

Module MA2321: Analysis in Several Real
Variables
Michaelmas Term 2016
Section 10: Second Order Partial Derivatives
and the Hessian Matrix

D. R. Wilkins

Copyright © David R. Wilkins 2015–2016

Contents

10 Second Order Partial Derivatives and the Hessian Matrix 151
10.1 Second Order Partial Derivatives 151
10.2 Maxima and Minima for Functions of Several Real Variables . 155

10 Second Order Partial Derivatives and the Hessian Matrix

10.1 Second Order Partial Derivatives

Let X be an open subset of \mathbb{R}^n and let $f: X \rightarrow \mathbb{R}$ be a real-valued function on X . We consider the second order partial derivatives of the function f defined by

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right).$$

We shall show that if the partial derivatives

$$\frac{\partial f}{\partial x_i}, \quad \frac{\partial f}{\partial x_j}, \quad \frac{\partial^2 f}{\partial x_i \partial x_j} \quad \text{and} \quad \frac{\partial^2 f}{\partial x_j \partial x_i}$$

all exist and are continuous then

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

First though we give a counterexample which demonstrates that there exist functions f for which

$$\frac{\partial^2 f}{\partial x_i \partial x_j} \neq \frac{\partial^2 f}{\partial x_j \partial x_i}.$$

Example Let $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ be the function defined by

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

For convenience of notation, let us write

$$\begin{aligned} f_x(x, y) &= \frac{\partial f(x, y)}{\partial x}, \\ f_y(x, y) &= \frac{\partial f(x, y)}{\partial y}, \\ f_{xy}(x, y) &= \frac{\partial^2 f(x, y)}{\partial x \partial y}, \\ f_{yx}(x, y) &= \frac{\partial^2 f(x, y)}{\partial y \partial x}. \end{aligned}$$

If $(x, y) \neq (0, 0)$ then

$$\begin{aligned} f_x &= \frac{(3x^2y - y^3)(x^2 + y^2) - 2x^2y(x^2 - y^2)}{(x^2 + y^2)^2} \\ &= \frac{3x^4y + 3x^2y^3 - x^2y^3 - y^5 - 2x^4y + 2x^2y^3}{(x^2 + y^2)^2} \\ &= \frac{x^4y + 4x^2y^3 - y^5}{(x^2 + y^2)^2}. \end{aligned}$$

Similarly

$$f_y = \frac{-xy^4 - 4x^3y^2 + x^5}{(x^2 + y^2)^2}.$$

(This can be deduced from the formula for f_x on noticing that $f(x, y)$ changes sign on interchanging the variables x and y .)

Differentiating again, when $(x, y) \neq (0, 0)$, we find that

$$\begin{aligned} f_{xy}(x, y) &= \frac{\partial f_y}{\partial x} \\ &= \frac{(-y^4 - 12x^2y^2 + 5x^4)(x^2 + y^2)}{(x^2 + y^2)^3} \\ &\quad + \frac{-4x(-xy^4 - 4x^3y^2 + x^5)}{(x^2 + y^2)^3} \\ &= \frac{-x^2y^4 - 12x^4y^2 + 5x^6 - y^6 - 12x^2y^4 + 5x^4y^2}{(x^2 + y^2)^3} \\ &\quad + \frac{4x^2y^4 + 16x^4y^2 - 4x^6}{(x^2 + y^2)^3} \\ &= \frac{x^6 + 9x^4y^2 - 9x^2y^4 - y^6}{(x^2 + y^2)^3}. \end{aligned}$$

Now the expression just obtained for f_{xy} when $(x, y) \neq (0, 0)$ changes sign when the variables x and y are interchanged. The same is true of the expression defining $f(x, y)$. It follows that f_{yx} . We conclude therefore that if $(x, y) \neq (0, 0)$ then

$$f_{xy} = f_{yx} = \frac{x^6 + 9x^4y^2 - 9x^2y^4 - y^6}{(x^2 + y^2)^3}.$$

Now if $(x, y) \neq (0, 0)$ and if $r = \sqrt{x^2 + y^2}$ then

$$|f_x(x, y)| = \frac{|x^4y + 4x^2y^3 - y^5|}{r^4} \leq \frac{6r^5}{r^4} = 6r.$$

It follows that

$$\lim_{(x,y) \rightarrow (0,0)} f_x(x, y) = 0.$$

Similarly

$$\lim_{(x,y) \rightarrow (0,0)} f_y(x, y) = 0.$$

However

$$\lim_{(x,y) \rightarrow (0,0)} f_{xy}(x, y)$$

does not exist. Indeed

$$\begin{aligned} \lim_{x \rightarrow 0} f_{xy}(x, 0) &= \lim_{x \rightarrow 0} f_{yx}(x, 0) = \lim_{x \rightarrow 0} \frac{x^6}{x^6} = 1, \\ \lim_{y \rightarrow 0} f_{xy}(0, y) &= \lim_{y \rightarrow 0} f_{yx}(0, y) = \lim_{y \rightarrow 0} \frac{-y^6}{y^6} = -1. \end{aligned}$$

Next we show that f_x , f_y , f_{xy} and f_{yx} all exist at $(0,0)$, and thus exist everywhere on \mathbb{R}^2 . Now $f(x, 0) = 0$ for all x , hence $f_x(0, 0) = 0$. Also $f(0, y) = 0$ for all y , hence $f_y(0, 0) = 0$. Thus

$$f_y(x, 0) = x, \quad f_x(0, y) = -y$$

for all $x, y \in \mathbb{R}$. We conclude that

$$\begin{aligned} f_{xy}(0, 0) &= \left. \frac{d(f_y(x, 0))}{dx} \right|_{x=0} = 1, \\ f_{yx}(0, 0) &= \left. \frac{d(f_x(0, y))}{dy} \right|_{y=0} = -1, \end{aligned}$$

Thus

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

at $(0, 0)$.

Observe that in this example the functions f_{xy} and f_{yx} are continuous throughout $\mathbb{R}^2 \setminus \{(0, 0)\}$ and are equal to one another there. Although the functions f_{xy} and f_{yx} are well-defined at $(0, 0)$, they are not continuous at $(0, 0)$ and $f_{xy}(0, 0) \neq f_{yx}(0, 0)$.

Theorem 10.1 *Let X be an open set in \mathbb{R}^2 and let $f: X \rightarrow \mathbb{R}$ be a real-valued function on X . Suppose that the partial derivatives*

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

exist and are continuous throughout X . Then the partial derivative

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

exists and is continuous on X , and

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

Proof Let

$$f_x(x, y) = \frac{\partial f}{\partial x}, \quad f_y(x, y) = \frac{\partial f}{\partial y}, \quad f_{xy} = \frac{\partial^2 f}{\partial x \partial y} \text{ and } f_{yx} = \frac{\partial^2 f}{\partial y \partial x}$$

and let (a, b) be a point of X . The set X is open in \mathbb{R}^n and therefore there exists some positive real number L such that $(a + h, b + k) \in X$ for all $(h, k) \in \mathbb{R}^2$ satisfying $|h| < L$ and $|k| < L$. Let

$$S(h, k) = f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b)$$

for all real numbers h and k satisfying $|h| < L$ and $|k| < L$. We use the Mean Value Theorem (Theorem 2.2) to prove the existence of real numbers u and v , where u lies between a and $a + h$ and v lies between b and $b + k$, for which

$$S(h, k) = hk \left. \frac{\partial^2 f}{\partial x \partial y} \right|_{(x,y)=(u,v)} = hk f_{xy}(u, v).$$

Let h be fixed, where $|h| < L$, and let $q: (b - L, b + L) \rightarrow \mathbb{R}$ be defined so that $q(t) = f(a + h, t) - f(a, t)$ for all real numbers t satisfying $b - L < t < b + L$. Then $S(h, k) = q(b + k) - q(b)$. But it follows from the Mean Value Theorem (Theorem 2.2) that there exists some real number v lying between b and $b + k$ for which $q(b + k) - q(b) = kq'(v)$. But $q'(v) = f_y(a + h, v) - f_y(a, v)$. It follows that

$$S(h, k) = k(f_y(a + h, v) - f_y(a, v)).$$

The Mean Value Theorem can now be applied to the function sending real numbers s in the interval $(a - L, a + L)$ to $f_y(s, v)$ to deduce the existence of a real number u lying between a and $a + h$ for which

$$S(h, k) = hk f_{xy}(u, v).$$

Now let some positive real number ε be given. The function f_{xy} is continuous. Therefore there exists some real number δ satisfying $0 < \delta < L$ such

that $|f_{xy}(a+h, b+k) - f_{xy}(a, b)| \leq \varepsilon$ whenever $|h| < \delta$ and $|k| < \delta$. It follows that

$$\left| \frac{S(h, k)}{hk} - f_{xy}(a, b) \right| \leq \varepsilon$$

for all real numbers h and k satisfying $0 < |h| < \delta$ and $0 < |k| < \delta$. Now

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{S(h, k)}{hk} &= \frac{1}{k} \lim_{h \rightarrow 0} \frac{f(a+h, b+k) - f(a, b+k)}{h} \\ &\quad - \frac{1}{k} \lim_{h \rightarrow 0} \frac{f(a+h, b) - f(a, b)}{h} \\ &= \frac{f_x(a, b+k) - f_x(a, b)}{k}. \end{aligned}$$

It follows that

$$\left| \frac{f_x(a, b+k) - f_x(a, b)}{k} - f_{xy}(a, b) \right| \leq \varepsilon$$

whenever $0 < |k| < \delta$. Thus the difference quotient $\frac{f_x(a, b+k) - f_x(a, b)}{k}$ tends to $f_{xy}(a, b)$ as k tends to zero, and therefore the second order partial derivative f_{yx} exists at the point (a, b) and

$$f_{yx}(a, b) = \lim_{k \rightarrow 0} \frac{f_x(a, b+k) - f_x(a, b)}{k} = f_{xy}(a, b),$$

as required. ■

Corollary 10.2 *Let X be an open set in \mathbb{R}^n and let $f: X \rightarrow \mathbb{R}$ be a real-valued function on X . Suppose that the partial derivatives*

$$\frac{\partial f}{\partial x_i} \text{ and } \frac{\partial^2 f}{\partial x_i \partial x_j}$$

exist and are continuous on X for all integers i and j between 1 and n . Then

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

for all integers i and j between 1 and n .

10.2 Maxima and Minima for Functions of Several Real Variables

Let $f: X \rightarrow \mathbb{R}$ be a real-valued function defined over some open subset X of \mathbb{R}^n whose first and second order partial derivatives exist and are continuous

throughout X . Suppose that f has a local minimum at some point \mathbf{p} of X , where $\mathbf{p} = (p_1, p_2, \dots, p_n)$. Now for each integer i between 1 and n the map

$$t \mapsto f(p_1, \dots, p_{i-1}, t, p_{i+1}, \dots, p_n)$$

has a local minimum at $t = p_i$, hence the derivative of this map vanishes there. Thus if f has a local minimum at \mathbf{p} then

$$\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\mathbf{p}} = 0.$$

In many situations the values of the second order partial derivatives of a twice-differentiable function of several real variables at a stationary point determines the qualitative behaviour of the function around that stationary point, in particular ensuring, in some situations, that the stationary point is a local minimum or a local maximum.

Lemma 10.3 *Let f be a continuous real-valued function defined throughout an open ball in \mathbb{R}^n of radius R about some point \mathbf{p} . Suppose that the partial derivatives of f of orders one and two exist and are continuous throughout this open ball. Then there exists some real number θ satisfying $0 < \theta < 1$ for which*

$$f(\mathbf{p} + \mathbf{h}) = f(\mathbf{p}) + \sum_{k=1}^n h_k \left. \frac{\partial f}{\partial x_k} \right|_{\mathbf{p}} + \frac{1}{2} \sum_{j,k=1}^n h_j h_k \left. \frac{\partial^2 f}{\partial x_j \partial x_k} \right|_{\mathbf{p} + \theta \mathbf{h}}$$

for all $\mathbf{h} \in \mathbb{R}^n$ satisfying $|\mathbf{h}| < \delta$.

Proof Let \mathbf{h} satisfy $|\mathbf{h}| < R$, and let

$$q(t) = f(\mathbf{p} + t\mathbf{h})$$

for all $t \in [0, 1]$. It follows from the Chain Rule for functions of several variables (Theorem 8.12) that

$$q'(t) = \sum_{j=1}^n h_j (\partial_j f)(\mathbf{p} + t\mathbf{h})$$

and

$$q''(t) = \sum_{j,k=1}^n h_j h_k (\partial_j \partial_k f)(\mathbf{p} + t\mathbf{h}),$$

where

$$(\partial_j f)(x_1, x_2, \dots, x_n) = \frac{\partial f(x_1, x_2, \dots, x_n)}{\partial x_j}$$

and

$$(\partial_j \partial_k f)(x_1, x_2, \dots, x_n) = \frac{\partial^2 f(x_1, x_2, \dots, x_n)}{\partial x_j \partial x_k}.$$

Now

$$q(1) = q(0) + q'(0) + \frac{1}{2}q''(\theta)$$

for some real number θ satisfying $0 < \theta < 1$. (see Proposition 2.3). It follows that

$$\begin{aligned} f(\mathbf{p} + \mathbf{h}) &= f(\mathbf{p}) + \sum_{k=1}^n h_k (\partial_k f)(\mathbf{p}) + \frac{1}{2} \sum_{j,k=1}^n h_j h_k (\partial_j \partial_k f)(\mathbf{p} + \theta \mathbf{h}) \\ &= f(\mathbf{p}) + \sum_{k=1}^n h_k \left. \frac{\partial f}{\partial x_k} \right|_{\mathbf{p}} + \frac{1}{2} \sum_{j,k=1}^n h_j h_k \left. \frac{\partial^2 f}{\partial x_j \partial x_k} \right|_{\mathbf{p} + \theta \mathbf{h}}, \end{aligned}$$

as required. ■

Let f be a real-valued function of several variables whose first second order partial derivatives exist and are continuous throughout some open neighbourhood of a given point \mathbf{p} , and let $R > 0$ be chosen such that the function f is defined throughout the open ball of radius R about the point \mathbf{p} . It follows from Lemma 10.3 that if

$$\left. \frac{\partial f}{\partial x_j} \right|_{\mathbf{p}} = 0$$

for $j = 1, 2, \dots, n$, and if $|\mathbf{h}| < R$ then

$$f(\mathbf{p} + \mathbf{h}) = f(\mathbf{p}) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n h_i h_j \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{p}+\theta \mathbf{h}}$$

for some θ satisfying $0 < \theta < 1$.

Let us denote by $(H_{i,j}(\mathbf{p}))$ the *Hessian matrix* at the point \mathbf{p} , defined by

$$H_{i,j}(\mathbf{p}) = \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{p}}.$$

If the partial derivatives of f of second order exist and are continuous then $H_{i,j}(\mathbf{p}) = H_{j,i}(\mathbf{p})$ for all i and j , by Corollary 10.2. Thus the Hessian matrix is symmetric.

We now recall some facts concerning symmetric matrices.

Let $(c_{i,j})$ be a symmetric $n \times n$ matrix.

The matrix $(c_{i,j})$ is said to be *positive semi-definite* if $\sum_{i=1}^n \sum_{j=1}^n c_{i,j} h_i h_j \geq 0$ for all $(h_1, h_2, \dots, h_n) \in \mathbb{R}^n$.

The matrix $(c_{i,j})$ is said to be *positive definite* if $\sum_{i=1}^n \sum_{j=1}^n c_{i,j} h_i h_j > 0$ for all non-zero $(h_1, h_2, \dots, h_n) \in \mathbb{R}^n$.

The matrix $(c_{i,j})$ is said to be *negative semi-definite* if $\sum_{i=1}^n \sum_{j=1}^n c_{i,j} h_i h_j \leq 0$ for all $(h_1, h_2, \dots, h_n) \in \mathbb{R}^n$.

The matrix $(c_{i,j})$ is said to be *negative definite* if $\sum_{i=1}^n \sum_{j=1}^n c_{i,j} h_i h_j < 0$ for all non-zero $(h_1, h_2, \dots, h_n) \in \mathbb{R}^n$.

The matrix $(c_{i,j})$ is said to be *indefinite* if it is neither positive semi-definite nor negative semi-definite.

Lemma 10.4 *Let $(c_{i,j})$ be a positive definite symmetric $n \times n$ matrix. Then there exists some $\varepsilon > 0$ with the following property: if all of the components of a symmetric $n \times n$ matrix $(b_{i,j})$ satisfy the inequality $|b_{i,j} - c_{i,j}| < \varepsilon$ then the matrix $(b_{i,j})$ is positive definite.*

Proof Let S^{n-1} be the unit $n - 1$ -sphere in \mathbb{R}^n defined by

$$S^{n-1} = \{(h_1, h_2, \dots, h_n) \in \mathbb{R}^n : h_1^2 + h_2^2 + \dots + h_n^2 = 1\}.$$

Observe that a symmetric $n \times n$ matrix $(b_{i,j})$ is positive definite if and only if

$$\sum_{i=1}^n \sum_{j=1}^n b_{i,j} h_i h_j > 0$$

for all $(h_1, h_2, \dots, h_n) \in S^{n-1}$. Now the matrix $(c_{i,j})$ is positive definite, by assumption. Therefore

$$\sum_{i=1}^n \sum_{j=1}^n c_{i,j} h_i h_j > 0$$

for all $(h_1, h_2, \dots, h_n) \in S^{n-1}$.

But S^{n-1} is a closed bounded set in \mathbb{R}^n , it therefore follows from Theorem 5.5 that there exists some $(k_1, k_2, \dots, k_n) \in S^{n-1}$ with the property that

$$\sum_{i=1}^n \sum_{j=1}^n c_{i,j} h_i h_j \geq \sum_{i=1}^n \sum_{j=1}^n c_{i,j} k_i k_j$$

for all $(h_1, h_2, \dots, h_n) \in S^{n-1}$. Let

$$A = \sum_{i=1}^n \sum_{j=1}^n c_{i,j} k_i k_j.$$

Then $A > 0$ and

$$\sum_{i=1}^n \sum_{j=1}^n c_{i,j} h_i h_j \geq A$$

for all $(h_1, h_2, \dots, h_n) \in S^{n-1}$. Set $\varepsilon = A/n^2$.

If $(b_{i,j})$ is a symmetric $n \times n$ matrix all of whose components satisfy $|b_{i,j} - c_{i,j}| < \varepsilon$ then

$$\left| \sum_{i=1}^n \sum_{j=1}^n (b_{i,j} - c_{i,j}) h_i h_j \right| < \varepsilon n^2 = A,$$

for all $(h_1, h_2, \dots, h_n) \in S^{n-1}$, hence

$$\sum_{i=1}^n \sum_{j=1}^n b_{i,j} h_i h_j > \sum_{i=1}^n \sum_{j=1}^n c_{i,j} h_i h_j - A \geq 0$$

for all $(h_1, h_2, \dots, h_n) \in S^{n-1}$. Thus the matrix $(b_{i,j})$ is positive-definite, as required. ■

Using the fact that a symmetric $n \times n$ matrix $(c_{i,j})$ is negative definite if and only if the matrix $(-c_{i,j})$ is positive-definite, we see that if $(c_{i,j})$ is a negative-definite matrix then there exists some $\varepsilon > 0$ with the following property: if all of the components of a symmetric $n \times n$ matrix $(b_{i,j})$ satisfy the inequality $|b_{i,j} - c_{i,j}| < \varepsilon$ then the matrix $(b_{i,j})$ is negative definite.

Let $f: X \rightarrow \mathbb{R}$ be a real-valued function whose partial derivatives of first and second order exist and are continuous throughout some open set X in \mathbb{R}^n . Let \mathbf{p} be a point of X . We have already observed that if the function f has a local maximum or a local minimum at \mathbf{p} then

$$\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\mathbf{p}} = 0 \quad (i = 1, 2, \dots, n).$$

We now study the behaviour of the function f around a point \mathbf{p} at which the first order partial derivatives vanish. We consider the Hessian matrix $(H_{i,j}(\mathbf{p}))$ defined by

$$H_{i,j}(\mathbf{p}) = \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{p}}.$$

Lemma 10.5 Let $f: X \rightarrow \mathbb{R}$ be a real-valued function whose partial derivatives of first and second order exist and are continuous throughout some open set X in \mathbb{R}^n , and let \mathbf{p} be a point of X at which

$$\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\mathbf{p}} = 0 \quad (i = 1, 2, \dots, n).$$

If f has a local minimum at a point \mathbf{p} of X then the Hessian matrix $(H_{i,j}(\mathbf{p}))$ at \mathbf{p} is positive semi-definite.

Proof The first order partial derivatives of f are zero at \mathbf{p} . It follows that, given any vector $\mathbf{h} \in \mathbb{R}^n$ which is sufficiently close to $\mathbf{0}$, there exists some θ satisfying $0 < \theta < 1$ (where θ depends on \mathbf{h}) such that

$$f(\mathbf{p} + \mathbf{h}) = f(\mathbf{p}) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n h_i h_j H_{i,j}(\mathbf{p} + \theta \mathbf{h}),$$

where

$$H_{i,j}(\mathbf{p} + \theta \mathbf{h}) = \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{p}+\theta \mathbf{h}}$$

(see Lemma 10.3).

It follows from this result that

$$\sum_{i=1}^n \sum_{j=1}^n h_i h_j H_{i,j}(\mathbf{p}) = \lim_{t \rightarrow 0} \frac{2(f(\mathbf{p} + t\mathbf{h}) - f(\mathbf{p}))}{t^2} \geq 0.$$

The result follows. ■

Let $f: X \rightarrow \mathbb{R}$ be a real-valued function whose partial derivatives of first and second order exist and are continuous throughout some open set X in \mathbb{R}^n , and let \mathbf{p} be a point at which the first order partial derivatives of f vanish. The above lemma shows that if the function f has a local minimum at \mathbf{h} then the Hessian matrix of f is positive semi-definite at \mathbf{p} . However the fact that the Hessian matrix of f is positive semi-definite at \mathbf{p} is not sufficient to ensure that f has a local minimum at \mathbf{p} , as the following example shows.

Example Consider the function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by $f(x, y) = x^2 - y^3$. Then the first order partial derivatives of f vanish at $(0, 0)$. The Hessian matrix of f at $(0, 0)$ is the matrix

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$$

and this matrix is positive semi-definite. However $(0, 0)$ is not a local minimum of f since $f(0, y) < f(0, 0)$ for all $y > 0$.

The following theorem shows that if the Hessian of the function f is positive definite at a point at which the first order partial derivatives of f vanish then f has a local minimum at that point.

Theorem 10.6 *Let $f: X \rightarrow \mathbb{R}$ be a real-valued function whose partial derivatives of first and second order exist and are continuous throughout some open set X in \mathbb{R}^n , and let \mathbf{p} be a point of X at which*

$$\left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\mathbf{p}} = 0 \quad (i = 1, 2, \dots, n).$$

Suppose that the Hessian matrix $(H_{i,j}(\mathbf{p}))$ at \mathbf{p} is positive definite. Then f has a local minimum at \mathbf{p} .

Proof The first order partial derivatives of f vanish at \mathbf{p} . It therefore follows from Taylor's Theorem that, for any $\mathbf{h} \in \mathbb{R}^n$ which is sufficiently close to $\mathbf{0}$, there exists some θ satisfying $0 < \theta < 1$ (where θ depends on \mathbf{h}) such that

$$f(\mathbf{p} + \mathbf{h}) = f(\mathbf{p}) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n h_i h_j H_{i,j}(\mathbf{p} + \theta \mathbf{h}),$$

where

$$H_{i,j}(\mathbf{p} + \theta \mathbf{h}) = \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{p}+\theta \mathbf{h}}$$

(see Lemma 10.3). Suppose that the Hessian matrix $(H_{i,j}(\mathbf{p}))$ is positive definite. It follows from Lemma 10.4 that there exists some $\varepsilon > 0$ such that if $|H_{i,j}(\mathbf{x}) - H_{i,j}(\mathbf{p})| < \varepsilon$ for all i and j then $(H_{i,j}(\mathbf{x}))$ is positive definite.

But it follows from the continuity of the second order partial derivatives of f that there exists some $\delta > 0$ such that $|H_{i,j}(\mathbf{x}) - H_{i,j}(\mathbf{p})| < \varepsilon$ whenever $|\mathbf{x} - \mathbf{p}| < \delta$. Thus if $|\mathbf{h}| < \delta$ then $(H_{i,j}(\mathbf{p} + \theta \mathbf{h}))$ is positive definite for all $\theta \in (0, 1)$ so that $f(\mathbf{p} + \mathbf{h}) > f(\mathbf{p})$. Thus \mathbf{p} is a local minimum of f . ■

A symmetric $n \times n$ matrix C is positive definite if and only if all its eigenvalues are strictly positive. In particular if $n = 2$ and if λ_1 and λ_2 are the eigenvalues of a symmetric 2×2 matrix C , then

$$\lambda_1 + \lambda_2 = \text{trace } C, \quad \lambda_1 \lambda_2 = \det C.$$

Thus a symmetric 2×2 matrix C is positive definite if and only if its trace and determinant are both positive.

Example Consider the function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

$$f(x, y) = 4x^2 + 3y^2 - 2xy - x^3 - x^2y - y^3.$$

Now

$$\left. \frac{\partial f(x, y)}{\partial x} \right|_{(x, y) = (0, 0)} = (0, 0), \quad \left. \frac{\partial f(x, y)}{\partial y} \right|_{(x, y) = (0, 0)} = (0, 0).$$

The Hessian matrix of f at $(0, 0)$ is

$$\begin{pmatrix} 8 & -2 \\ -2 & 6 \end{pmatrix}.$$

The trace and determinant of this matrix are 14 and 44 respectively. Hence this matrix is positive definite. We conclude from Theorem 10.6 that the function f has a local minimum at $(0, 0)$.