbf{x}) \, d\mathbf{x} = \int_{D'} f[\mathbf{g}(\mathbf{u})] |\det \mathbf{J}_{\mathbf{g}}(\mathbf{u})| \, d\mathbf{u}, \quad (7.67)$$

where  $D' = \mathbf{h}(D)$ ,  $\mathbf{u} = \mathbf{h}(\mathbf{x})$ ,  $\mathbf{g}$  is the inverse function of  $\mathbf{h}$ , and

$$\det \mathbf{J}_{\mathbf{g}}(\mathbf{u}) = \frac{\partial(g_1, g_2, \dots, g_n)}{\partial(u_1, u_2, \dots, u_n)}, \quad (7.68)$$

where  $g_i$  and  $u_i$  are, respectively, the  $i$ th elements of  $\mathbf{g}$  and  $\mathbf{u}$  ( $i = 1, 2, \dots, n$ ).

*Proof.* See, for example, Corwin and Szczarba (1982, Theorem 6.2), or Sagan (1974, Theorem 115.1).  $\square$

**EXAMPLE 7.9.4.** Consider the integral  $\int_D x_1 x_2^2 \, dx_1 \, dx_2$ , where  $D$  is bounded by the four parabolas,  $x_2^2 = x_1$ ,  $x_2^2 = 3x_1$ ,  $x_1^2 = x_2$ ,  $x_1^2 = 4x_2$ . Let  $u_1 = x_2^2/x_1$ ,  $u_2 = x_1^2/x_2$ . The inverse transformation is given by

$$x_1 = (u_1 u_2^2)^{1/3}, \quad x_2 = (u_1^2 u_2)^{1/3}.$$

From formula (7.68) we have

$$\frac{\partial(g_1, g_2)}{\partial(u_1, u_2)} = \frac{\partial(x_1, x_2)}{\partial(u_1, u_2)} = -\frac{1}{3}.$$

By applying formula (7.67) we obtain

$$\int \int_D x_1 x_2^2 \, dx_1 \, dx_2 = \frac{1}{3} \int \int_{D'} u_1^{5/3} u_2^{4/3} \, du_1 \, du_2,$$

where  $D'$  is a rectangular region in the  $u_1 u_2$  space bounded by the lines  $u_1 = 1, 3$ ;  $u_2 = 1, 4$ . Hence,

$$\begin{aligned} \int \int_D x_1 x_2^2 \, dx_1 \, dx_2 &= \frac{1}{3} \int_1^3 u_1^{5/3} \, du_1 \int_1^4 u_2^{4/3} \, du_2 \\ &= \frac{3}{56} (3^{8/3} - 1) (4^{7/3} - 1). \end{aligned}$$



### 7.10. DIFFERENTIATION UNDER THE INTEGRAL SIGN

Suppose that  $f(x_1, x_2, \dots, x_n)$  is a real-valued function defined on  $D \subset R^n$ . If some of the  $x_i$ 's, for example,  $x_{m+1}, x_{m+2}, \dots, x_n$  ( $n > m$ ), are integrated out, we obtain a function that depends only on the remaining variables. In this section we discuss conditions under which the latter function is differentiable. For simplicity, we shall only consider functions of  $n = 2$  variables.

**Theorem 7.10.1.** Let  $f: D \rightarrow R$ , where  $D \subset R^2$  contains the two-dimensional cell  $c_2(a, b) = \{(x_1, x_2) | a_1 \leq x_1 \leq b_1, a_2 \leq x_2 \leq b_2\}$ . Suppose that  $f$  is continuous and has a continuous first-order partial derivative with respect to  $x_2$  in  $D$ . Then, for  $a_2 < x_2 < b_2$ ,

$$\frac{d}{dx_2} \int_{a_1}^{b_1} f(x_1, x_2) dx_1 = \int_{a_1}^{b_1} \frac{\partial f(x_1, x_2)}{\partial x_2} dx_1. \quad (7.69)$$

*Proof.* Let  $h(x_2)$  be defined on  $[a_2, b_2]$  as

$$h(x_2) = \int_{a_1}^{b_1} \frac{\partial f(x_1, x_2)}{\partial x_2} dx_1, \quad a_2 \leq x_2 \leq b_2.$$

Since  $\partial f / \partial x_2$  is continuous, then by Lemma 7.9.2,  $h(x_2)$  is continuous on  $[a_2, b_2]$ . Now, let  $t$  be such that  $a_2 < t < b_2$ . By integrating  $h(x_2)$  over the interval  $[a_2, t]$  we obtain

$$\int_{a_2}^t h(x_2) dx_2 = \int_{a_2}^t \left[ \int_{a_1}^{b_1} \frac{\partial f(x_1, x_2)}{\partial x_2} dx_1 \right] dx_2. \quad (7.70)$$

The order of integration in (7.70) can be reversed by Theorem 7.9.4. We then have

$$\begin{aligned} \int_{a_2}^t h(x_2) dx_2 &= \int_{a_1}^{b_1} \left[ \int_{a_2}^t \frac{\partial f(x_1, x_2)}{\partial x_2} dx_2 \right] dx_1 \\ &= \int_{a_1}^{b_1} [f(x_1, t) - f(x_1, a_2)] dx_1 \\ &= \int_{a_1}^{b_1} f(x_1, t) dx_1 - \int_{a_1}^{b_1} f(x_1, a_2) dx_1 \\ &= F(t) - F(a_2), \end{aligned} \quad (7.71)$$

where  $F(y) = \int_{a_1}^{b_1} f(x_1, y) dx_1$ . If we now apply Theorem 6.4.8 and differentiate the two sides of (7.71) with respect to  $t$  we obtain  $h(t) = F'(t)$ , that is,

$$\int_{a_1}^{b_1} \frac{\partial f(x_1, t)}{\partial t} dx_1 = \frac{d}{dt} \int_{a_1}^{b_1} f(x_1, t) dx_1. \quad (7.72)$$

Formula (7.69) now follows from formula (7.72) on replacing  $t$  with  $x_2$ .  $\square$

**Theorem 7.10.2.** Let  $f$  and  $D$  be the same as in Theorem 7.10.1. Furthermore, let  $\lambda(x_2)$  and  $\theta(x_2)$  be functions defined and having continuous derivatives on  $[a_2, b_2]$  such that  $a_1 \leq \lambda(x_2) \leq \theta(x_2) \leq b_1$  for all  $x_2$  in  $[a_2, b_2]$ . Then the function  $G: [a_2, b_2] \rightarrow R$  defined by

$$G(x_2) = \int_{\lambda(x_2)}^{\theta(x_2)} f(x_1, x_2) dx_1$$

is differentiable for  $a_2 < x_2 < b_2$ , and

$$\frac{dG}{dx_2} = \int_{\lambda(x_2)}^{\theta(x_2)} \frac{\partial f(x_1, x_2)}{\partial x_2} dx_1 + \theta'(x_2) f[\theta(x_2), x_2] - \lambda'(x_2) f[\lambda(x_2), x_2].$$

*Proof.* Let us write  $G(x_2)$  as  $H(\lambda, \theta, x_2)$ . Since both of  $\lambda$  and  $\theta$  depend on  $x_2$ , then by applying the total derivative formula [see formula (7.12)] to  $H$  we obtain

$$\frac{dH}{dx_2} = \frac{\partial H}{\partial \lambda} \frac{d\lambda}{dx_2} + \frac{\partial H}{\partial \theta} \frac{d\theta}{dx_2} + \frac{\partial H}{\partial x_2}. \quad (7.73)$$

Now, by Theorem 6.4.8,

$$\begin{aligned} \frac{\partial H}{\partial \theta} &= \frac{\partial}{\partial \theta} \int_{\lambda}^{\theta} f(x_1, x_2) dx_1 = f(\theta, x_2), \\ \frac{\partial H}{\partial \lambda} &= \frac{\partial}{\partial \lambda} \int_{\lambda}^{\theta} f(x_1, x_2) dx_1 \\ &= - \frac{\partial}{\partial \lambda} \int_{\theta}^{\lambda} f(x_1, x_2) dx_1 = -f(\lambda, x_2). \end{aligned}$$

Furthermore, by Theorem 7.10.1,

$$\frac{\partial H}{\partial x_2} = \frac{\partial}{\partial x_2} \int_{\lambda}^{\theta} f(x_1, x_2) dx_1 = \int_{\lambda}^{\theta} \frac{\partial f(x_1, x_2)}{\partial x_2} dx_1.$$

By making the proper substitution in formula (7.73) we finally conclude that

$$\begin{aligned} \frac{d}{dx_2} \int_{\lambda(x_2)}^{\theta(x_2)} f(x_1, x_2) dx_1 &= \int_{\lambda(x_2)}^{\theta(x_2)} \frac{\partial f(x_1, x_2)}{\partial x_2} dx_1 + \theta'(x_2) f[\theta(x_2), x_2] \\ &\quad - \lambda'(x_2) f[\lambda(x_2), x_2]. \quad \square \end{aligned}$$

EXAMPLE 7.10.1.

$$\begin{aligned} \frac{d}{dx_2} \int_{x_2^2}^{\cos x_2} (x_1 x_2^2 - 1) e^{-x_1} dx_1 &= \int_{x_2^2}^{\cos x_2} 2x_1 x_2 e^{-x_1} dx_1 \\ &\quad - \sin x_2 (x_2^2 \cos x_2 - 1) e^{-\cos x_2} \\ &\quad - 2x_2 (x_2^4 - 1) e^{-x_2^2}. \end{aligned}$$

Theorems 7.10.1 and 7.10.2 can be used to evaluate certain integrals of the form  $\int_a^b f(x) dx$ . For example, consider the integral

$$I = \int_0^\pi x^2 \cos x dx.$$

Define the function

$$F(x_2) = \int_0^\pi \cos(x_1 x_2) dx_1,$$

where  $x_2 \geq 1$ . Then

$$F(x_2) = \frac{1}{x_2} \sin(x_1 x_2) \Big|_{x_1=0}^{x_1=\pi} = \frac{1}{x_2} \sin(\pi x_2).$$

If we now differentiate  $F(x_2)$  two times, we obtain

$$F''(x_2) = \frac{2 \sin(\pi x_2) - 2\pi x_2 \cos(\pi x_2) - \pi^2 x_2^2 \sin(\pi x_2)}{x_2^3}.$$

Thus

$$\int_0^\pi x_1^2 \cos(x_1 x_2) dx_1 = \frac{2\pi x_2 \cos(\pi x_2) + \pi^2 x_2^2 \sin(\pi x_2) - 2 \sin(\pi x_2)}{x_2^3}.$$

By replacing  $x_2$  with 1 we obtain

$$I = \int_0^\pi x_1^2 \cos x_1 \, dx_1 = -2\pi.$$

### 7.11. APPLICATIONS IN STATISTICS

Multidimensional calculus provides a theoretical framework for the study of multivariate distributions, that is, joint distributions of several random variables. It can also be used to estimate the parameters of a statistical model. We now provide details of some of these applications.

Let  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$  be a random vector. The distribution of  $\mathbf{X}$  is characterized by its cumulative distribution function, namely,

$$F(\mathbf{x}) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n), \quad (7.74)$$

where  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ . If  $F(\mathbf{x})$  is continuous and has an  $n$ th-order mixed partial derivative with respect to  $x_1, x_2, \dots, x_n$ , then the function

$$f(\mathbf{x}) = \frac{\partial^n F(\mathbf{x})}{\partial x_1 \partial x_2 \cdots \partial x_n},$$

is called the density function of  $\mathbf{X}$ . In this case, formula (7.74) can be written in the form

$$F(\mathbf{x}) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(\mathbf{z}) \, d\mathbf{z}.$$

where  $\mathbf{z} = (z_1, z_2, \dots, z_n)'$ . If the random variable  $X_i$  ( $i = 1, 2, \dots, n$ ) is considered separately, then its distribution function is called the  $i$ th marginal distribution of  $\mathbf{X}$ . Its density function  $f_i(x_i)$ , called the  $i$ th marginal density function, can be obtained by integrating out the remaining  $n - 1$  variables from  $f(\mathbf{x})$ . For example, if  $\mathbf{X} = (X_1, X_2)'$ , then the marginal density function of  $X_1$  is

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) \, dx_2.$$

Similarly, the marginal density function of  $X_2$  is

$$f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) \, dx_1.$$

In particular, if  $X_1, X_2, \dots, X_n$  are independent random variables, then the density function of  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$  is the product of all the associated marginal density functions, that is,  $f(\mathbf{x}) = \prod_{i=1}^n f_i(x_i)$ .

If only  $n-2$  variables are integrated out from  $f(\mathbf{x})$ , we obtain the so-called bivariate density function of the remaining two variables. For example, if  $\mathbf{X} = (X_1, X_2, X_3, X_4)'$ , the bivariate density function of  $x_1$  and  $x_2$  is

$$f_{12}(x_1, x_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_1, x_2, x_3, x_4) dx_3 dx_4.$$

Now, the mean of  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$  is  $\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)'$ , where

$$\mu_i = \int_{-\infty}^{\infty} x_i f_i(x_i) dx_i, \quad i = 1, 2, \dots, n.$$

The variance-covariance matrix of  $\mathbf{X}$  is the  $n \times n$  matrix  $\boldsymbol{\Sigma} = (\sigma_{ij})$ , where

$$\sigma_{ij} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_i - \mu_i)(x_j - \mu_j) f_{ij}(x_i, x_j) dx_i dx_j,$$

where  $\mu_i$  and  $\mu_j$  are the means of  $X_i$  and  $X_j$ , respectively, and  $f_{ij}(x_i, x_j)$  is the bivariate density function of  $X_i$  and  $X_j$ ,  $i \neq j$ . If  $i = j$ , then  $\sigma_{ii}$  is the variance of  $X_i$ , where

$$\sigma_{ii} = \int_{-\infty}^{\infty} (x_i - \mu_i)^2 f_i(x_i) dx_i, \quad i = 1, 2, \dots, n.$$

### 7.11.1. Transformations of Random Vectors

In this section we consider a multivariate extension of formula (6.73) regarding the density function of a function of a single random variable. This is given in the next theorem.

**Theorem 7.11.1.** Let  $\mathbf{X}$  be a random vector with a continuous density function  $f(\mathbf{x})$ . Let  $\mathbf{g}: D \rightarrow R^n$ , where  $D$  is an open subset of  $R^n$  such that  $P(\mathbf{X} \in D) = 1$ . Suppose that  $\mathbf{g}$  satisfies the conditions of the inverse function theorem (Theorem 7.6.1), namely the following:

- i.  $\mathbf{g}$  has continuous first-order partial derivatives in  $D$ .
- ii. The Jacobian matrix  $\mathbf{J}_{\mathbf{g}}(\mathbf{x})$  is nonsingular in  $D$ , that is,

$$\det[\mathbf{J}_{\mathbf{g}}(\mathbf{x})] = \frac{\partial(g_1, g_2, \dots, g_n)}{\partial(x_1, x_2, \dots, x_n)} \neq 0$$

for all  $\mathbf{x} \in D$ , where  $g_i$  is the  $i$ th element of  $\mathbf{g}$  ( $i = 1, 2, \dots, n$ ).

Then the density function of  $\mathbf{Y} = \mathbf{g}(\mathbf{X})$  is given by

$$h(\mathbf{y}) = f[\mathbf{g}^{-1}(\mathbf{y})] |\det[\mathbf{J}_{\mathbf{g}^{-1}}(\mathbf{y})]|,$$

where  $\mathbf{g}^{-1}$  is the inverse function of  $\mathbf{g}$ .

*Proof.* By Theorem 7.6.1, the inverse function of  $\mathbf{g}$  exists. Let us therefore write  $\mathbf{X} = \mathbf{g}^{-1}(\mathbf{Y})$ . Now, the cumulative distribution function of  $\mathbf{Y}$  is

$$\begin{aligned} H(\mathbf{y}) &= P[g_1(\mathbf{X}) \leq y_1, \quad g_2(\mathbf{X}) \leq y_2, \dots, \quad g_n(\mathbf{X}) \leq y_n] \\ &= \int_{A_n} f(\mathbf{x}) \, d\mathbf{x}, \end{aligned} \quad (7.75)$$

where  $A_n = \{\mathbf{x} \in D | g_i(\mathbf{x}) \leq y_i, \quad i = 1, 2, \dots, n\}$ . If we make the change of variable  $\mathbf{w} = \mathbf{g}(\mathbf{x})$  in formula (7.75), then, by applying Theorem 7.9.8 with  $\mathbf{g}^{-1}(\mathbf{w})$  used instead of  $\mathbf{g}(\mathbf{u})$ , we obtain

$$\int_{A_n} f(\mathbf{x}) \, d\mathbf{x} = \int_{B_n} f[\mathbf{g}^{-1}(\mathbf{w})] |\det[\mathbf{J}_{\mathbf{g}^{-1}}(\mathbf{w})]| \, d\mathbf{w},$$

where  $B_n = \mathbf{g}(A_n) = \{\mathbf{g}(\mathbf{x}) | g_i(\mathbf{x}) \leq y_i, \quad i = 1, 2, \dots, n\}$ . Thus

$$H(\mathbf{y}) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \cdots \int_{-\infty}^{y_n} f[\mathbf{g}^{-1}(\mathbf{w})] |\det[\mathbf{J}_{\mathbf{g}^{-1}}(\mathbf{w})]| \, d\mathbf{w}.$$

It follows that the density function of  $\mathbf{Y}$  is

$$h(\mathbf{y}) = f[\mathbf{g}^{-1}(\mathbf{y})] \left| \frac{\partial(g_1^{-1}, g_2^{-1}, \dots, g_n^{-1})}{\partial(y_1, y_2, \dots, y_n)} \right|, \quad (7.76)$$

where  $g_i^{-1}$  is the  $i$ th element of  $\mathbf{g}^{-1}$  ( $i = 1, 2, \dots, n$ ).  $\square$

EXAMPLE 7.11.1. Let  $\mathbf{X} = (X_1, X_2)'$ , where  $X_1$  and  $X_2$  are independent random variables that have the standard normal distribution. Here, the density function of  $\mathbf{X}$  is the product of the density functions of  $X_1$  and  $X_2$ . Thus

$$f(\mathbf{x}) = \frac{1}{2\pi} \exp\left[-\frac{1}{2}(x_1^2 + x_2^2)\right], \quad -\infty < x_1, x_2 < \infty.$$

Let  $\mathbf{Y} = (Y_1, Y_2)'$  be defined as

$$\begin{aligned} Y_1 &= X_1 + X_2, \\ Y_2 &= X_1 - 2X_2. \end{aligned}$$

In this case, the set  $D$  in Theorem 7.11.1 is  $R^2$ ,  $g_1(\mathbf{x}) = x_1 + x_2$ ,  $g_2(\mathbf{x}) = x_1 - 2x_2$ ,  $g_1^{-1}(\mathbf{y}) = x_1 = \frac{1}{3}(2y_1 + y_2)$ ,  $g_2^{-1}(\mathbf{y}) = x_2 = \frac{1}{3}(y_1 - y_2)$ , and

$$\frac{\partial(g_1^{-1}, g_2^{-1})}{\partial(y_1, y_2)} = \det \begin{bmatrix} \frac{2}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{bmatrix} = -\frac{1}{3}.$$

Hence, by formula (7.76), the density function of  $\mathbf{y}$  is

$$\begin{aligned} h(\mathbf{y}) &= \frac{1}{2\pi} \exp \left[ -\frac{1}{2} \left( \frac{2y_1 + y_2}{3} \right)^2 - \frac{1}{2} \left( \frac{y_1 - y_2}{3} \right)^2 \right] \times \frac{1}{3} \\ &= \frac{1}{6\pi} \exp \left[ -\frac{1}{18} (5y_1^2 + 2y_1y_2 + 2y_2^2) \right], \quad -\infty < y_1, y_2 < \infty. \end{aligned}$$

EXAMPLE 7.11.2. Suppose that it is desired to determine the density function of the random variable  $V = X_1 + X_2$ , where  $X_1 \geq 0$ ,  $X_2 \geq 0$ , and  $\mathbf{X} = (X_1, X_2)'$  has a continuous density function  $f(x_1, x_2)$ . This can be accomplished in two ways:

- i. Let  $Q(v)$  denote the cumulative distribution function of  $V$  and let  $q(v)$  be its density function. Then

$$\begin{aligned} Q(v) &= P(X_1 + X_2 \leq v) \\ &= \iint_A f(x_1, x_2) dx_1 dx_2, \end{aligned}$$

where  $A = \{(x_1, x_2) | x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq v\}$ . We can write  $Q(v)$  as

$$Q(v) = \int_0^v \left[ \int_0^{v-x_2} f(x_1, x_2) dx_1 \right] dx_2.$$

If we now apply Theorem 7.10.2, we obtain

$$\begin{aligned} q(v) &= \frac{dQ}{dv} = \int_0^v \frac{\partial}{\partial v} \left[ \int_0^{v-x_2} f(x_1, x_2) dx_1 \right] dx_2 \\ &= \int_0^v f(v - x_2, x_2) dx_2. \end{aligned} \tag{7.77}$$

- ii. Consider the following transformation:

$$\begin{aligned} Y_1 &= X_1 + X_2, \\ Y_2 &= X_2. \end{aligned}$$

Then

$$\begin{aligned}X_1 &= Y_1 - Y_2, \\X_2 &= Y_2.\end{aligned}$$

By Theorem 7.11.1, the density function of  $\mathbf{Y} = (Y_1, Y_2)'$  is

$$\begin{aligned}h(y_1, y_2) &= f(y_1 - y_2, y_2) \left| \frac{\partial(x_1, x_2)}{\partial(y_1, y_2)} \right| \\&= f(y_1 - y_2, y_2) \left| \det \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \right| \\&= f(y_1 - y_2, y_2), \quad y_1 \geq y_2 \geq 0.\end{aligned}$$

By integrating  $y_2$  out we obtain the marginal density function of  $Y_1 = V$ , namely,

$$q(v) = \int_0^v f(y_1 - y_2, y_2) dy_2 = \int_0^v f(v - x_2, x_2) dx_2.$$

This is identical to the density function given in formula (7.77).

### 7.11.2. Maximum Likelihood Estimation

Let  $X_1, X_2, \dots, X_n$  be a sample of size  $n$  from a population whose distribution depends on a set of  $p$  parameters, namely  $\theta_1, \theta_2, \dots, \theta_p$ . We can regard this sample as forming a random vector  $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ . Suppose that  $\mathbf{X}$  has the density function  $f(\mathbf{x}, \boldsymbol{\theta})$ , where  $\mathbf{x} = (x_1, x_2, \dots, x_n)'$  and  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_p)'$ . This density function is usually referred to as the likelihood function of  $\mathbf{X}$ ; we denote it by  $L(\mathbf{x}, \boldsymbol{\theta})$ .

For a given sample, the maximum likelihood estimate of  $\boldsymbol{\theta}$ , denoted by  $\hat{\boldsymbol{\theta}}$ , is the value of  $\boldsymbol{\theta}$  that maximizes  $L(\mathbf{x}, \boldsymbol{\theta})$ . If  $L(\mathbf{x}, \boldsymbol{\theta})$  has partial derivatives with respect to  $\theta_1, \theta_2, \dots, \theta_p$ , then  $\hat{\boldsymbol{\theta}}$  is often obtained by solving the equations

$$\frac{\partial L(\mathbf{x}, \hat{\boldsymbol{\theta}})}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, p.$$

In most situations, it is more convenient to work with the natural logarithm of  $L(\mathbf{x}, \boldsymbol{\theta})$ ; its maxima are attained at the same points as those of  $L(\mathbf{x}, \boldsymbol{\theta})$ . Thus  $\hat{\boldsymbol{\theta}}$  satisfies the equation

$$\frac{\partial \log [L(\mathbf{x}, \hat{\boldsymbol{\theta}})]}{\partial \theta_i} = 0, \quad i = 1, 2, \dots, p. \quad (7.78)$$

Equations (7.78) are known as the *likelihood equations*.



EXAMPLE 7.11.3. Suppose that  $X_1, X_2, \dots, X_n$  form a sample of size  $n$  from a normal distribution with an unknown mean  $\mu$  and a variance  $\sigma^2$ . Here,  $\boldsymbol{\theta} = (\mu, \sigma^2)'$ , and the likelihood function is given by

$$L(\mathbf{x}, \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right],$$

Let  $L^*(\mathbf{x}, \boldsymbol{\theta}) = \log L(\mathbf{x}, \boldsymbol{\theta})$ . Then

$$L^*(\mathbf{x}, \boldsymbol{\theta}) = -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 - \frac{n}{2} \log(2\pi\sigma^2).$$

The likelihood equations in formula (7.78) are of the form

$$\frac{\partial L^*}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \hat{\mu}) = 0 \quad (7.79)$$

$$\frac{\partial L^*}{\partial \sigma^2} = \frac{1}{2\hat{\sigma}^4} \sum_{i=1}^n (x_i - \hat{\mu})^2 - \frac{n}{2\hat{\sigma}^2} = 0. \quad (7.80)$$

Equations (7.79) and (7.80) can be written as

$$n(\bar{x} - \hat{\mu}) = 0, \quad (7.81)$$

$$\sum_{i=1}^n (x_i - \hat{\mu})^2 - n\hat{\sigma}^2 = 0, \quad (7.82)$$

where  $\bar{x} = (1/n)\sum_{i=1}^n x_i$ . If  $n \geq 2$ , then equations (7.81) and (7.82) have the solution

$$\begin{aligned} \hat{\mu} &= \bar{x}, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2. \end{aligned}$$

These are the maximum likelihood estimates of  $\mu$  and  $\sigma^2$ , respectively.

It can be verified that  $\hat{\mu}$  and  $\hat{\sigma}^2$  are indeed the values of  $\mu$  and  $\sigma^2$  that maximize  $L^*(\mathbf{x}, \boldsymbol{\theta})$ . To show this, let us consider the Hessian matrix  $\mathbf{A}$  of second-order partial derivatives of  $L^*$  (see formula 7.34),

$$\mathbf{A} = \begin{bmatrix} \frac{\partial^2 L^*}{\partial \mu^2} & \frac{\partial^2 L^*}{\partial \mu \partial \sigma^2} \\ \frac{\partial^2 L^*}{\partial \mu \partial \sigma^2} & \frac{\partial^2 L^*}{\partial \sigma^4} \end{bmatrix}.$$

Hence, for  $\mu = \hat{\mu}$  and  $\sigma^2 = \hat{\sigma}^2$ ,

$$\begin{aligned}\frac{\partial^2 L^*}{\partial \mu^2} &= -\frac{n}{\hat{\sigma}^2}, \\ \frac{\partial^2 L^*}{\partial \mu \partial \sigma^2} &= -\frac{1}{\hat{\sigma}^4} \sum_{i=1}^n (x_i - \hat{\mu}) = 0, \\ \frac{\partial^2 L^*}{\partial \sigma^4} &= -\frac{n}{2\hat{\sigma}^4}.\end{aligned}$$

Thus  $\partial^2 L^* / \partial \mu^2 < 0$  and  $\det(\mathbf{A}) = n^2 / 2\hat{\sigma}^6 > 0$ . Therefore, by Corollary 7.7.1,  $(\hat{\mu}, \hat{\sigma}^2)$  is a point of local maximum of  $L^*$ . Since it is the only maximum, it must also be the absolute maximum.

Maximum likelihood estimators have interesting asymptotic properties. For more information on these properties, see, for example, Bickel and Doksum (1977, Section 4.4).

### 7.11.3. Comparison of Two Unbiased Estimators

Let  $X_1$  and  $X_2$  be two unbiased estimators of a parameter  $\mu$ . Suppose that  $\mathbf{X} = (X_1, X_2)'$  has the density function  $f(x_1, x_2)$ ,  $-\infty < x_1, x_2 < \infty$ . To compare these estimators, we may consider the probability that one estimator, for example,  $X_1$ , is closer to  $\mu$  than the other,  $X_2$ , that is,

$$p = P[|X_1 - \mu| < |X_2 - \mu|].$$

This probability can be expressed as

$$p = \int \int_D f(x_1, x_2) dx_1 dx_2, \quad (7.83)$$

where  $D = \{(x_1, x_2) | |x_1 - \mu| < |x_2 - \mu|\}$ . Let us now make the following change of variables using polar coordinates:

$$x_1 - \mu = r \cos \theta, \quad x_2 - \mu = r \sin \theta.$$

By applying formula (7.67), the integral in (7.83) can be written as

$$\begin{aligned}p &= \int \int_D \tilde{g}(r, \theta) \left| \frac{\partial(x_1, x_2)}{\partial(r, \theta)} \right| dr d\theta \\ &= \int \int_D \tilde{g}(r, \theta) r dr d\theta,\end{aligned}$$

where  $\tilde{g}(r, \theta) = f(\mu + r \cos \theta, \mu + r \sin \theta)$  and

$$D' \left\{ (r, \theta) \mid 0 \leq r < \infty, \frac{\pi}{4} \leq \theta \leq \frac{3\pi}{4}, \frac{5\pi}{4} \leq \theta \leq \frac{7\pi}{4} \right\}.$$

In particular, if  $\mathbf{X}$  has the bivariate normal density, then

$$\begin{aligned} f(x_1, x_2) &= \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \\ &\times \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \frac{(x_1 - \mu)^2}{\sigma_1^2} - \frac{2\rho(x_1 - \mu)(x_2 - \mu)}{\sigma_1\sigma_2} \right. \right. \\ &\quad \left. \left. + \frac{(x_2 - \mu)^2}{\sigma_2^2} \right] \right\}, \quad -\infty < x_1, x_2 < \infty, \end{aligned}$$

and

$$\begin{aligned} \tilde{g}(r, \theta) &= \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \\ &\times \exp \left\{ -\frac{r^2}{2(1-\rho^2)} \left[ \frac{\cos^2 \theta}{\sigma_1^2} - \frac{2\rho \cos \theta \sin \theta}{\sigma_1\sigma_2} + \frac{\sin^2 \theta}{\sigma_2^2} \right] \right\}, \end{aligned}$$

where  $\sigma_1^2$  and  $\sigma_2^2$  are the variances of  $X_1$  and  $X_2$ , respectively, and  $\rho$  is their correlation coefficient. In this case,

$$p = 2 \int_{\pi/4}^{3\pi/4} \left[ \int_0^\infty \tilde{g}(r, \theta) r dr \right] d\theta.$$

It can be shown (see Lowerre, 1983) that

$$p = 1 - \frac{1}{\pi} \text{Arctan} \left[ \frac{2\sigma_1\sigma_2(1-\rho^2)^{1/2}}{\sigma_2^2 - \sigma_1^2} \right] \quad (7.84)$$

if  $\sigma_2 > \sigma_1$ . A large value of  $p$  indicates that  $X_1$  is closer to  $\mu$  than  $X_2$ , which means that  $X_1$  is a better estimator of  $\mu$  than  $X_2$ .

#### 7.11.4. Best Linear Unbiased Estimation

Let  $X_1, X_2, \dots, X_n$  be independent and identically distributed random variables with a common mean  $\mu$  and a common variance  $\sigma^2$ . An estimator of

the form  $\hat{\phi} = \sum_{i=1}^n a_i X_i$ , where the  $a_i$ 's are constants, is said to be a linear estimator of  $\mu$ . This estimator is unbiased if  $E(\hat{\phi}) = \mu$ , that is, if  $\sum_{i=1}^n a_i = 1$ , since  $E(X_i) = \mu$  for  $i = 1, 2, \dots, n$ . The variance of  $\hat{\phi}$  is given by

$$\text{Var}(\hat{\phi}) = \sigma^2 \sum_{i=1}^n a_i^2.$$

The smaller the variance of  $\hat{\phi}$ , the more efficient  $\hat{\phi}$  is as an estimator of  $\mu$ . In particular, if  $a_1, a_2, \dots, a_n$  are chosen so that  $\text{Var}(\hat{\phi})$  attains a minimum value, then  $\hat{\phi}$  will have the smallest variance among all unbiased linear estimators of  $\mu$ . In this case,  $\hat{\phi}$  is called the *best linear unbiased estimator* (BLUE) of  $\mu$ .

Thus to find the BLUE of  $\mu$  we need to minimize the function  $f = \sum_{i=1}^n a_i^2$  subject to the constraint  $\sum_{i=1}^n a_i = 1$ . This minimization problem can be solved using the method of Lagrange multipliers. Let us therefore write  $F$  [see formula (7.41)] as

$$F = \sum_{i=1}^n a_i^2 + \lambda \left( \sum_{i=1}^n a_i - 1 \right),$$

$$\frac{\partial F}{\partial a_i} = 2a_i + \lambda = 0, \quad i = 1, 2, \dots, n.$$

Hence,  $a_i = -\lambda/2$  ( $i = 1, 2, \dots, n$ ). Using the constraint  $\sum_{i=1}^n a_i = 1$ , we conclude that  $\lambda = -2/n$ . Thus  $a_i = 1/n$ ,  $i = 1, 2, \dots, n$ . To verify that this solution minimizes  $f$ , we need to consider the signs of  $\Delta_1, \Delta_2, \dots, \Delta_{n-1}$ , where  $\Delta_i$  is the determinant of  $\mathbf{B}_i$  (see Section 7.8). Here,  $\mathbf{B}_1$  is an  $(n+1) \times (n+1)$  matrix of the form

$$\mathbf{B}_1 = \begin{bmatrix} 2\mathbf{I}_n & \mathbf{1}_n \\ \mathbf{1}'_n & 0 \end{bmatrix}.$$

It follows that

$$\Delta_1 = \det(\mathbf{B}_1) = -\frac{n2^n}{2} < 0,$$

$$\Delta_2 = -\frac{(n-1)2^{n-1}}{2} < 0,$$

$$\vdots$$

$$\Delta_{n-1} = -2^2 < 0.$$

Since the number of constraints,  $m = 1$ , is odd, then by the sufficient

conditions described in Section 7.8 we must have a local minimum when  $a_i = 1/n$ ,  $i = 1, 2, \dots, n$ . Since this is the only local minimum in  $R^n$ , it must be the absolute minimum. Note that for such values of  $a_1, a_2, \dots, a_n$ ,  $\hat{\phi}$  is the sample mean  $\bar{X}_n$ . We conclude that the sample mean is the most efficient (in terms of variance) unbiased linear estimator of  $\mu$ .

### 7.11.5. Optimal Choice of Sample Sizes in Stratified Sampling

In stratified sampling, a finite population of  $N$  units is divided into  $r$  subpopulations, called strata, of sizes  $N_1, N_2, \dots, N_r$ . From each stratum a random sample is drawn, and the drawn samples are obtained independently in the different strata. Let  $n_i$  be the size of the sample drawn from the  $i$ th stratum ( $i = 1, 2, \dots, r$ ). Let  $y_{ij}$  denote the response value obtained from the  $j$ th unit within the  $i$ th stratum ( $i = 1, 2, \dots, r$ ;  $j = 1, 2, \dots, n_i$ ). The population mean  $\bar{Y}$  is

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^r \sum_{j=1}^{N_i} y_{ij} = \frac{1}{N} \sum_{i=1}^r N_i \bar{Y}_i,$$

where  $\bar{Y}_i$  is the true mean for the  $i$ th stratum ( $i = 1, 2, \dots, r$ ). A stratified estimate of  $\bar{Y}$  is  $\bar{y}_{\text{st}}$  (st for stratified), where

$$\bar{y}_{\text{st}} = \frac{1}{N} \sum_{i=1}^r N_i \bar{y}_i,$$

in which  $\bar{y}_i = (1/n_i) \sum_{j=1}^{n_i} y_{ij}$  is the mean of the sample from the  $i$ th stratum ( $i = 1, 2, \dots, r$ ). If, in every stratum,  $\bar{y}_i$  is unbiased for  $\bar{Y}_i$ , then  $\bar{y}_{\text{st}}$  is an unbiased estimator of  $\bar{Y}$ . The variance of  $\bar{y}_{\text{st}}$  is

$$\text{Var}(\bar{y}_{\text{st}}) = \frac{1}{N^2} \sum_{i=1}^r N_i^2 \text{Var}(\bar{y}_i).$$

Since  $\bar{y}_i$  is the mean of a random sample from a finite population, then its variance is given by (see Cochran, 1963, page 22)

$$\text{Var}(\bar{y}_i) = \frac{S_i^2}{n_i} (1 - f_i), \quad i = 1, 2, \dots, r,$$

where  $f_i = n_i/N_i$ , and

$$S_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (y_{ij} - \bar{Y}_i)^2.$$

Hence,

$$\text{Var}(\bar{y}_{\text{st}}) = \sum_{i=1}^r \frac{1}{n_i} L_i^2 S_i^2 (1 - f_i),$$

where  $L_i = N_i/N$  ( $i = 1, 2, \dots, r$ ).

The sample sizes  $n_1, n_2, \dots, n_r$  can be chosen by the sampler in an optimal way, the optimality criterion being the minimization of  $\text{Var}(\bar{y}_{\text{st}})$  for a specified cost of taking the samples. Here, the cost is defined by the formula

$$\text{cost} = c_0 + \sum_{i=1}^r c_i n_i,$$

where  $c_i$  is the cost per unit in the  $i$ th stratum ( $i = 1, 2, \dots, r$ ) and  $c_0$  is the overhead cost. Thus the optimal choice of the sample sizes is reduced to finding the values of  $n_1, n_2, \dots, n_r$  that minimize  $\sum_{i=1}^r (1/n_i) L_i^2 S_i^2 (1 - f_i)$  subject to the constraint

$$\sum_{i=1}^r c_i n_i = d - c_0, \quad (7.85)$$

where  $d$  is a constant. Using the method of Lagrange multipliers, we write

$$\begin{aligned} F &= \sum_{i=1}^r \frac{1}{n_i} L_i^2 S_i^2 (1 - f_i) + \lambda \left( \sum_{i=1}^r c_i n_i + c_0 - d \right) \\ &= \sum_{i=1}^r \frac{1}{n_i} L_i^2 S_i^2 - \sum_{i=1}^r \frac{1}{N_i} L_i^2 S_i^2 + \lambda \left( \sum_{i=1}^r c_i n_i + c_0 - d \right). \end{aligned}$$

Differentiating with respect to  $n_i$  ( $i = 1, 2, \dots, r$ ), we obtain

$$\frac{\partial F}{\partial n_i} = -\frac{1}{n_i^2} L_i^2 S_i^2 + \lambda c_i = 0, \quad i = 1, 2, \dots, r,$$

Thus

$$n_i = (\lambda c_i)^{-1/2} L_i S_i, \quad i = 1, 2, \dots, r.$$

By substituting  $n_i$  in the equality constraint (7.85) we get

$$\sqrt{\lambda} = \frac{\sum_{i=1}^r \sqrt{c_i} L_i S_i}{d - c_0}.$$

Therefore,

$$n_i = \frac{(d - c_0) N_i S_i}{\sqrt{c_i} \sum_{j=1}^r \sqrt{c_j} N_j S_j}, \quad i = 1, 2, \dots, r. \quad (7.86)$$

It is easy to verify (using the sufficient conditions in Section 7.8) that the values of  $n_1, n_2, \dots, n_r$  given by equation (7.86) minimize  $\text{Var}(\bar{y}_{st})$  under the constraint of equality (7.85). We conclude that  $\text{Var}(\bar{y}_{st})$  is minimized when  $n_i$  is proportional to  $(1/\sqrt{c_i})N_i S_i$  ( $i = 1, 2, \dots, r$ ). Consequently,  $n_i$  must be large if the corresponding stratum is large, if the cost of sampling per unit in that stratum is low, or if the variability within the stratum is large.

### FURTHER READING AND ANNOTATED BIBLIOGRAPHY

- Bickel, P. J., and K. A. Doksum (1977). *Mathematical Statistics*, Holden-Day, San Francisco. (Chap. 1 discusses distribution theory for transformation of random vectors.)
- Brownlee, K. A. (1965). *Statistical Theory and Methodology*, 2nd ed. Wiley, New York. (See Section 9.8 with regard to the Behrens–Fisher test.)
- Cochran, W. G. (1963). *Sampling Techniques*, 2nd ed. Wiley, New York. (This is a classic book on sampling theory as developed for use in sample surveys.)
- Corwin, L. J., and R. H. Szczarba (1982). *Multivariate Calculus*. Marcel Dekker, New York. (This is a useful book that provides an introduction to multivariable calculus. The topics covered include continuity, differentiation, multiple integrals, line and surface integrals, differential forms, and infinite series.)
- Fulks, W. (1978). *Advanced Calculus*, 3rd ed. Wiley, New York. (Chap. 8 discusses limits and continuity for a multivariable function; Chap. 10 covers the inverse function theorem; Chap. 11 discusses multiple integration.)
- Gillespie, R. P. (1954). *Partial Differentiation*. Oliver and Boyd, Edinburgh, Scotland. (This concise book provides a brief introduction to multivariable calculus. It covers partial differentiation, Taylor’s theorem, and maxima and minima of functions of several variables.)
- Kaplan, W. (1991). *Advanced Calculus*, 4th ed. Addison-Wesley, Redwood City, California. (Topics pertaining to multivariable calculus are treated in several chapters including Chaps. 2, 3, 4, 5, and 6.)
- Kaplan, W., and D. J. Lewis (1971). *Calculus and Linear Algebra*, Vol. II. Wiley, New York. (Chap. 12 gives a brief introduction to differential calculus of a multivariable function; Chap. 13 covers multiple integration.)
- Lindgren, B. W. (1976). *Statistical Theory*, 3rd ed. Macmillan, New York. (Multivariate transformations are discussed in Chap. 10.)
- Lowerre, J. M. (1983). “An integral of the bivariate normal and an application.” *Amer. Statist.*, **37**, 235–236.
- Rudin, W. (1964). *Principles of Mathematical Analysis*, 2nd ed. McGraw-Hill, New York. (Chap. 9 includes a study of multivariable functions.)

- Sagan, H. (1974). *Advanced Calculus*. Houghton Mifflin, Boston. (Chap. 9 covers differential calculus of a multivariable function; Chap. 10 deals with the inverse function and implicit function theorems; Chap. 11 discusses multiple integration.)
- Satterthwaite, F. E. (1946). "An approximate distribution of estimates of variance components." *Biometrics Bull.*, **2**, 110–114.
- Taylor, A. E., and W. R. Mann (1972). *Advanced Calculus*, 2nd ed. Wiley, New York. (This book contains several chapters on multivariable calculus with many helpful exercises.)
- Thibaudeau, Y., and G. P. H. Styan (1985). "Bounds for Chakrabarti's measure of imbalance in experimental design." In *Proceedings of the First International Tampere Seminar on Linear Statistical Models and Their Applications*, T. Pukkila and S. Puntanen, eds. University of Tampere, Tampere, Finland, pp. 323–347.
- Wen, L. (2001). "A counterexample for the two-dimensional density function." *Amer. Math. Monthly*, **108**, 367–368.

## EXERCISES

### In Mathematics

**7.1.** Let  $f(x_1, x_2)$  be a function defined on  $R^2$  as

$$f(x_1, x_2) = \begin{cases} \frac{|x_1|}{x_2^2} \exp\left(-\frac{|x_1|}{x_2^2}\right), & x_2 \neq 0, \\ 0, & x_2 = 0. \end{cases}$$

- (a) Show that  $f(x_1, x_2)$  has a limit equal to zero as  $\mathbf{x} = (x_1, x_2)' \rightarrow \mathbf{0}$  along any straight line through the origin.
- (b) Show that  $f(x_1, x_2)$  does not have a limit as  $\mathbf{x} \rightarrow \mathbf{0}$ .

**7.2.** Prove Lemma 7.3.1.

**7.3.** Prove Lemma 7.3.2.

**7.4.** Prove Lemma 7.3.3.

**7.5.** Consider the function

$$f(x_1, x_2) = \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2}, & (x_1, x_2) \neq (0, 0), \\ 0, & (x_1, x_2) = (0, 0). \end{cases}$$

- (a) Show that  $f(x_1, x_2)$  is not continuous at the origin.



(b) Show that the partial derivatives of  $f(x_1, x_2)$  with respect to  $x_1$  and  $x_2$  exist at the origin.

[Note: This exercise shows that a multivariable function does not have to be continuous at a point in order for its partial derivatives to exist at that point.]

**7.6.** The function  $f(x_1, x_2, \dots, x_k)$  is said to be homogeneous of degree  $n$  in  $x_1, x_2, \dots, x_k$  if for any nonzero scalar  $t$ ,

$$f(tx_1, tx_2, \dots, tx_k) = t^n f(x_1, x_2, \dots, x_k)$$

for all  $\mathbf{x} = (x_1, x_2, \dots, x_k)'$  in the domain of  $f$ . Show that if  $f(x_1, x_2, \dots, x_k)$  is homogeneous of degree  $n$ , then

$$\sum_{i=1}^k x_i \frac{\partial f}{\partial x_i} = n f$$

[Note: This result is known as Euler's theorem for homogeneous functions.]

**7.7.** Consider the function

$$f(x_1, x_2) = \begin{cases} \frac{x_1^2 x_2}{x_1^4 + x_2^2}, & (x_1, x_2) \neq (0, 0), \\ 0, & (x_1, x_2) = (0, 0). \end{cases}$$

(a) Is  $f$  continuous at the origin? Why or why not?

(b) Show that  $f$  has a directional derivative in every direction at the origin.

**7.8.** Let  $S$  be a surface defined by the equation  $f(\mathbf{x}) = c_0$ , where  $\mathbf{x} = (x_1, x_2, \dots, x_k)'$  and  $c_0$  is a constant. Let  $C$  denote a curve on  $S$  given by the equations  $x_1 = g_1(t)$ ,  $x_2 = g_2(t)$ ,  $\dots$ ,  $x_k = g_k(t)$ , where  $g_1, g_2, \dots, g_k$  are differentiable functions. Let  $s$  be the arc length of  $C$  measured from some fixed point in such a way that  $s$  increases with  $t$ . The curve can then be parameterized, using  $s$  instead of  $t$ , in the form  $x_1 = h_1(s)$ ,  $x_2 = h_2(s)$ ,  $\dots$ ,  $x_k = h_k(s)$ . Suppose that  $f$  has partial derivatives with respect to  $x_1, x_2, \dots, x_k$ .

Show that the directional derivative of  $f$  at a point  $\mathbf{x}$  on  $C$  in the direction of  $\mathbf{v}$ , where  $\mathbf{v}$  is a unit tangent vector to  $C$  at  $\mathbf{x}$  (in the direction of increasing  $s$ ), is equal to  $df/ds$ .

**7.9.** Use Taylor's expansion in a neighborhood of the origin to obtain a second-order approximation for each of the following functions:

- (a)  $f(x_1, x_2) = \exp(x_2 \sin x_1)$ .  
 (b)  $f(x_1, x_2, x_3) = \sin(e^{x_1} + x_2^2 + x_3^3)$ .  
 (c)  $f(x_1, x_2) = \cos(x_1 x_2)$ .

**7.10.** Suppose that  $f(x_1, x_2)$  and  $g(x_1, x_2)$  are continuously differentiable functions in a neighborhood of a point  $\mathbf{x}_0 = (x_{10}, x_{20})'$ . Consider the equation  $u_1 = f(x_1, x_2)$ . Suppose that  $\partial f / \partial x_1 \neq 0$  at  $\mathbf{x}_0$ .

- (a) Show that

$$\frac{\partial x_1}{\partial x_2} = - \frac{\partial f}{\partial x_2} \bigg/ \frac{\partial f}{\partial x_1}$$

in a neighborhood of  $\mathbf{x}_0$ .

- (b) Suppose that in a neighborhood of  $\mathbf{x}_0$ ,

$$\frac{\partial(f, g)}{\partial(x_1, x_2)} = 0.$$

Show that

$$\frac{\partial g}{\partial x_1} \frac{\partial x_1}{\partial x_2} + \frac{\partial g}{\partial x_2} = 0,$$

that is,  $g$  is actually independent of  $x_2$  in a neighborhood of  $\mathbf{x}_0$ .

- (c) Deduce from (b) that there exists a function  $\phi: D \rightarrow R$ , where  $D \subset R$  is a neighborhood of  $f(\mathbf{x}_0)$ , such that

$$g(x_1, x_2) = \phi[f(x_1, x_2)]$$

throughout a neighborhood of  $\mathbf{x}_0$ . In this case, the functions  $f$  and  $g$  are said to be functionally dependent.

- (d) Show that if  $f$  and  $g$  are functionally dependent, then

$$\frac{\partial(f, g)}{\partial(x_1, x_2)} = 0.$$

[Note: From (b), (c), and (d) we conclude that  $f$  and  $g$  are functionally dependent on a set  $\Delta \subset R^2$  if and only if  $\partial(f, g) / \partial(x_1, x_2) = 0$  in  $\Delta$ .]

**7.11.** Consider the equation

$$x_1 \frac{\partial u}{\partial x_1} + x_2 \frac{\partial u}{\partial x_2} + x_3 \frac{\partial u}{\partial x_3} = nu.$$

Let  $\xi_1 = x_1/x_3$ ,  $\xi_2 = x_2/x_3$ ,  $\xi_3 = x_3$ . Use this change of variables to show that the equation can be written as

$$\xi_3 \frac{\partial u}{\partial \xi_3} = nu.$$

Deduce that  $u$  is of the form

$$u = x_3^n F\left(\frac{x_1}{x_3}, \frac{x_2}{x_3}\right).$$

**7.12.** Let  $u_1$  and  $u_2$  be defined as

$$\begin{aligned} u_1 &= x_1(1 - x_2^2)^{1/2} + x_2(1 - x_1^2)^{1/2}, \\ u_2 &= (1 - x_1^2)^{1/2}(1 - x_2^2)^{1/2} - x_1x_2. \end{aligned}$$

Show that  $u_1$  and  $u_2$  are functionally dependent.

**7.13.** Let  $\mathbf{f}: R^3 \rightarrow R^3$  be defined as

$$\mathbf{u} = \mathbf{f}(\mathbf{x}), \quad \mathbf{x} = (x_1, x_2, x_3)', \quad \mathbf{u} = (u_1, u_2, u_3)',$$

where  $u_1 = x_1^3$ ,  $u_2 = x_2^3$ ,  $u_3 = x_3^3$ .

(a) Show that the Jacobian matrix of  $\mathbf{f}$  is not nonsingular in any subset  $D \subset R^3$  that contains points on any of the coordinate planes.

(b) Show that  $\mathbf{f}$  has a unique inverse everywhere in  $R^3$  including any subset  $D$  of the type described in (a).

[Note: This exercise shows that the nonvanishing of the Jacobian determinant in Theorem 7.6.1 (inverse function theorem) is a sufficient condition for the existence of an inverse function, but is not necessary.]

**7.14.** Consider the equations

$$g_1(x_1, x_2, y_1, y_2) = 0,$$

$$g_2(x_1, x_2, y_1, y_2) = 0,$$

where  $g_1$  and  $g_2$  are differentiable functions defined on a set  $D \subset R^4$ . Suppose that  $\partial(g_1, g_2)/\partial(x_1, x_2) \neq 0$  in  $D$ . Show that

$$\begin{aligned} \frac{\partial x_1}{\partial y_1} &= - \frac{\partial(g_1, g_2)}{\partial(y_1, x_2)} \bigg/ \frac{\partial(g_1, g_2)}{\partial(x_1, x_2)}, \\ \frac{\partial x_2}{\partial y_1} &= - \frac{\partial(g_1, g_2)}{\partial(x_1, y_1)} \bigg/ \frac{\partial(g_1, g_2)}{\partial(x_1, x_2)}. \end{aligned}$$

- 7.15.** Let  $f(x_1, x_2, x_3) = 0$ ,  $g(x_1, x_2, x_3) = 0$ , where  $f$  and  $g$  are differentiable functions defined on a set  $D \subset \mathbb{R}^3$ . Suppose that

$$\frac{\partial(f, g)}{\partial(x_2, x_3)} \neq 0, \quad \frac{\partial(f, g)}{\partial(x_3, x_1)} \neq 0, \quad \frac{\partial(f, g)}{\partial(x_1, x_2)} \neq 0$$

in  $D$ . Show that

$$\frac{dx_1}{\partial(f, g)/\partial(x_2, x_3)} = \frac{dx_2}{\partial(f, g)/\partial(x_3, x_1)} = \frac{dx_3}{\partial(f, g)/\partial(x_1, x_2)}.$$

- 7.16.** Determine the stationary points of the following functions and check for local minima and maxima:

- (a)  $f = x_1^2 + x_2^2 + x_1 + x_2 + x_1x_2$ .  
 (b)  $f = 2\alpha x_1^2 - x_1x_2 + x_2^2 + x_1 - x_2 + 1$ , where  $\alpha$  is a scalar. Can  $\alpha$  be chosen so that the stationary point is (i) a point of local minimum; (ii) a point of local maximum; (iii) a saddle point?  
 (c)  $f = x_1^3 - 6x_1x_2 + 3x_2^2 - 24x_1 + 4$ .  
 (d)  $f = x_1^4 + x_2^4 - 2(x_1 - x_2)^2$ .

- 7.17.** Consider the function

$$f = \frac{1 + p + \sum_{i=1}^m p_i}{(1 + p^2 + \sum_{i=1}^m p_i^2)^{1/2}},$$

which is defined on the region

$$C = \{(p_1, p_2, \dots, p_m) | 0 < p \leq p_i \leq 1, i = 1, 2, \dots, m\},$$

where  $p$  is a known constant. Show that

- (a)  $\partial f / \partial p_i$ , for  $i = 1, 2, \dots, m$ , vanish at exactly one point in  $C$ .  
 (b) The gradient vector  $\nabla f = (\partial f / \partial p_1, \partial f / \partial p_2, \dots, \partial f / \partial p_m)'$  does not vanish anywhere on the boundary of  $C$ .  
 (c)  $f$  attains its absolute maximum in the interior of  $C$  at the point  $(p_1^o, p_2^o, \dots, p_m^o)$ , where

$$p_i^o = \frac{1 + p^2}{1 + p}, \quad i = 1, 2, \dots, m.$$

[Note: The function  $f$  was considered in an article by Thibaudeau and Styan (1985) concerning a measure of imbalance for experimental designs.]

- 7.18.** Show that the function  $f = (x_2 - x_1^2)(x_2 - 2x_1^2)$  does not have a local maximum or minimum at the origin, although it has a local minimum for  $t = 0$  along every straight line given by the equations  $x_1 = at$ ,  $x_2 = bt$ , where  $a$  and  $b$  are constants.

**7.19.** Find the optimal values of the function  $f = x_1^2 + 12x_1x_2 + 2x_2^2$  subject to  $4x_1^2 + x_2^2 = 25$ . Determine the nature of the optima.

**7.20.** Find the minimum distance from the origin to the curve of intersection of the surfaces,  $x_3(x_1 + x_2) = -2$  and  $x_1x_2 = 1$ .

**7.21.** Apply the method of Lagrange multipliers to show that

$$(x_1^2x_2^2x_3^2)^{1/3} \leq \frac{1}{3}(x_1^2 + x_2^2 + x_3^2)$$

for all values of  $x_1, x_2, x_3$ .

[Hint: Find the maximum value of  $f = x_1^2x_2^2x_3^2$  subject to  $x_1^2 + x_2^2 + x_3^2 = c^2$ , where  $c$  is a constant.]

**7.22.** Prove Theorem 7.9.3.

**7.23.** Evaluate the following integrals:

(a)  $\int \int_D x_2 \sqrt{x_1} dx_1 dx_2$ , where

$$D = \{(x_1, x_2) | x_1 > 0, x_2 > x_1^2, x_2 < 2 - x_1^2\}.$$

(b)  $\int_0^1 [\int_0^{\sqrt{1-x_2^2}} (\frac{2}{3}x_1 + \frac{4}{3}x_2) dx_1] dx_2$ .

**7.24.** Show that if  $f(x_1, x_2)$  is continuous, then

$$\int_0^2 \left[ \int_{x_1^2}^{4x_1 - x_1^2} f(x_1, x_2) dx_2 \right] dx_1 = \int_0^4 \left[ \int_{2 - \sqrt{4-x_2}}^{\sqrt{x_2}} f(x_1, x_2) dx_1 \right] dx_2.$$

**7.25.** Consider the integral

$$I = \int_0^1 \left[ \int_{1-x_1}^{1-x_1^2} f(x_1, x_2) dx_2 \right] dx_1.$$

(a) Write an equivalent expression for  $I$  by reversing the order of integration.

(b) If  $g(x_1) = \int_{1-x_1}^{1-x_1^2} f(x_1, x_2) dx_2$ , find  $dg/dx_1$ .

**7.26.** Evaluate  $\int \int_D x_1x_2 dx_1 dx_2$ , where  $D$  is a region enclosed by the four parabolas  $x_2^2 = x_1$ ,  $x_2^2 = 2x_1$ ,  $x_1^2 = x_2$ ,  $x_1^2 = 2x_2$ .

[Hint: Use a proper change of variables.]

**7.27.** Evaluate  $\iiint_D (x_1^2 + x_2^2) dx_1 dx_2 dx_3$ , where  $D$  is a sphere of radius 1 centered at the origin.

[Hint: Make a change of variables using spherical polar coordinates of the form

$$x_1 = r \sin \theta \cos \phi,$$

$$x_2 = r \sin \theta \sin \phi,$$

$$x_3 = r \cos \theta,$$

$$0 \leq r \leq 1, 0 \leq \theta \leq \pi, 0 \leq \phi \leq 2\pi.]$$

**7.28.** Find the value of the integral

$$I = \int_0^{\sqrt{3}} \frac{dx}{(1+x^2)^3}.$$

[Hint: Consider the integral  $\int_0^{\sqrt{3}} dx/(a+x^2)$ , where  $a > 0$ .]

### In Statistics

**7.29.** Suppose that the random vector  $\mathbf{X} = (X_1, X_2)'$  has the density function

$$f(x_1, x_2) = \begin{cases} x_1 + x_2, & 0 < x_1 < 1, 0 < x_2 < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

(a) Are the random variables  $X_1$  and  $X_2$  independent?

(b) Find the expected value of  $X_1 X_2$ .

**7.30.** Consider the density function  $f(x_1, x_2)$  of  $\mathbf{X} = (X_1, X_2)'$ , where

$$f(x_1, x_2) = \begin{cases} 1, & -x_2 < x_1 < x_2, 0 < x_2 < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Show that  $X_1$  and  $X_2$  are uncorrelated random variables [that is,  $E(X_1 X_2) = E(X_1)E(X_2)$ ], but are not independent.

**7.31.** The density function of  $\mathbf{X} = (X_1, X_2)'$  is given by

$$f(x_1, x_2) = \begin{cases} \frac{1}{\Gamma(\alpha)\Gamma(\beta)} x_1^{\alpha-1} x_2^{\beta-1} e^{-x_1-x_2}, & 0 < x_1, x_2 < \infty, \\ 0 & \text{elsewhere.} \end{cases}$$

where  $\alpha > 0$ ,  $\beta > 0$ , and  $\Gamma(m)$  is the gamma function  $\Gamma(m) = \int_0^\infty e^{-x} x^{m-1} dx$ ,  $m > 0$ . Suppose that  $Y_1$  and  $Y_2$  are random variables defined as

$$Y_1 = \frac{X_1}{X_1 + X_2},$$

$$Y_2 = X_1 + X_2.$$

- (a) Find the joint density function of  $Y_1$  and  $Y_2$ .
- (b) Find the marginal densities of  $Y_1$  and  $Y_2$ .
- (c) Are  $Y_1$  and  $Y_2$  independent?

**7.32.** Suppose that  $\mathbf{X} = (X_1, X_2)'$  has the density function

$$f(x_1, x_2) = \begin{cases} 10x_1x_2^2, & 0 < x_1 < x_2, 0 < x_2 < 1, \\ 0 & \text{elsewhere.} \end{cases}$$

Find the density function of  $W = X_1X_2$ .

**7.33.** Find the density function of  $W = (X_1^2 + X_2^2)^{1/2}$  given that  $\mathbf{X} = (X_1, X_2)'$  has the density function

$$f(x_1, x_2) = \begin{cases} 4x_1x_2 e^{-x_1^2 - x_2^2}, & x_1 > 0, x_2 > 0, \\ 0 & \text{elsewhere.} \end{cases}$$

**7.34.** Let  $X_1, X_2, \dots, X_n$  be independent random variables that have the exponential density  $f(x) = e^{-x}$ ,  $x > 0$ . Let  $Y_1, Y_2, \dots, Y_n$  be  $n$  random variables defined as

$$\begin{aligned} Y_1 &= X_1, \\ Y_2 &= X_1 + X_2, \\ &\vdots \\ Y_n &= X_1 + X_2 + \dots + X_n. \end{aligned}$$

Find the density of  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$ , and then deduce the marginal density of  $Y_n$ .

**7.35.** Prove formula (7.84).

**7.36.** Let  $X_1$  and  $X_2$  be independent random variables such that  $W_1 = (6/\sigma_1^2)X_1$  and  $W_2 = (8/\sigma_2^2)X_2$  have the chi-squared distribution with

six and eight degrees of freedom, respectively, where  $\sigma_1^2$  and  $\sigma_2^2$  are unknown parameters. Let  $\theta = \frac{1}{7}\sigma_1^2 + \frac{1}{9}\sigma_2^2$ . An unbiased estimator of  $\theta$  is given by  $\hat{\theta} = \frac{1}{7}X_1 + \frac{1}{9}X_2$ , since  $X_1$  and  $X_2$  are unbiased estimators of  $\sigma_1^2$  and  $\sigma_2^2$ , respectively.

Using Satterthwaite's approximation (see Satterthwaite, 1946), it can be shown that  $\eta\hat{\theta}/\theta$  is approximately distributed as a chi-squared variate with  $\eta$  degrees of freedom, where  $\eta$  is given by

$$\eta = \frac{\left(\frac{1}{7}\sigma_1^2 + \frac{1}{9}\sigma_2^2\right)^2}{\frac{1}{6}\left(\frac{1}{7}\sigma_1^2\right)^2 + \frac{1}{8}\left(\frac{1}{9}\sigma_2^2\right)^2},$$

which can be written as

$$\eta = \frac{8(9 + 7\lambda)^2}{108 + 49\lambda^2},$$

where  $\lambda = \sigma_2^2/\sigma_1^2$ . It follows that the probability

$$p = P\left(\frac{\eta\hat{\theta}}{\chi_{0.025, \eta}^2} < \theta < \frac{\eta\hat{\theta}}{\chi_{0.975, \eta}^2}\right),$$

where  $\chi_{\alpha, \eta}^2$  denotes the upper  $100\alpha\%$  point of the chi-squared distribution with  $\eta$  degrees of freedom, is approximately equal to 0.95. Compute the exact value of  $p$  using double integration, given that  $\lambda = 2$ . Compare the result with the 0.95 value.

[Notes: (1) The density function of a chi-squared random variable with  $n$  degrees of freedom is given in Example 6.9.6. (2) In general,  $\eta$  is unknown. It can be estimated by  $\hat{\eta}$  which results from replacing  $\lambda$  with  $\hat{\lambda} = X_2/X_1$  in the formula for  $\eta$ . (3) The estimator  $\hat{\theta}$  is used in the Behrens–Fisher test statistic for comparing the means of two populations with unknown variances  $\sigma_1^2$  and  $\sigma_2^2$ , which are assumed to be unequal. If  $\bar{Y}_1$  and  $\bar{Y}_2$  are the means of two independent samples of sizes  $n_1 = 7$  and  $n_2 = 9$ , respectively, randomly chosen from these populations, then  $\theta$  is the variance of  $\bar{Y}_1 - \bar{Y}_2$ . In this case,  $X_1$  and  $X_2$  represent the corresponding sample variances. The Behrens–Fisher  $t$ -statistic is then given by

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\hat{\theta}}}.$$

If the two population means are equal,  $t$  has approximately the  $t$ -distribution with  $\eta$  degrees of freedom. For more details about the Behrens–Fisher test, see for example, Brownlee (1965, Section 9.8).]



- 7.37.** Suppose that a parabola of the form  $\mu = \beta_0 + \beta_1 x + \beta_2 x^2$  is fitted to a set of paired data,  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ . Obtain estimates of  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$  by minimizing  $\sum_{i=1}^n [y_i - (\beta_0 + \beta_1 x_i + \beta_2 x_i^2)]^2$  with respect to  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ .

[Note: The estimates obtained in this manner are the least-squares estimates of  $\beta_0$ ,  $\beta_1$ , and  $\beta_2$ .]

- 7.38.** Suppose that we have  $k$  disjoint events  $A_1, A_2, \dots, A_k$  such that the probability of  $A_i$  is  $p_i$  ( $i = 1, 2, \dots, k$ ) and  $\sum_{i=1}^k p_i = 1$ . Furthermore, suppose that among  $n$  independent trials there are  $X_1, X_2, \dots, X_k$  outcomes associated with  $A_1, A_2, \dots, A_k$ , respectively. The joint probability that  $X_1 = x_1, X_2 = x_2, \dots, X_k = x_k$  is given by the likelihood function

$$L(\mathbf{x}, \mathbf{p}) = \frac{n!}{x_1! x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k},$$

where  $x_i = 0, 1, 2, \dots, n$  for  $i = 1, 2, \dots, k$  such that  $\sum_{i=1}^k x_i = n$ ,  $\mathbf{x} = (x_1, x_2, \dots, x_k)'$ ,  $\mathbf{p} = (p_1, p_2, \dots, p_k)'$ . This defines a joint distribution for  $X_1, X_2, \dots, X_k$  known as the multinomial distribution.

Find the maximum likelihood estimates of  $p_1, p_2, \dots, p_k$  by maximizing  $L(\mathbf{x}, \mathbf{p})$  subject to  $\sum_{i=1}^k p_i = 1$ .

[Hint: Maximize the natural logarithm of  $p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$  subject to  $\sum_{i=1}^k p_i = 1$ .]

- 7.39.** Let  $\phi(y)$  be a positive, even, and continuous function on  $(-\infty, \infty)$  such that  $\phi(y)$  is strictly decreasing on  $(0, \infty)$ , and  $\int_{-\infty}^{\infty} \phi(y) dy = 1$ . Consider the following bivariate density function:

$$f(x, y) = \begin{cases} 1 + x/\phi(y), & -\phi(y) \leq x < 0, \\ 1 - x/\phi(y), & 0 \leq x \leq \phi(y), \\ 0 & \text{otherwise.} \end{cases}$$

- (a) Show that  $f(x, y)$  is continuous for  $-\infty < x, y < \infty$ .  
 (b) Let  $F(x, y)$  be the corresponding cumulative distribution function,

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(s, t) ds dt.$$

Show that if  $0 < \Delta x < \phi(0)$ , then

$$\begin{aligned} F(\Delta x, 0) - F(0, 0) &\geq \int_0^{\phi^{-1}(\Delta x)} \int_0^{\Delta x} \left[ 1 - \frac{s}{\phi(t)} \right] ds dt \\ &\geq \frac{1}{2} \Delta x \phi^{-1}(\Delta x), \end{aligned}$$

where  $\phi^{-1}$  is the inverse function of  $\phi(y)$  for  $0 \leq y < \infty$ .

(c) Use part (b) to show that

$$\lim_{\Delta x \rightarrow 0^+} \frac{F(\Delta x, 0) - F(0, 0)}{\Delta x} = \infty.$$

Hence,  $\partial F(x, y)/\partial x$  does not exist at  $(0, 0)$ .

(d) Deduce from part (c) that the equality

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y}$$

does not hold in this example.

[*Note:* This example was given by Wen (2001) to demonstrate that continuity of  $f(x, y)$  is not sufficient for the existence of  $\partial F/\partial x$ , and hence for the validity of the equality in part (d).]