

Module MAU23203: Analysis in Several Real
Variables

Michaelmas Term 2021

Section 10: Second Order Partial Derivatives
and the Hessian Matrix

D. R. Wilkins

Copyright © Trinity College Dublin 2020–21

Contents

| | |
|---|-----------|
| 10 Second Order Partial Derivatives and the Hessian Matrix | 89 |
| 10.1 Second Order Partial Derivatives | 89 |
| 10.2 Local Maxima and Minima | 92 |

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

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

Now it would be incorrect to assert that if the second order partial derivatives of a real-valued function f of real variables x_1, x_2, \dots, x_n all exist at some point of the domain of the function then

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

are equal for all values of i and j . A standard counterexample is provided by the function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ that is defined so that

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

Calculations applying the basic definitions and the standard rules of differential calculus show that the second order partial derivatives of this function f at every point of its domain \mathbb{R}^2 . However the second order partial derivatives are not continuous at the point $(0, 0)$, and moreover

$$\frac{\partial^2 f}{\partial x \partial y} = 1 \quad \text{and} \quad \frac{\partial^2 f}{\partial y \partial x} = -1.$$

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

$$f_x(x, y) = \frac{\partial f}{\partial x}, \quad f_y(x, y) = \frac{\partial f}{\partial y},$$

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

and let (a, b) be a point of X . The set X is open in \mathbb{R}^2 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

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

for all real numbers h and k satisfying $|h| < L$ and $|k| < L$. First consider h to 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 $S(h, k) = q(b + k) - q(b)$. It then follows from the Mean Value Theorem (Theorem 7.5) 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

$$\begin{aligned} S(h, k) &= k(f_y(a + h, v) - f_y(a, v)) \\ &= hk f_{xy}(u, v) \\ &= hk \left. \frac{\partial^2 f}{\partial x \partial y} \right|_{(x, y) = (u, v)}. \end{aligned}$$

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. \blacksquare

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 Local Maxima and Minima

Definition A function $\varphi: X \rightarrow \mathbb{R}^p$, defined over an open set X in \mathbb{R}^n and mapping that open set into \mathbb{R}^p for some positive integers n and p , is said to be *k times continuously differentiable* if the partial derivatives of the components of the functions φ of all orders less than or equal to k exist and are continuous throughout the domain X of the function φ .

Let $f: X \rightarrow \mathbb{R}$ be a twice continuously differentiable real-valued function defined over some open subset X of \mathbb{R}^n . (In other words, let f be a real-valued function defined on an open set X in \mathbb{R}^n whose first and second order partial derivatives exist and are continuous throughout the domain X of the function f .) 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 $t = p_i$. It follows that 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 continuously 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.

Proposition 10.3 *Let f be a twice continuously differentiable real-valued function defined over an open ball in \mathbb{R}^n of radius δ centred on some point \mathbf{p} of \mathbb{R}^n . Then, given any vector \mathbf{h} in \mathbb{R}^n satisfying $|\mathbf{h}| < \delta$, there exists some real number θ satisfying $0 < \theta < 1$ for which*

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

Proof Let \mathbf{h} satisfy $|\mathbf{h}| < \delta$, and let $q(t) = f(\mathbf{p} + t\mathbf{h})$ for all real numbers t in some appropriately chosen open interval in the real line that contains the real numbers 0 and 1. The function q is the composition function in which the function f follows the function that sends real numbers t in the domain

of q to the point $\mathbf{p} + t\mathbf{h}$ of \mathbb{R}^n . It follows, on applying the Chain Rule for differentiable functions of several real variables (Theorem 8.20) that

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

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 7.10). Consequently

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

as required. ■

Let f be a twice continuously differentiable real-valued function defined over an open ball of radius δ about some given point \mathbf{p} of \mathbb{R}^n . It follows from Proposition 10.3 that if

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

for $j = 1, 2, \dots, n$, and if $|\mathbf{h}| < \delta$ then there exists some real number θ satisfying $0 < \theta < 1$ for which

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

Let f be a real-valued function defined over an open set in \mathbb{R}^n whose second order partial derivative are defined at a point \mathbf{p} of its domain. Let us denote by $(H_{i,j}(\mathbf{p}))$ the *Hessian matrix* at the point \mathbf{p} , defined by

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

Suppose now that the function f is twice continuously differentiable on its domain. Then $H_{i,j}(\mathbf{p}) = H_{j,i}(\mathbf{p})$ for all integers i and j between 1 and n , by Corollary 10.2, and 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 positive real number ε that is small enough to ensure that any symmetric $n \times n$ matrix $(b_{i,j})$ whose components all satisfy the inequality $|b_{i,j} - c_{i,j}| < \varepsilon$ 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 5.10 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 coefficients satisfy the inequality $|b_{i,j} - c_{i,j}| < \varepsilon$ 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 twice continuously differentiable real-valued function defined over some open set X in \mathbb{R}^n , and let \mathbf{p} be a point of the open set 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 twice continuously differentiable real-valued function defined over an open set X in \mathbb{R}^n , and let \mathbf{p} be a point of the open set 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 the point \mathbf{p} 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 Proposition 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 twice continuously differentiable real-valued function defined over some open set in \mathbb{R}^n , and let \mathbf{p} be a point of the domain of f at which the first order partial derivatives of f are zero. The above lemma shows that if the function f has a local minimum at \mathbf{p} 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 $f(x, y) = x^2 - y^3$. The first order partial derivatives of f are zero at $(0, 0)$. The Hessian matrix of f at $(0, 0)$ is the matrix

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

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

The following theorem shows that if the Hessian matrix 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 twice continuously differentiable real-valued function defined over 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 $(H_{i,j}(\mathbf{p}))$ of the function f at the point \mathbf{p} is positive definite. Then f has a local minimum at \mathbf{p} .

Proof The first order partial derivatives of f take the value 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 Proposition 10.3). Suppose that the Hessian matrix $(H_{i,j}(\mathbf{p}))$ is positive definite. Then there exists some positive real number ε small enough to

ensure 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 (see Lemma 10.4).

But it follows from the continuity of the second order partial derivatives of f that there exists some positive real number δ small enough to ensure that $\mathbf{x} \in X$ and $|H_{i,j}(\mathbf{x}) - H_{i,j}(\mathbf{p})| < \varepsilon$ for all integers i and j between 1 and n 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 the function 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

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

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 $(0, 0)$.