MULTIVARIABLE AND VECTOR ANALYSIS

W W L CHEN

Chapter 4

HIGHER ORDER DERIVATIVES

4.1. Iterated Partial Derivatives

In this chapter, we shall be concerned with functions of the type $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$. We shall consider iterated partial derivatives of the form

$$\frac{\partial^2 f}{\partial x_i^2} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_i}\right) \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right),$$

where $i, j = 1, \ldots, n$. An immediate question that arises is whether

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

when $i \neq j$.

Example 4.1.1. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by

$$f(x, y) = \begin{cases} \dfrac{xy(x^2 - y^2)}{x^2 + y^2} & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{if } (x, y) = (0, 0). \end{cases}$$

It is easily seen that

$$\frac{\partial f}{\partial x} = \frac{x^4 y + 4x^2 y^3 - y^5}{(x^2 + y^2)^2} \qquad \text{and} \qquad \frac{\partial f}{\partial y} = \frac{x^5 - 4x^3 y^2 - xy^4}{(x^2 + y^2)^2}$$

whenever $(x, y) \neq (0, 0)$. Furthermore,

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{x \to 0} \frac{f(x, 0) - f(0, 0)}{x - 0} = 0 \qquad \text{and} \qquad \frac{\partial f}{\partial y}(0, 0) = \lim_{y \to 0} \frac{f(0, y) - f(0, 0)}{y - 0} = 0.$$

Note, however, that

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) = \lim_{x\to 0} \frac{\dfrac{\partial f}{\partial y}(x,0) - \dfrac{\partial f}{\partial y}(0,0)}{x-0} = \lim_{x\to 0} \frac{x-0}{x-0} = 1,$$

while

$$\frac{\partial^2 f}{\partial y \partial x}(0,0) = \lim_{y\to 0} \frac{\dfrac{\partial f}{\partial x}(0,y) - \dfrac{\partial f}{\partial x}(0,0)}{y-0} = \lim_{y\to 0} \frac{-y-0}{y-0} = -1,$$

so that

$$\frac{\partial^2 f}{\partial x \partial y}(0,0) \neq \frac{\partial^2 f}{\partial y \partial x}(0,0).$$

It can further be checked that

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = \frac{x^6 + 9x^4y^2 - 9x^2y^4 - y^6}{(x^2+y^2)^3}$$

whenever $(x,y) \neq (0,0)$. Clearly at least one of these two iterated second partial derivatives

$$\frac{\partial^2 f}{\partial x \partial y} \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y \partial x}$$

is not continuous at $(0,0)$.

THEOREM 4A. *Suppose that the function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^2$ is an open set, has continuous iterated second partial derivatives. Then*

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$$

holds everywhere in A.

† PROOF. Suppose that $(x_0, y_0) \in A$ is chosen. Since A is open, there exists an open disc $D(x_0, y_0, r) \subseteq A$. For every $(x,y) \in D(x_0, y_0, r)$, consider the expression

$$S(x,y) = f(x,y) - f(x,y_0) - f(x_0,y) + f(x_0,y_0).$$

For every fixed y, write

$$g_y(x) = f(x,y) - f(x,y_0),$$

so that

$$S(x,y) = g_y(x) - g_y(x_0).$$

By the Mean value theorem on g_y, there exists $\widetilde{x}$ between x_0 and x such that

$$g_y(x) - g_y(x_0) = (x-x_0)\frac{\partial g_y}{\partial x}(\widetilde{x}) = (x-x_0)\left(\frac{\partial f}{\partial x}(\widetilde{x},y) - \frac{\partial f}{\partial x}(\widetilde{x},y_0)\right).$$

By the Mean value theorem on $\partial f/\partial x$, there exists $\widetilde{y}$ between y_0 and y such that

$$\frac{\partial f}{\partial x}(\widetilde{x},y) - \frac{\partial f}{\partial x}(\widetilde{x},y_0) = (y-y_0)\frac{\partial^2 f}{\partial y \partial x}(\widetilde{x},\widetilde{y}).$$

Hence

$$\frac{\partial^2 f}{\partial y \partial x}(\widetilde{x}, \widetilde{y}) = \frac{S(x,y)}{(x-x_0)(y-y_0)}.$$

Since $\partial^2 f/\partial y \partial x$ is continuous at (x_0, y_0), and since $(\widetilde{x}, \widetilde{y}) \to (x_0, y_0)$ as $(x,y) \to (x_0, y_0)$, we must have

$$\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) = \lim_{(x,y)\to(x_0,y_0)} \frac{S(x,y)}{(x-x_0)(y-y_0)}.$$

A similar argument with the roles of the two variables x and y reversed gives

$$\frac{\partial^2 f}{\partial x \partial y}(x_0, y_0) = \lim_{(x,y)\to(x_0,y_0)} \frac{S(x,y)}{(x-x_0)(y-y_0)}.$$

Hence

$$\frac{\partial^2 f}{\partial y \partial x}(x_0, y_0) = \frac{\partial^2 f}{\partial x \partial y}(x_0, y_0)$$

as required. ○

4.2. Taylor's Theorem

Recall that in the theory of real valued functions of one real variable, Taylor's theorem states that for a smooth function,

$$f(x) = f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2!}(x-x_0)^2 + \ldots + \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + R_k(x), \tag{1}$$

where the remainder term

$$R_k(x) = \int_{x_0}^{x} \frac{(x-t)^k}{k!} f^{(k+1)}(t)\,\mathrm{d}t \tag{2}$$

satisfies

$$\lim_{x\to x_0} \frac{R_k(x)}{(x-x_0)^k} = 0.$$

Remark. We usually prove this result by first using the Fundamental theorem of integral calculus to obtain

$$f(x) - f(x_0) = \int_{x_0}^{x} f'(t)\,\mathrm{d}t.$$

Integrating by parts, we obtain

$$\int_{x_0}^{x} f'(t)\,\mathrm{d}t = \Big[(t-x)f'(t)\Big]_{x_0}^{x} - \int_{x_0}^{x} (t-x)f''(t)\,\mathrm{d}t = f'(x_0)(x-x_0) + \int_{x_0}^{x} (x-t)f''(t)\,\mathrm{d}t.$$

Hence

$$f(x) - f(x_0) = f'(x_0)(x-x_0) + \int_{x_0}^{x} (x-t)f''(t)\,\mathrm{d}t,$$

proving (1) and (2) for $k = 1$. The proof is now completed by induction on k and using integrating by parts on the integral (2).

Our goal in this section is to obtain Taylor approximations for functions of the type $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set. Suppose first of all that $\mathbf{x}_0 \in A$, and that f is differentiable at $\mathbf{x}_0$. For any $\mathbf{x} \in A$, let

$$R_1(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}_0) - (\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0),$$

where $(\mathbf{D}f)(\mathbf{x}_0)$ denotes the total derivative of f at $\mathbf{x}_0$, and where $\mathbf{x} - \mathbf{x}_0$ is interpreted as a column matrix. Since f is differentiable at $\mathbf{x}_0$, we have

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{|f(\mathbf{x}) - f(\mathbf{x}_0) - (\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0;$$

in other words,

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{|R_1(\mathbf{x})|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

Note that

$$(\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}(\mathbf{x}_0) \right) (x_i - X_i), \tag{3}$$

where $\mathbf{x}_0 = (X_1, \ldots, X_n)$.

We have therefore proved the following result on first-order Taylor approximations.

THEOREM 4B. *Suppose that the function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set, is differentiable at $\mathbf{x}_0 \in A$. Then for every $\mathbf{x} \in A$, we have*

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}(\mathbf{x}_0) \right) (x_i - X_i) + R_1(\mathbf{x}),$$

where $\mathbf{x}_0 = (X_1, \ldots, X_n)$, and where

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{|R_1(\mathbf{x})|}{\|\mathbf{x} - \mathbf{x}_0\|} = 0.$$

For second-order Taylor approximation, we have the following result.

THEOREM 4C. *Suppose that the function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set, has continuous iterated second partial derivatives. Suppose further that $\mathbf{x}_0 \in A$. Then for every $\mathbf{x} \in A$, we have*

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i}(\mathbf{x}_0) \right) (x_i - X_i) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0) \right) (x_i - X_i)(x_j - X_j) + R_2(\mathbf{x}), \tag{4}$$

where $\mathbf{x}_0 = (X_1, \ldots, X_n)$, and where

$$\lim_{\mathbf{x}\to\mathbf{x}_0} \frac{|R_2(\mathbf{x})|}{\|\mathbf{x} - \mathbf{x}_0\|^2} = 0. \tag{5}$$

† Sketch of Proof. We shall attempt to demonstrate the theorem by making the extra assumption that f has continuous iterated third partial derivatives. Consider the function

$$L : [0,1] \to \mathbb{R}^n : t \mapsto (1-t)\mathbf{x}_0 + t\mathbf{x};$$

here L denotes the line segment joining $\mathbf{x}_0$ and $\mathbf{x}$, and we shall make the extra assumption that this line segment lies in A. Then consider the composition $g = f \circ L : [0,1] \to \mathbb{R}$, where $g(t) = f((1-t)\mathbf{x}_0 + t\mathbf{x})$ for every $t \in [0,1]$. We now apply (1) and (2) to the function g to obtain

$$g(1) = g(0) + g'(0) + \frac{g''(0)}{2} + R_2,$$

where

$$R_2 = \int_0^1 \frac{(t-1)^2}{2} g'''(t)\,\mathrm{d}t.$$

Applying the Chain rule, we have

$$\begin{aligned} g'(t) = (\mathbf{D}f)(L(t))(\mathbf{D}L)(t) &= \begin{pmatrix} \dfrac{\partial f}{\partial x_1}(L(t)) & \dots & \dfrac{\partial f}{\partial x_n}(L(t)) \end{pmatrix} \begin{pmatrix} x_1 - X_1 \\ \vdots \\ x_n - X_n \end{pmatrix} \\ &= \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(L(t)) \right) (x_i - X_i), \end{aligned}$$

so that

$$g'(0) = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(L(0)) \right) (x_i - X_i) = \sum_{i=1}^n \left(\frac{\partial f}{\partial x_i}(\mathbf{x}_0) \right) (x_i - X_i).$$

Note that

$$g'(t) = \sum_{j=1}^n \left(\left(\frac{\partial f}{\partial x_j} \circ L \right)(t) \right) (x_j - X_j).$$

It follows from the Chain rule and the arithmetic of derivatives that

$$\begin{aligned} g''(t) &= \sum_{j=1}^n \left(\left(\mathbf{D}\frac{\partial f}{\partial x_j} \right)(L(t))(\mathbf{D}L)(t) \right) (x_j - X_j) \\ &= \sum_{j=1}^n \left(\begin{pmatrix} \dfrac{\partial^2 f}{\partial x_1 \partial x_j}(L(t)) & \dots & \dfrac{\partial^2 f}{\partial x_n \partial x_j}(L(t)) \end{pmatrix} \begin{pmatrix} x_1 - X_1 \\ \vdots \\ x_n - X_n \end{pmatrix} \right) (x_j - X_j) \\ &= \sum_{j=1}^n \left(\sum_{i=1}^n \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(L(t)) \right) (x_i - X_i) \right) (x_j - X_j) \\ &= \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(L(t)) \right) (x_i - X_i)(x_j - X_j), \end{aligned}$$

so that

$$g''(0) = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(L(0)) \right) (x_i - X_i)(x_j - X_j) = \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0) \right) (x_i - X_i)(x_j - X_j).$$

Note that

$$g''(t) = \sum_{j=1}^n \sum_{k=1}^n \left(\left(\frac{\partial^2 f}{\partial x_j \partial x_k} \circ L \right)(t) \right) (x_j - X_j)(x_k - X_k).$$

It can be shown, using the Chain rule and the arithmetic of derivatives, that

$$\begin{aligned} g'''(t) &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \left(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(L(t)) \right) (x_i - X_i)(x_j - X_j)(x_k - X_k) \\ &= \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \left(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}((1-t)\mathbf{x}_0 + t\mathbf{x}) \right) (x_i - X_i)(x_j - X_j)(x_k - X_k). \end{aligned}$$

Writing $R_2 = R_2(\mathbf{x})$, we have established (4), where

$$R_2(\mathbf{x}) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \int_0^1 \frac{(t-1)^2}{2} \left(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}((1-t)\mathbf{x}_0 + t\mathbf{x}) \right) (x_i - X_i)(x_j - X_j)(x_k - X_k)\,\mathrm{d}t.$$

The function

$$\frac{(t-1)^2}{2} \left(\frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}((1-t)\mathbf{x}_0 + t\mathbf{x}) \right)$$

is continuous, and hence bounded by M, say, in $[0,1]$. Also

$$|x_i - X_i|, |x_j - X_j|, |x_k - X_k| \le \|\mathbf{x} - \mathbf{x}_0\|,$$

so $|R_2(\mathbf{x})| \le n^3 M \|\mathbf{x} - \mathbf{x}_0\|^3$, and so (5) follows. ◯

The second-order term that arises in Theorem 4C is of particular importance in the determination of the nature of stationary points later.

Definition. The quadratic function

$$\mathbf{H}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0) \right) (x_i - X_i)(x_j - X_j) \tag{6}$$

is called the Hessian of f at $\mathbf{x}_0$.

Remark. The expression (4) can be rewritten in the form

$$f(\mathbf{x}) = f(\mathbf{x}_0) + (\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + \mathbf{H}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + R_2(\mathbf{x}), \tag{7}$$

where $(\mathbf{D}f)(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$, the matrix product of the total derivative $(\mathbf{D}f)(\mathbf{x}_0)$ with the column matrix $\mathbf{x} - \mathbf{x}_0$, is given by (3), and where the Hessian $\mathbf{H}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$ is given by (6).

Example 4.2.1. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x,y) = x^2y + 3y - 2$ for every $(x,y) \in \mathbb{R}^2$, near the point $(x_0, y_0) = (1, -2)$. Clearly $f(1,-2) = -10$. We have

$$\frac{\partial f}{\partial x} = 2xy \qquad \text{and} \qquad \frac{\partial f}{\partial y} = x^2 + 3.$$

Also

$$\frac{\partial^2 f}{\partial x^2} = 2y \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2} = 0 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = 2x.$$

Hence

$$\frac{\partial f}{\partial x}(1,-2) = -4 \qquad \text{and} \qquad \frac{\partial f}{\partial y}(1,-2) = 4.$$

Also

$$\frac{\partial^2 f}{\partial x^2}(1,-2) = -4 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2}(1,-2) = 0 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x \partial y}(1,-2) = \frac{\partial^2 f}{\partial y \partial x}(1,-2) = 2.$$

Since

$$\begin{aligned} f(x,y) &= f(1,-2) + \left(\left(\frac{\partial f}{\partial x}(1,-2) \right)(x-1) + \left(\frac{\partial f}{\partial y}(1,-2) \right)(y+2) \right) \\ &\quad + \frac{1}{2} \left(\left(\frac{\partial^2 f}{\partial x^2}(1,-2) \right)(x-1)^2 + \left(\frac{\partial^2 f}{\partial x \partial y}(1,-2) \right)(x-1)(y+2) \right. \\ &\qquad \left. + \left(\frac{\partial^2 f}{\partial y \partial x}(1,-2) \right)(x-1)(y+2) + \left(\frac{\partial^2 f}{\partial y^2}(1,-2) \right)(y+2)^2 \right) + R_2(x,y) \\ &= -10 - 4(x-1) + 4(y+2) - 2(x-1)^2 + 2(x-1)(y+2) + R_2(x,y), \end{aligned}$$

it follows that the second-order Taylor approximation of f at $(1,-2)$ is given by

$$-10 - 4(x-1) + 4(y+2) - 2(x-1)^2 + 2(x-1)(y+2),$$

and the Hessian of f at $(1,-2)$ is given by

$$-2(x-1)^2 + 2(x-1)(y+2).$$

Example 4.2.2. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x,y) = \mathrm{e}^x \cos y$ for every $(x,y) \in \mathbb{R}^2$, near the point $(x_0, y_0) = (0,0)$. Clearly $f(0,0) = 1$. We have

$$\frac{\partial f}{\partial x} = \mathrm{e}^x \cos y \qquad \text{and} \qquad \frac{\partial f}{\partial y} = -\mathrm{e}^x \sin y.$$

Also

$$\frac{\partial^2 f}{\partial x^2} = \mathrm{e}^x \cos y \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2} = -\mathrm{e}^x \cos y \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = -\mathrm{e}^x \sin y.$$

Hence

$$\frac{\partial f}{\partial x}(0,0) = 1 \qquad \text{and} \qquad \frac{\partial f}{\partial y}(0,0) = 0.$$

Also

$$\frac{\partial^2 f}{\partial x^2}(0,0) = 1 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2}(0,0) = -1 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x \partial y}(0,0) = \frac{\partial^2 f}{\partial y \partial x}(0,0) = 0.$$

Since

$$\begin{aligned} f(x,y) &= f(0,0) + \left(\left(\frac{\partial f}{\partial x}(0,0) \right)(x-0) + \left(\frac{\partial f}{\partial y}(0,0) \right)(y-0) \right) \\ &\quad + \frac{1}{2} \left(\left(\frac{\partial^2 f}{\partial x^2}(0,0) \right)(x-0)^2 + \left(\frac{\partial^2 f}{\partial x \partial y}(0,0) \right)(x-0)(y-0) \right. \\ &\qquad \left. + \left(\frac{\partial^2 f}{\partial y \partial x}(0,0) \right)(x-0)(y-0) + \left(\frac{\partial^2 f}{\partial y^2}(0,0) \right)(y-0)^2 \right) + R_2(x,y) \\ &= 1 + x + \frac{1}{2}x^2 - \frac{1}{2}y^2 + R_2(x,y), \end{aligned}$$

it follows that the second-order Taylor approximation of f at $(0,0)$ is given by

$$1+x+\frac{1}{2}x^2-\frac{1}{2}y^2,$$

and the Hessian of f at $(0,0)$ is given by

$$\frac{1}{2}x^2-\frac{1}{2}y^2.$$

EXAMPLE 4.2.3. Consider the function $f:\mathbb{R}^3\to\mathbb{R}$, defined by $f(x,y,z)=x^2y+xz^3+y^2z^2$ for every $(x,y,z)\in\mathbb{R}^3$, near the point $(x_0,y_0,z_0)=(1,1,1)$. Clearly $f(1,1,1)=3$. We have

$$\frac{\partial f}{\partial x}=2xy+z^3 \qquad\text{and}\qquad \frac{\partial f}{\partial y}=x^2+2yz^2 \qquad\text{and}\qquad \frac{\partial f}{\partial z}=3xz^2+2y^2z.$$

Also

$$\frac{\partial^2 f}{\partial x^2}=2y \qquad\text{and}\qquad \frac{\partial^2 f}{\partial y^2}=2z^2 \qquad\text{and}\qquad \frac{\partial^2 f}{\partial z^2}=6xz+2y^2.$$

Furthermore,

$$\frac{\partial^2 f}{\partial x\partial y}=\frac{\partial^2 f}{\partial y\partial x}=2x \qquad\text{and}\qquad \frac{\partial^2 f}{\partial x\partial z}=\frac{\partial^2 f}{\partial z\partial x}=3z^2 \qquad\text{and}\qquad \frac{\partial^2 f}{\partial y\partial z}=\frac{\partial^2 f}{\partial z\partial y}=4yz.$$

Hence

$$\frac{\partial f}{\partial x}(1,1,1)=3 \qquad\text{and}\qquad \frac{\partial f}{\partial y}(1,1,1)=3 \qquad\text{and}\qquad \frac{\partial f}{\partial z}(1,1,1)=5.$$

Also

$$\frac{\partial^2 f}{\partial x^2}(1,1,1)=2 \qquad\text{and}\qquad \frac{\partial^2 f}{\partial y^2}(1,1,1)=2 \qquad\text{and}\qquad \frac{\partial^2 f}{\partial z^2}(1,1,1)=8.$$

Furthermore,

$$\frac{\partial^2 f}{\partial x\partial y}(1,1,1)=2 \qquad\text{and}\qquad \frac{\partial^2 f}{\partial x\partial z}(1,1,1)=3 \qquad\text{and}\qquad \frac{\partial^2 f}{\partial y\partial z}(1,1,1)=4.$$

Since

$$\begin{aligned}
f(x,y,z)&=f(1,1,1)+\left(\left(\frac{\partial f}{\partial x}(1,1,1)\right)(x-1)+\left(\frac{\partial f}{\partial y}(1,1,1)\right)(y-1)+\left(\frac{\partial f}{\partial z}(1,1,1)\right)(z-1)\right)\\
&\qquad+\frac{1}{2}\Bigg(\left(\frac{\partial^2 f}{\partial x^2}(1,1,1)\right)(x-1)^2+\left(\frac{\partial^2 f}{\partial y^2}(1,1,1)\right)(y-1)^2+\left(\frac{\partial^2 f}{\partial z^2}(1,1,1)\right)(z-1)^2\\
&\qquad\qquad+2\left(\frac{\partial^2 f}{\partial x\partial y}(1,1,1)\right)(x-1)(y-1)+2\left(\frac{\partial^2 f}{\partial x\partial z}(1,1,1)\right)(x-1)(z-1)\\
&\qquad\qquad+2\left(\frac{\partial^2 f}{\partial y\partial z}(1,1,1)\right)(y-1)(z-1)\Bigg)+R_2(x,y,z)\\
&=3+3(x-1)+3(y-1)+5(z-1)+(x-1)^2+(y-1)^2+4(z-1)^2\\
&\qquad+2(x-1)(y-1)+3(x-1)(z-1)+4(y-1)(z-1)+R_2(x,y,z),
\end{aligned}$$

it follows that the second-order Taylor approximation of f at $(1,1,1)$ is given by

$$3+3(x-1)+3(y-1)+5(z-1)+(x-1)^2+(y-1)^2+4(z-1)^2+2(x-1)(y-1)+3(x-1)(z-1)+4(y-1)(z-1),$$

and the Hessian of f at $(1,1,1)$ is given by

$$(x-1)^2+(y-1)^2+4(z-1)^2+2(x-1)(y-1)+3(x-1)(z-1)+4(y-1)(z-1).$$

4.3. Stationary Points

In this section, we study stationary points using an approach which allows us to generalize our technique for functions of two real variables. Throughout this section, we shall consider functions of the type $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set. We shall assume that f has continuous iterated second partial derivatives.

DEFINITION. A point $\mathbf{x}_0 \in A$ is said to be a stationary point of f if the total derivative $(\mathbf{D}f)(\mathbf{x}_0) = \mathbf{0}$, where $\mathbf{0}$ denotes the zero $1 \times n$ matrix.

REMARK. In other words, $\mathbf{x}_0 \in A$ is a stationary point of f if

$$\frac{\partial f}{\partial x_i}(\mathbf{x}_0) = 0$$

for every $i = 1, \ldots, n$.

DEFINITION. A point $\mathbf{x}_0 \in A$ is said to be a (local) maximum of f if there exists a neighbourhood U of $\mathbf{x}_0$ such that $f(\mathbf{x}) \leq f(\mathbf{x}_0)$ for every $\mathbf{x} \in U$.

DEFINITION. A point $\mathbf{x}_0 \in A$ is said to be a (local) minimum of f if there exists a neighbourhood U of $\mathbf{x}_0$ such that $f(\mathbf{x}) \geq f(\mathbf{x}_0)$ for every $\mathbf{x} \in U$.

DEFINITION. A stationary point $\mathbf{x}_0 \in A$ that is not a maximum or minimum of f is said to be a saddle point of f.

Our first task is to show that if f is differentiable, then every maximum or minimum of f is a stationary point of f. Note that this may not be the case if the function f is not differentiable, as can be observed for the function $f : \mathbb{R} \to \mathbb{R} : x \mapsto |x|$ at the point $x = 0$.

THEOREM 4D. *Suppose that the function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set, is differentiable. Suppose further that $\mathbf{x}_0 \in A$ is a maximum or minimum of f. Then $\mathbf{x}_0$ is a stationary point of f.*

† PROOF. Suppose that $\mathbf{x}_0 \in A$ is a maximum of f. Consider the restriction of f to a line through $\mathbf{x}_0$. More precisely, consider the points $\mathbf{x}_0 + t\mathbf{h} \in \mathbb{R}^n$, where $\mathbf{0} \neq \mathbf{h} \in \mathbb{R}^n$ is fixed. Since A is open, there exists an open interval I containing $t = 0$ and such that $\{\mathbf{x}_0 + t\mathbf{h} : t \in I\} \subseteq A$. Consider now the line segment

$$L : I \to \mathbb{R}^n : t \mapsto \mathbf{x}_0 + t\mathbf{h}.$$

Since the function f has a maximum at $\mathbf{x}_0$, it follows that the function

$$g = f \circ L : I \to \mathbb{R},$$

where $g(t) = f(\mathbf{x}_0 + t\mathbf{h})$ for every $t \in I$, has a maximum at $t = 0$. By the Chain rule, g is differentiable. Since

$$g'(0) = \lim_{t \to 0} \frac{g(t) - g(0)}{t - 0},$$

it clearly follows that

$$g'(0) = \lim_{t\to 0+} \frac{g(t)-g(0)}{t-0} \le 0 \qquad \text{and} \qquad g'(0) = \lim_{t\to 0-} \frac{g(t)-g(0)}{t-0} \ge 0,$$

and so $g'(0) = 0$, whence $(\mathbf{D}g)(0) = 0$. Again, by the Chain rule, we have

$$(\mathbf{D}g)(0) = (\mathbf{D}f)(L(0))(\mathbf{D}L)(0).$$

It is easy to check that $(\mathbf{D}L)(0) = \mathbf{h}$, and so $(\mathbf{D}f)(L(0))\mathbf{h} = 0$. Since $\mathbf{h} \neq \mathbf{0}$ is arbitrary, we must have $(\mathbf{D}f)(\mathbf{x}_0) = (\mathbf{D}f)(L(0)) = \mathbf{0}$. The case when $\mathbf{x}_0 \in A$ is a minimum of f can be studied by considering the function $-f$. ○

It is a consequence of (7) that if f has a stationary point at $\mathbf{x}_0 \in A$, then

$$f(\mathbf{x}) = f(\mathbf{x}_0) + \mathbf{H}f(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0) + R_2(\mathbf{x}). \tag{8}$$

It follows that the Hessian $\mathbf{H}f(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)$ plays a crucial role in the determination of the nature of the stationary point. Recall that the Hessian is given by (6). Let us write $\mathbf{h} = \mathbf{x}-\mathbf{x}_0$ and $\mathbf{h} = (h_1, \ldots, h_n)$. Then (6) becomes

$$\mathbf{H}f(\mathbf{x}_0)(\mathbf{h}) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0)\right) h_i h_j = \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_{ij} h_i h_j,$$

where, for every $i,j = 1,\ldots,n$, we have

$$\alpha_{ij} = \frac{1}{2}\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}_0).$$

A function of the type

$$g(\mathbf{h}) = g(h_1,\ldots,h_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_{ij} h_i h_j \tag{9}$$

is called a quadratic function. Note that if we write

$$B = \begin{pmatrix} \beta_{11} & \ldots & \beta_{1n} \\ \vdots & & \vdots \\ \beta_{n1} & \ldots & \beta_{nn} \end{pmatrix} \qquad \text{and} \qquad \mathbf{h} = \begin{pmatrix} h_1 \\ \vdots \\ h_n \end{pmatrix},$$

then

$$g(\mathbf{h}) = \mathbf{h}^t B \mathbf{h}.$$

Clearly for any real number $\lambda \in \mathbb{R}$, we have

$$g(\lambda\mathbf{h}) = (\lambda\mathbf{h})^t B(\lambda\mathbf{h}) = \lambda^2 \mathbf{h}^t B \mathbf{h} = \lambda^2 g(\mathbf{h}); \tag{10}$$

hence the term "quadratic".

Definition. A quadratic function $g : \mathbb{R}^n \to \mathbb{R}$ is said to be positive definite if $g(\mathbf{0}) = 0$ and $g(\mathbf{h}) > 0$ for every non-zero $\mathbf{h} \in \mathbb{R}^n$.

Definition. A quadratic function $g : \mathbb{R}^n \to \mathbb{R}$ is said to be negative definite if $g(\mathbf{0}) = 0$ and $g(\mathbf{h}) < 0$ for every non-zero $\mathbf{h} \in \mathbb{R}^n$.

THEOREM 4E. *Suppose that the function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set, has continuous iterated second partial derivatives. Suppose further that $\mathbf{x}_0 \in A$ is a stationary point of f.*
(a) If the Hessian $\mathbf{H}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = \mathbf{H}f(\mathbf{x}_0)(\mathbf{h})$ is positive definite, then f has a minimum at $\mathbf{x}_0$.
(b) If the Hessian $\mathbf{H}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = \mathbf{H}f(\mathbf{x}_0)(\mathbf{h})$ is negative definite, then f has a maximum at $\mathbf{x}_0$.

EXAMPLE 4.3.1. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by

$$f(x, y) = x^3 + y^3 - 3x - 12y + 4.$$

Then

$$(\mathbf{D}f)(x, y) = \begin{pmatrix} \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix} = \begin{pmatrix} 3x^2 - 3 & 3y^2 - 12 \end{pmatrix}.$$

For stationary points, we need $3x^2 - 3 = 0$ and $3y^2 - 12 = 0$, so there are four stationary points $(\pm 1, \pm 2)$. Now

$$\frac{\partial^2 f}{\partial x^2} = 6x \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2} = 6y \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x \partial y} = 0.$$

At the stationary point $(1, 2)$, we have

$$\frac{\partial^2 f}{\partial x^2}(1, 2) = 6 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2}(1, 2) = 12 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x \partial y}(1, 2) = 0.$$

Hence the Hessian of f at $(1, 2)$ is given by

$$\begin{aligned} &\frac{1}{2}\left(\left(\frac{\partial^2 f}{\partial x^2}(1,2)\right)(x-1)^2 + \left(\frac{\partial^2 f}{\partial y^2}(1,2)\right)(y-2)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}(1,2)\right)(x-1)(y-2)\right) \\ &\quad = 3(x-1)^2 + 6(y-2)^2 \end{aligned}$$

and is positive definite. It follows that f has a minimum at $(1, 2)$. At the stationary point $(-1, -2)$, we have

$$\frac{\partial^2 f}{\partial x^2}(-1, -2) = -6 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2}(-1, -2) = -12 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x \partial y}(-1, -2) = 0.$$

Hence the Hessian of f at $(-1, -2)$ is given by

$$\begin{aligned} &\frac{1}{2}\left(\left(\frac{\partial^2 f}{\partial x^2}(-1,-2)\right)(x+1)^2 + \left(\frac{\partial^2 f}{\partial y^2}(-1,-2)\right)(y+2)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}(-1,-2)\right)(x+1)(y+2)\right) \\ &\quad = -3(x+1)^2 - 6(y+2)^2 \end{aligned}$$

and is negative definite. It follows that f has a maximum at $(-1, -2)$. At the stationary point $(1, -2)$, we have

$$\frac{\partial^2 f}{\partial x^2}(1, -2) = 6 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial y^2}(1, -2) = -12 \qquad \text{and} \qquad \frac{\partial^2 f}{\partial x \partial y}(1, -2) = 0.$$

Hence the Hessian of f at $(1, -2)$ is given by

$$\begin{aligned} &\frac{1}{2}\left(\left(\frac{\partial^2 f}{\partial x^2}(1,-2)\right)(x-1)^2 + \left(\frac{\partial^2 f}{\partial y^2}(1,-2)\right)(y+2)^2 + 2\left(\frac{\partial^2 f}{\partial x \partial y}(1,-2)\right)(x-1)(y+2)\right) \\ &\quad = 3(x-1)^2 - 6(y+2)^2. \end{aligned}$$

Let us investigate the function

$$\mathbf{H}f(1,-2)(h_1,h_2) = 3h_1^2 - 6h_2^2$$

more closely. Note that

$$\mathbf{H}f(1,-2)(h_1,0) = 3h_1^2 \geq 0 \qquad \text{and} \qquad \mathbf{H}f(1,-2)(0,h_2) = 0 - 6h_2^2 \leq 0.$$

In this case, Theorem 4E does not give any conclusion. In fact, both stationary points $(1,-2)$ and $(-1,2)$ are saddle points.

Remark. To prove Theorem 4E, we need the following result in linear algebra. Suppose that a quadratic function of the type (9) is positive definite. Then there exists a constant $M > 0$ such that for every $\mathbf{h} \in \mathbb{R}^n$, we have

$$g(\mathbf{h}) \geq M\|\mathbf{h}\|^2. \tag{11}$$

To see this, consider the restriction

$$g_S : S \to \mathbb{R} : \mathbf{h} \mapsto g(\mathbf{h})$$

of the function g to the unit sphere $S = \{\mathbf{h} \in \mathbb{R}^n : \|\mathbf{h}\| = 1\}$. The function g_S is continuous in S and has a minimum value $M > 0$, say. Then for any non-zero $\mathbf{h} \in \mathbb{R}^n$, we have, noting (10), that

$$g(\mathbf{h}) = g\left(\|\mathbf{h}\|\frac{\mathbf{h}}{\|\mathbf{h}\|}\right) = \|\mathbf{h}\|^2 g\left(\frac{\mathbf{h}}{\|\mathbf{h}\|}\right) = \|\mathbf{h}\|^2 g_S\left(\frac{\mathbf{h}}{\|\mathbf{h}\|}\right) \geq M\|\mathbf{h}\|^2.$$

Hence (11) holds for any non-zero $\mathbf{h} \in \mathbb{R}^n$. Clearly it also holds for $\mathbf{h} = \mathbf{0}$.

† Sketch of Proof of Theorem 4E. We shall attempt to demonstrate the theorem by making the extra assumption that iterated third partial derivatives exist and are continuous. At a stationary point $\mathbf{x}_0$, the expression (8) is valid, and can be rewritten in the form

$$f(\mathbf{x}) - f(\mathbf{x}_0) = \mathbf{H}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) + R_2(\mathbf{x}),$$

where $R_2(\mathbf{x})$ satisfies (5). Suppose that $\mathbf{H}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$ is positive definite. Then by our remark on linear algebra, there exists $M > 0$ such that

$$\mathbf{H}f(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) \geq M\|\mathbf{x} - \mathbf{x}_0\|^2$$

for every $\mathbf{x} \in \mathbb{R}^n$. On the other hand, it follows from (5) that

$$|R_2(\mathbf{x})| \leq \frac{1}{2}M\|\mathbf{x} - \mathbf{x}_0\|^2,$$

provided that $\|\mathbf{x} - \mathbf{x}_0\|$ is sufficiently small. Hence

$$f(\mathbf{x}) - f(\mathbf{x}_0) \geq \frac{1}{2}M\|\mathbf{x} - \mathbf{x}_0\|^2 \geq 0,$$

provided that $\|\mathbf{x} - \mathbf{x}_0\|$ is sufficiently small, whence f has a minimum at $\mathbf{x}_0$. The negative definite case can be studied by considering the function $-f$. ◯

4.4. Functions of Two Variables

We now attempt to link the Hessian to the discriminant. Suppose that a function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^2$ is an open set, has continuous iterated second partial derivatives. Suppose further that f has a stationary point at (x_0, y_0). Then the Hessian is given by

$$\begin{aligned}
&\frac{1}{2}\Bigg(\left(\frac{\partial^2 f}{\partial x^2}(x_0,y_0)\right)(x-x_0)^2 + \left(\frac{\partial^2 f}{\partial x \partial y}(x_0,y_0)\right)(x-x_0)(y-y_0) \\
&\qquad + \left(\frac{\partial^2 f}{\partial y \partial x}(x_0,y_0)\right)(x-x_0)(y-y_0) + \left(\frac{\partial^2 f}{\partial y^2}(x_0,y_0)\right)(y-y_0)^2\Bigg) \\
&= \frac{1}{2}\begin{pmatrix} x-x_0 & y-y_0 \end{pmatrix}\begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(x_0,y_0) & \dfrac{\partial^2 f}{\partial y \partial x}(x_0,y_0) \\ \dfrac{\partial^2 f}{\partial x \partial y}(x_0,y_0) & \dfrac{\partial^2 f}{\partial y^2}(x_0,y_0) \end{pmatrix}\begin{pmatrix} x-x_0 \\ y-y_0 \end{pmatrix}.
\end{aligned}$$

REMARK. We need the following result in linear algebra. The quadratic function

$$g(x,y) = \begin{pmatrix} x & y \end{pmatrix}\begin{pmatrix} a & b \\ b & c \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix},$$

where $a, b, c \in \mathbb{R}$, is positive definite if and only if $a > 0$ and $ac - b^2 > 0$. To see this, note that

$$g(x,y) = ax^2 + 2bxy + cy^2.$$

Suppose first of all that $a > 0$ and $ac - b^2 > 0$. Completing squares, we have

$$g(x,y) = a\left(x + \frac{b}{a}y\right)^2 + \left(c - \frac{b^2}{a}\right)y^2 \geq 0, \tag{12}$$

with equality only when

$$y = 0 \qquad \text{and} \qquad x + \frac{b}{a}y = 0;$$

in other words, when $(x, y) = (0, 0)$. Suppose now that $a = 0$. Then $g(x, y) = 2bxy + cy^2$ clearly cannot be positive definite (why?). It follows that if $g(x, y)$ is positive definite, then $a \neq 0$ and (12) holds, with strict inequality whenever $(x, y) \neq (0, 0)$. Setting $y = 0$, we conclude that we must have $a > 0$. Setting $x = -by/a$, we conclude that we must have $ac - b^2 > 0$.

We have essentially proved the following result.

THEOREM 4F. *Suppose that the function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^2$ is an open set, has continuous iterated second partial derivatives. Suppose further that $(x_0, y_0) \in A$ is a stationary point of f.*
(a) If

$$\frac{\partial^2 f}{\partial x^2}(x_0,y_0) > 0 \qquad \textit{and} \qquad \Delta = \det\begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(x_0,y_0) & \dfrac{\partial^2 f}{\partial y \partial x}(x_0,y_0) \\ \dfrac{\partial^2 f}{\partial x \partial y}(x_0,y_0) & \dfrac{\partial^2 f}{\partial y^2}(x_0,y_0) \end{pmatrix} > 0,$$

then f has a minimum at (x_0, y_0).
(b) If

$$\frac{\partial^2 f}{\partial x^2}(x_0,y_0) < 0 \qquad \textit{and} \qquad \Delta = \det\begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2}(x_0,y_0) & \dfrac{\partial^2 f}{\partial y \partial x}(x_0,y_0) \\ \dfrac{\partial^2 f}{\partial x \partial y}(x_0,y_0) & \dfrac{\partial^2 f}{\partial y^2}(x_0,y_0) \end{pmatrix} > 0,$$

then f has a maximum at (x_0, y_0).

Remark. The reader may wish to re-examine Example 4.3.1 using this result.

4.5. Constrained Maxima and Minima

In this last section, we consider the problem of finding maxima and minima of functions of n variables, where the veriables are not always independent of each other but are subject to some constraints. In the case of one constraint, we have the following useful result.

THEOREM 4G. *Suppose that the functions $f : A \to \mathbb{R}$ and $g : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set, have continuous partial derivatives. Suppose next that $c \in \mathbb{R}$ is fixed, and $S = \{\mathbf{x} \in A : g(\mathbf{x}) = c\}$. Suppose further that the function $f|_S$, the restriction of f to S, has a maximum or minimum at $\mathbf{x}_0 \in S$, and that $(\nabla g)(\mathbf{x}_0) \neq \mathbf{0}$. Then there exists a real number $\lambda \in \mathbb{R}$ such that $(\nabla f)(\mathbf{x}_0) = \lambda(\nabla g)(\mathbf{x}_0)$.*

Remarks. (1) The restriction of f to $S \subseteq A$ is the function $f|_S : S \to \mathbb{R} : \mathbf{x} \mapsto f(\mathbf{x})$.

(2) The number λ is called the Lagrange multiplier.

(3) Note that $(\nabla g)(\mathbf{x}_0)$ is a vector which is orthogonal to the surface S at $\mathbf{x}_0$. It follows that if f has a maximum or minimum at $\mathbf{x}_0$, then $(\nabla f)(\mathbf{x}_0)$ must be orthogonal to the surface S at $\mathbf{x}_0$.

† Sketch of Proof of Theorem 4G. We shall only consider the case $n = 3$. Suppose that $I \subseteq \mathbb{R}$ is an open interval containing the number 0. Suppose further that

$$L : I \to \mathbb{R}^3 : t \mapsto L(t) = (L_1(t), L_2(t), L_3(t))$$

is a path on S, with $L(0) = \mathbf{x}_0$, so that $L(t) \in S$ for every $t \in I$. Consider first of all the function $h = g \circ L : I \to \mathbb{R}$. Clearly $h(t) = g(L(t)) = c$ for every $t \in I$. It follows that

$$(\mathbf{D}h)(0) = \frac{\mathrm{d}h}{\mathrm{d}t}(0) = 0.$$

On the other hand, it follows from the Chain rule that

$$(\mathbf{D}h)(0) = (\mathbf{D}g)(L(0))(\mathbf{D}L)(0) = (\nabla g)(\mathbf{x}_0) \cdot (L_1'(0), L_2'(0), L_3'(0)),$$

so that $(\nabla g)(\mathbf{x}_0)$ is perpendicular to $(L_1'(0), L_2'(0), L_3'(0))$, a tangent vector to S at $\mathbf{x}_0$. Since L is arbitrary, it follows that $(\nabla g)(\mathbf{x}_0)$ must be perpendicular to the tangent plane to S at $\mathbf{x}_0$. Consider next the function $k = f \circ L : I \to \mathbb{R}$. If $f|_S$ has a maximum or minimum at $\mathbf{x}_0$, then clearly k has a maximum or minimum at $t = 0$. It follows that

$$(\mathbf{D}k)(0) = \frac{\mathrm{d}k}{\mathrm{d}t}(0) = 0.$$

On the other hand, it follows from the Chain rule that

$$(\mathbf{D}k)(0) = (\mathbf{D}f)(L(0))(\mathbf{D}L)(0) = (\nabla f)(\mathbf{x}_0) \cdot (L_1'(0), L_2'(0), L_3'(0)),$$

so that $(\nabla f)(\mathbf{x}_0)$ is perpendicular to $(L_1'(0), L_2'(0), L_3'(0))$. Since L is arbitrary, it follows as before that $(\nabla f)(\mathbf{x}_0)$ must also be perpendicular to the tangent plane to S at $\mathbf{x}_0$. Since $(\nabla f)(\mathbf{x}_0)$ and $(\nabla g)(\mathbf{x}_0) \neq \mathbf{0}$ are perpendicular to the same plane, there exists a real number $\lambda \in \mathbb{R}$ such that $(\nabla f)(\mathbf{x}_0) = \lambda(\nabla g)(\mathbf{x}_0)$. ○

EXAMPLE 4.5.1. We wish to find the distance from the origin to the plane $x-2y-2z=3$. To do this, we consider the function

$$f:\mathbb{R}^3\to\mathbb{R}:(x,y,z)\mapsto x^2+y^2+z^2,$$

which represents the square of the distance from the origin to a point $(x,y,z)\in\mathbb{R}^3$. The points (x,y,z) under consideration are subject to the constraint $g(x,y,z)=3$, where

$$g:\mathbb{R}^3\to\mathbb{R}:(x,y,z)\mapsto x-2y-2z.$$

We now wish to minimize f subject to this constraint. Using the Lagrange multiplier method, we know that the minimum is attained at a point (x,y,z) which satisfies

$$(\nabla f)(x,y,z)=\lambda(\nabla g)(x,y,z)$$

for some real number $\lambda\in\mathbb{R}$. Note that

$$(\nabla f)(x,y,z)=(2x,2y,2z)\qquad\text{and}\qquad(\nabla g)(x,y,z)=(1,-2,-2).$$

Hence we need to solve the equations

$$(2x,2y,2z)=\lambda(1,-2,-2)\qquad\text{and}\qquad x-2y-2z=3.$$

Substituting the former into the latter, we obtain $\lambda=2/3$. This gives $(x,y,z)=(1/3,-2/3,-2/3)$. Clearly $f(x,y,z)=1$ at this point. Hence the minimum distance is equal to 1, the square root of $f(x,y,z)$ at this point.

EXAMPLE 4.5.2. We wish to find the volume of the largest rectangular box with edges parallel to the coordinate axes and inscribed in the ellipsoid

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1. \tag{13}$$

Clearly the box is given by $[-x,x]\times[-y,y]\times[-z,z]$ for some positive $x,y,z\in\mathbb{R}$ satisfying (13), with volume equal to $8xyz$. We therefore wish to maximize the function

$$f:\mathbb{R}^3\to\mathbb{R}:(x,y,z)\mapsto 8xyz,$$

subject to the constraint $g(x,y,z)=1$, where

$$g:\mathbb{R}^3\to\mathbb{R}:(x,y,z)\mapsto \frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}.$$

Using the Lagrange multiplier method, we know that the maximum is attained at a point (x,y,z) which satisfies

$$(\nabla f)(x,y,z)=\lambda(\nabla g)(x,y,z)$$

for some real number $\lambda\in\mathbb{R}$. Note that

$$(\nabla f)(x,y,z)=(8yz,8xz,8xy)\qquad\text{and}\qquad(\nabla g)(x,y,z)=\left(\frac{2x}{a^2},\frac{2y}{b^2},\frac{2z}{c^2}\right).$$

Hence we need to solve the equations (13) and

$$(8yz,8xz,8xy)=\lambda\left(\frac{2x}{a^2},\frac{2y}{b^2},\frac{2z}{c^2}\right). \tag{14}$$

Since $x, y, z > 0$, it follows from (14) that

$$8xyz = \frac{2\lambda x^2}{a^2} = \frac{2\lambda y^2}{b^2} = \frac{2\lambda z^2}{c^2}, \tag{15}$$

so that combining with (13), we have

$$24xyz = \frac{2\lambda x^2}{a^2} + \frac{2\lambda y^2}{b^2} + \frac{2\lambda z^2}{c^2} = 2\lambda\left(\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}\right) = 2\lambda,$$

whence $\lambda = 12xyz$. Substituting this into the left hand side of (15), we deduce that

$$\frac{x^2}{a^2} = \frac{y^2}{b^2} = \frac{z^2}{c^2} = \frac{1}{3},$$

giving

$$(x, y, z) = \left(\frac{a}{\sqrt{3}}, \frac{b}{\sqrt{3}}, \frac{c}{\sqrt{3}}\right) \qquad \text{and} \qquad f(x, y, z) = \frac{8abc}{3\sqrt{3}}.$$

Remark. In Example 4.5.2, we clearly have constrained minima at points such as $(a, 0, 0)$. Note, however, that we have dispensed with such trivial cases by considering only positive values of x, y, z. Note that (15) is obtained only under such a specialization.

In the case of more than one constraint, we have the following generalized version of Theorem 4G.

THEOREM 4H. *Suppose that the functions $f : A \to \mathbb{R}$ and $g_i : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}^n$ is an open set and $i = 1, \ldots, k$, have continuous partial derivatives. Suppose next that $c_1, \ldots, c_k \in \mathbb{R}$ are fixed, and $S = \{\mathbf{x} \in A : g_i(\mathbf{x}) = c_i \text{ for every } i = 1, \ldots, k\}$. Suppose further that the function $f|_S$, the restriction of f to S, has a maximum or minimum at $\mathbf{x}_0 \in S$, and that $(\nabla g_1)(\mathbf{x}_0), \ldots, (\nabla g_k)(\mathbf{x}_0)$ are linearly independent over $\mathbb{R}$. Then there exist real numbers $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that*

$$(\nabla f)(\mathbf{x}_0) = \lambda_1(\nabla g_1)(\mathbf{x}_0) + \ldots + \lambda_k(\nabla g_k)(\mathbf{x}_0).$$

Example 4.5.3. We wish to find the distance from the origin to the intersection of $xy = 12$ and $x + 2z = 0$. To do this, we consider the function

$$f : \mathbb{R}^3 \to \mathbb{R} : (x, y, z) \mapsto x^2 + y^2 + z^2,$$

which represents the square of the distance from the origin to a point $(x, y, z) \in \mathbb{R}^3$. The points (x, y, z) under consideration are subject to the constraints $g_1(x, y, z) = 12$ and $g_2(x, y, z) = 0$, where

$$g_1 : \mathbb{R}^3 \to \mathbb{R} : (x, y, z) \mapsto xy \qquad \text{and} \qquad g_2 : \mathbb{R}^3 \to \mathbb{R} : (x, y, z) \mapsto x + 2z.$$

We now wish to minimize f subject to these constraints. Using the Lagrange multiplier method, we know that the minimum is attained at a point (x, y, z) which satisfies

$$(\nabla f)(x, y, z) = \lambda_1(\nabla g_1)(x, y, z) + \lambda_2(\nabla g_2)(x, y, z)$$

for some real numbers $\lambda_1, \lambda_2 \in \mathbb{R}$. Note that

$$(\nabla f)(x, y, z) = (2x, 2y, 2z) \qquad \text{and} \qquad (\nabla g_1)(x, y, z) = (y, x, 0) \qquad \text{and} \qquad (\nabla g_2)(x, y, z) = (1, 0, 2).$$

Hence we need to solve the equations

$$(2x, 2y, 2z) = \lambda_1(y, x, 0) + \lambda_2(1, 0, 2) \qquad \text{and} \qquad xy = 12 \qquad \text{and} \qquad x + 2z = 0.$$

Eliminating λ_1 and λ_2 from this system of five equations, we conclude (after a fair bit of calculation) that

$$(x, y, z) = \left(\pm 2 \sqrt[4]{\frac{36}{5}}, \pm 6 \sqrt[4]{\frac{5}{36}}, \mp \sqrt[4]{\frac{36}{5}} \right) \qquad \text{and} \qquad f(x, y, z) = 12\sqrt{5}.$$

Hence the minimum distance is equal to $\sqrt{12\sqrt{5}}$, the square root of $f(x, y, z)$ at this point.

Problems for Chapter 4

1. For each of the following functions, find the second-order Taylor approximation at the given point:
   a) $f(x, y) = x\cos(xy) + y\sin(xy)$; $(x_0, y_0) = (0, 0)$
   b) $f(x, y, z) = \mathrm{e}^{xz}y^2 + \sin y \cos z + x^2 z$; $(x_0, y_0, z_0) = (0, \pi, 0)$

2. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$f(x, y) = x^3 - y^3 - 3xy + 4.$$

   a) Show that the total derivatives $(\mathbf{D}f)(-1, 1) = 0$ and $(\mathbf{D}f)(0, 0) = 0$.
   b) Find the second-order Taylor approximations to $f(x, y)$ at the points $(-1, 1)$ and $(0, 0)$.
   c) Find the Hessians $(\mathbf{H}f)(-1, 1)$ and $(\mathbf{H}f)(0, 0)$.
   d) Is $(\mathbf{H}f)(-1, 1)$ positive definite? Negative definite? Comment on the result.
   e) Is $(\mathbf{H}f)(0, 0)$ positive definite? Negative definite?
   f) Find the discriminant of f at $(0, 0)$.
   g) Comment on your observations in (e), (f) and the second part of (a).

3. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by

$$f(x, y) = \frac{x^4 - 4x^3 + 4x^2 - 3}{1 + y^2}.$$

   a) Find the total derivative $(\mathbf{D}f)(x, y)$.
   b) Show that the three stationary points are $(0, 0)$, $(1, 0)$ and $(2, 0)$.
   c) Evaluate the partial derivatives $\partial^2 f/\partial x^2$, $\partial^2 f/\partial y^2$ and $\partial^2 f/\partial x \partial y$, and find the Hessian of f at each of the stationary points.
   d) Show that the Hessian of f at $(0, 0)$ and at $(2, 0)$ are positive definite.
   e) Find the discriminant of f at $(1, 0)$.
   f) Classify the stationary points.

4. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x, y) = x^3 + y^3 + 9x^2 + 9y^2 + 12xy$.
   a) Show that $(0, 0)$, $(-10, -10)$, $(-4, 2)$ and $(2, -4)$ are stationary points.
   b) Find the Hessian of f at $(0, 0)$ and show that it is positive definite.
   c) Find the Hessian of f at $(-10, -10)$ and show that it is negative definite.
   d) Classify the stationary points $(0, 0)$ and $(-10, -10)$.
   e) Find the discriminant of f at the other two stationary points, and classify these stationary points.

5. Consider the function $f : \mathbb{R}^3 \to \mathbb{R}$, defined by $f(x, y, z) = x^2 + y^2 + z^2 - 6xy + 8xz - 10yz$.
   a) Show that $(\mathbf{D}f)(x, y, z) = \mathbf{0}$ leads to a system of three linear equations with unique solution $(x, y, z) = (0, 0, 0)$.
   b) Without any calculation, can you write down the Hessian of f at $(0, 0, 0)$?
   c) If you cannot do (b), then proceed to calculate the Hessian of f at $(0, 0, 0)$. Then try to understand the surprise (assuming that your calculation is correct).

6. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x, y) = 4x^2 - 12xy + 9y^2$.
   a) Show that f has infinitely many stationary points.
   b) Show that the Hessian at any stationary point of f is given by the same function $(2x - 3y)^2$.
   c) Can you classify these stationary points?
   [HINT: Dispense with the theorems and have some fun instead.]

7. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$, defined by $f(x, y) = (y - x^2)(y - 2x^2)$.
   a) Show that $(0, 0)$ is a stationary point of f.
   b) Find the Hessian of f at $(0, 0)$. Is it positive definite? Negative definite?
   c) Show that on any line through the origin, f has a minimum at $(0, 0)$.
   [HINT: Consider three cases: $y = 0$, $x = 0$ and $y = \alpha x$ where α is any non-zero real number.]
   d) Draw a picture of the two parabolas $y = x^2$ and $y = 2x^2$ on the plane. Note that $f(x, y)$ is a product of two factors which are non-zero at any point (x, y) not on the parabolas. Shade in one colour the region in $\mathbb{R}^2$ for which $f(x, y) > 0$, and in another colour the region in $\mathbb{R}^2$ for which $f(x, y) < 0$. Convince yourself that f has a saddle point at $(0, 0)$.

8. Follow the steps indicated below to find the shortest distance from the origin to the hyperbola $x^2 + 8xy + 7y^2 = 225$. Write $f(x, y) = x^2 + y^2$ and $g(x, y) = x^2 + 8xy + 7y^2$. We shall minimize $f(x, y)$ subject to the constraint $g(x, y) = 225$.
   a) Let λ be a Lagrange multiplier. Show that the equation $(\nabla f)(x, y) - \lambda(\nabla g)(x, y) = 0$ can be rewritten as a system of two homogeneous linear equations in x and y, where some of the coefficients depend on λ.
   b) Clearly $(x, y) \neq (0, 0)$. It follows that the system of homogeneous linear equations in (a) has non-trivial solution, and so the determinant of the corresponding matrix is zero. Use this fact to find two roots for λ.
   c) Show that one of the roots λ in (b) leads to no real solution of the system, while the other root λ leads to a solution. Use this solution to minimize $f(x, y)$.
   d) What is the shortest distance from the origin to the hyperbola?

9. Find the point on the paraboloid $z = x^2 + y^2$ which is closest to the point $(3, -6, 4)$.

10. Find the extreme values of z on the surface $2x^2 + 3y^2 + z^2 - 12xy + 4xz = 35$.