Module MA2321: Analysis in Several Real Variables
Michaelmas Term 2015
Section 7

D. R. Wilkins

Contents

7 Differentiation of Functions of Several Real Variables

7.1 Linear Transformations

The space $\mathbb{R}^n$ consisting of all n-tuples $(x_1, x_2, \ldots, x_n)$ of real numbers is a vector space over the field $\mathbb{R}$ of real numbers, where addition and multiplication by scalars are defined by

$$\begin{aligned} (x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) &= (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n), \\ t(x_1, x_2, \ldots, x_n) &= (tx_1, tx_2, \ldots, tx_n) \end{aligned}$$

for all $(x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ and $t \in \mathbb{R}$.

Definition A map $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is said to be a *linear transformation* if

$$T(\mathbf{x} + \mathbf{y}) = T\mathbf{x} + T\mathbf{y}, \qquad T(t\mathbf{x}) = tT\mathbf{x}$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $t \in \mathbb{R}$.

Every linear transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is represented by an $m \times n$ matrix $(T_{i,j})$. Indeed let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ be the standard basis vectors of $\mathbb{R}^n$ defined by

$$\mathbf{e}_1 = (1, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, \ldots, 0), \ldots, \mathbf{e}_n = (0, 0, \ldots, 1).$$

Thus if $\mathbf{x} \in \mathbb{R}^n$ is represented by the n-tuple $(x_1, x_2, \ldots, x_n)$ then

$$\mathbf{x} = \sum_{j=1}^{n} x_j \mathbf{e}_j.$$

Similarly let $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_m$ be the standard basis vectors of $\mathbb{R}^m$ defined by

$$\mathbf{f}_1 = (1, 0, \ldots, 0), \quad \mathbf{f}_2 = (0, 1, \ldots, 0), \ldots, \mathbf{f}_m = (0, 0, \ldots, 1).$$

Thus if $\mathbf{v} \in \mathbb{R}^m$ is represented by the n-tuple $(v_1, v_2, \ldots, v_m)$ then

$$\mathbf{v} = \sum_{i=1}^{m} v_i \mathbf{f}_i.$$

Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation. Define $T_{i,j}$ for all integers i between 1 and m and for all integers j between 1 and n such that

$$T\mathbf{e}_j = \sum_{i=1}^{m} T_{i,j} \mathbf{f}_i.$$

Using the linearity of T, we see that if $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ then

$$T\mathbf{x} = T\left(\sum_{j=1}^{n} x_j \mathbf{e}_j\right) = \sum_{j=1}^{n} (x_j\, T\mathbf{e}_j) = \sum_{i=1}^{m}\left(\sum_{j=1}^{n} T_{i,j} x_j\right)\mathbf{f}_i.$$

Thus the ith component of $T\mathbf{x}$ is

$$T_{i1}x_1 + T_{i2}x_2 + \cdots + T_{in}x_n.$$

Writing out this identity in matrix notation, we see that if $T\mathbf{x} = \mathbf{v}$, where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \qquad \mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix},$$

then

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} & \ldots & T_{1n} \\ T_{21} & T_{22} & \ldots & T_{2n} \\ \vdots & \vdots & & \vdots \\ T_{m1} & T_{m2} & \ldots & T_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Recall that the *length* (or *norm*) of an element $\mathbf{x} \in \mathbb{R}^n$ is defined such that

$$|\mathbf{x}|^2 = x_1^2 + x_2^2 + \cdots + x_n^2.$$

Definition Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, and let $(T_{i,j})$ be the $m \times n$ matrix representing this linear transformation with respect to the standard bases of $\mathbb{R}^m$ and $\mathbb{R}^m$. The *Hilbert-Schmidt norm* $\|T\|_{\mathrm{HS}}$ of the linear transformation is then defined so that

$$\|T\|_{\mathrm{HS}} = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} T_{i,j}^2}.$$

Note that the Hilbert-Schmidt norm is just the Euclidean norm on the real vector space of dimension mn whose elements are $m \times n$ matrices representing linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$ with respect to the standard bases of these vector spaces. Therefore it has the standard properties of the Euclidean norm. In particular it follows from the Triangle Inequality (Lemma 6.2) that

$$\|T + U\|_{\mathrm{HS}} \leq \|T\|_{\mathrm{HS}} + \|U\|_{\mathrm{HS}} \quad \text{and} \quad \|sT\|_{\mathrm{HS}} = |s|\, \|T\|_{\mathrm{HS}}$$

for all linear transformations T and U from $\mathbb{R}^n$ to $\mathbb{R}^m$ and for all real numbers s.

Lemma 7.1 *Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Then T is uniformly continuous on $\mathbb{R}^n$. Moreover*

$$|T\mathbf{x} - T\mathbf{y}| \leq \|T\|_{\mathrm{HS}} |\mathbf{x} - \mathbf{y}|$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where $\|T\|_{HS}$ is the Hilbert-Schmidt norm of the linear transformation T.

Proof Let $\mathbf{v} = T\mathbf{x} - T\mathbf{y}$, where $\mathbf{v} \in \mathbb{R}^m$ is represented by the m-tuple $(v_1, v_2, \ldots, v_m)$. Then

$$v_i = T_{i1}(x_1 - y_1) + T_{i2}(x_2 - y_2) + \cdots + T_{in}(x_n - y_n)$$

for all integers i between 1 and m. It follows from Schwarz' Inequality (Lemma 6.1) that

$$v_i^2 \leq \left(\sum_{j=1}^n T_{i,j}^2 \right) \left(\sum_{j=1}^n (x_j - y_j)^2 \right) = \left(\sum_{j=1}^n T_{i,j}^2 \right) |\mathbf{x} - \mathbf{y}|^2.$$

Hence

$$|\mathbf{v}|^2 = \sum_{i=1}^m v_i^2 \leq \left(\sum_{i=1}^m \sum_{j=1}^n T_{i,j}^2 \right) |\mathbf{x} - \mathbf{y}|^2 = \|T\|_{\mathrm{HS}} |\mathbf{x} - \mathbf{y}|^2.$$

Thus $|T\mathbf{x} - T\mathbf{y}| \leq \|T\|_{\mathrm{HS}} |\mathbf{x} - \mathbf{y}|$. It follows from this that T is uniformly continuous. Indeed let some positive real number ε be given. We can then choose δ so that $\|T\|_{\mathrm{HS}}\, \delta < \varepsilon$. If $\mathbf{x}$ and $\mathbf{y}$ are elements of $\mathbb{R}^n$ which satisfy the condition $|\mathbf{x} - \mathbf{y}| < \delta$ then $|T\mathbf{x} - T\mathbf{y}| < \varepsilon$. This shows that $T: \mathbb{R}^n \to \mathbb{R}^m$ is uniformly continuous on $\mathbb{R}^n$, as required. ∎

Lemma 7.2 *Let $T: \mathbb{R}^n \to \mathbf{R}^m$ be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and let $S: \mathbf{R}^m \to \mathbf{R}^p$ be a linear transformation from $\mathbf{R}^m$ to $\mathbf{R}^p$. Then the Hilbert-Schmidt norm of the composition of the linear operators T and S satisfies the inequality $\|ST\|_{\mathrm{HS}} \leq \|S\|_{\mathrm{HS}}\, \|T\|_{\mathrm{HS}}$.*

Proof The composition ST of the linear operators is represented by the product of the corresponding matrices. Thus the component $(ST)_{k,j}$ in the kth row and the jth column of the $p \times n$ matrix representing the linear transformation ST satisfies

$$(ST)_{k,j} = \sum_{i=1}^m S_{k,i} T_{i,j}.$$

It follows from Schwarz' Inequality (Lemma 6.1) that

$$(ST)_{k,j}^2 \leq \left(\sum_{i=1}^m S_{k,i}^2 \right) \left(\sum_{i=1}^m T_{i,j}^2 \right).$$

Summing over k, we find that

$$\sum_{k=1}^p (ST)_{k,j}^2 \leq \left(\sum_{k=1}^p \sum_{i=1}^m S_{k,i}^2 \right) \left(\sum_{i=1}^m T_{i,j}^2 \right) = \|S\|_{\mathrm{HS}}^2 \left(\sum_{i=1}^m T_{i,j}^2 \right).$$

Then summing over j, we find that

$$\|ST\|_{\mathrm{HS}}^2 = \sum_{k=1}^p \sum_{j=1}^n (ST)_{k,j}^2 \leq \|S\|_{\mathrm{HS}}^2 \left(\sum_{i=1}^m \sum_{j=1}^n T_{i,j}^2 \right) \leq \|S\|_{\mathrm{HS}}\|^2 \|T\|_{\mathrm{HS}}\|^2.$$

On taking square roots, we find that $\|ST\|_{\mathrm{HS}} \leq \|S\|_{\mathrm{HS}} \, \|T\|_{\mathrm{HS}}$, as required. ∎

7.2 Review of Differentiability for Functions of One Real Variable

Let $f\colon I \to \mathbb{R}$ be a real-valued function defined on some open interval I in $\mathbb{R}$. Let a be an element of I. Recall that the function f is *differentiable* at a if and only if

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

exists, and the value of this limit (if it exists) is known as the *derivative* of f at a (denoted by $f'(a)$).

We wish to define the notion of differentiability for functions of more than one variable. However we cannot immediately generalize the above definition as it stands (because this would require us to divide one element in $\mathbb{R}^n$ by another, which we cannot do since the operation of division is not defined on $\mathbb{R}^n$). We shall therefore reformulate the above definition of differentiability for functions of one real variable, exhibiting a criterion which is equivalent to the definition of differentiability given above and which can be easily generalized to functions of more than one real variable. This criterion is provided by the following lemma.

Lemma 7.3 *Let $f\colon I \to \mathbb{R}$ be a real-valued function defined on some open interval I in $\mathbb{R}$. Let a be an element of I. The function f is differentiable at a with derivative $f'(a)$ (where $f'(a)$ is some real number) if and only if*

$$\lim_{h \to 0} \frac{1}{|h|} \left(f(a+h) - f(a) - f'(a)h \right) = 0.$$

Proof It follows directly from the definition of the limit of a function that

$$\lim_{h\to 0} \frac{f(a+h)-f(a)}{h} = f'(a)$$

if and only if

$$\lim_{h\to 0} \left| \frac{f(a+h)-f(a)}{h} - f'(a) \right| = 0.$$

But

$$\left| \frac{f(a+h)-f(a)}{h} - f'(a) \right| = \left| \frac{1}{|h|} \left(f(a+h) - f(a) - f'(a)h\right) \right|.$$

It follows immediately from this that the function f is differentiable at a with derivative $f'(a)$ if and only if

$$\lim_{h\to 0} \frac{1}{|h|} \left(f(a+h) - f(a) - f'(a)h\right) = 0. \quad \blacksquare$$

Now let us observe that, for any real number c, the map $h \mapsto ch$ defines a linear transformation from $\mathbb{R}$ to $\mathbb{R}$. Conversely, every linear transformation from $\mathbb{R}$ to $\mathbb{R}$ is of the form $h \mapsto ch$ for some $c \in \mathbb{R}$. Because of this, we may regard the derivative $f'(a)$ of f at a as representing a linear transformation $h \mapsto f'(a)h$, characterized by the property that the map

$$x \mapsto f(a) + f'(a)(x-a)$$

provides a 'good' approximation to f around a in the sense that

$$\lim_{h\to 0} \frac{e(a,h)}{|h|} = 0,$$

where

$$e(a,h) = f(a+h) - f(a) - f'(a)h$$

(i.e., e(a,h) measures the difference between $f(a+h)$ and the value $f(a) + f'(a)h$ of the approximation at $a+h$, and thus provides a measure of the error of this approximation). We shall generalize the notion of differentiability to functions f from $\mathbb{R}^n$ to $\mathbb{R}^m$ by defining the derivative $(Df)_{\mathbf{p}}$ of f at $\mathbf{p}$ to be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ characterized by the property that the map

$$\mathbf{x} \mapsto f(\mathbf{p}) + (Df)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})$$

provides a 'good' approximation to f around $\mathbf{p}$.

7.3 Derivatives of Functions of Several Variables

Definition Let V be an open subset of $\mathbb{R}^n$ and let $\varphi: V \to \mathbb{R}^m$ be a map from V into $\mathbb{R}^m$. Let $\mathbf{p}$ be a point of V. The function φ is said to be *differentiable* at $\mathbf{p}$, with *derivative* $T: \mathbb{R}^n \to \mathbb{R}^m$ if and only if there exists a linear transformation $T: R^n \to R^m$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ with the property that

$$\lim_{\mathbf{h} \to \mathbf{0}} \frac{1}{|\mathbf{h}|} (\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - T\mathbf{h}) = \mathbf{0}.$$

If φ is differentiable at $\mathbf{p}$ then the derivative $T: \mathbb{R}^n \to \mathbb{R}^m$ of φ at $\mathbf{p}$ may be denoted by $(D\varphi)_{\mathbf{p}}$, or by $(D\varphi)(\mathbf{p})$, or by $f'(\mathbf{p})$.

The derivative $(D\varphi)_{\mathbf{p}}$ of φ at $\mathbf{p}$ is sometimes referred to as the *total derivative* of φ at $\mathbf{p}$. If φ is differentiable at every point of V then we say that φ is differentiable on V.

Lemma 7.4 *Let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$. Then T is differentiable at each point $\mathbf{p}$ of $\mathbb{R}^n$, and $(DT)_{\mathbf{p}} = T$.*

Proof This follows immediately from the identity $T(\mathbf{p} + \mathbf{h}) - T\mathbf{p} - T\mathbf{h} = \mathbf{0}$. ∎

Lemma 7.5 *Let V be an open subset of $\mathbb{R}^n$, let $\varphi: V \to \mathbb{R}^m$ be a map from V into $\mathbb{R}^m$, let $T: \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let $\mathbf{p}$ be a point of V. Then φ is differentiable at $\mathbf{p}$, with derivative T, if and only if, given any strictly positive real number ε, there exists some strictly positive real number δ such that $\mathbf{p} + \mathbf{h} \in V$ and*

$$|\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - T\mathbf{h}| \leq \varepsilon |\mathbf{h}|$$

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

Proof Suppose that the function $\varphi: \mathbb{R}^n \to \mathbb{R}^m$ satisfies the criterion described in the statement of the lemma. Let some strictly positive real number ε be given. Take some real number ε' satisfying $0 < \varepsilon' < \varepsilon$. Then there exists some strictly positive real number δ such that $\mathbf{p} + \mathbf{h} \in V$ and

$$|\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - T\,\mathbf{h}| \leq \varepsilon' |\mathbf{h}|$$

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

$$\frac{1}{|\mathbf{h}|} (\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - T\mathbf{h}) \leq \varepsilon' < \varepsilon$$

whenever $0 < |\mathbf{h}| < \delta$, and therefore

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} (\varphi(\mathbf{p}+\mathbf{h}) - \varphi(\mathbf{p}) - T\mathbf{h}) = \mathbf{0}.$$

It then follows that the function φ is differentiable at $\mathbf{p}$, with derivative T.

Conversely, the function φ is differentiable at $\mathbf{p}$, with derivative T, then

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} (\varphi(\mathbf{p}+\mathbf{h}) - \varphi(\mathbf{p}) - T\mathbf{h}) = \mathbf{0}$$

then it follows from the definition of limits that, given any strictly positive real number ε, there exists some strictly positive real number δ such that the condition set out in the statement of the lemma is satisfied, as required. ∎

It follows from Lemma 7.5 that if a function $\varphi\colon V \to \mathbb{R}^m$ defined over an open set V in $\mathbb{R}^n$ is differentiable at a point $\mathbf{p}$ of V, then, given any positive real number ε there exists a positive real number δ such that

$$|\varphi(\mathbf{p}+\mathbf{h}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}\,\mathbf{h}| \le \varepsilon|\mathbf{h}|$$

for all $\mathbf{h} \in \mathbb{R}^n$ satisfying $|\mathbf{h}| < \delta$, where $(D\varphi)_{\mathbf{p}}\colon \mathbf{R}^n \to \mathbb{R}^m$ denotes the derivative of φ at the point $\mathbf{p}$. In that case

$$\varphi(\mathbf{p}+\mathbf{h}) = \varphi(\mathbf{p}) + (D\varphi)_{\mathbf{p}}\,\mathbf{h} + e(\mathbf{p},\mathbf{h}),$$

where

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{e(\mathbf{p},\mathbf{h})}{|\mathbf{h}|} = \mathbf{0}.$$

Thus if φ is differentiable at $\mathbf{p}$ then the map $\lambda\colon V \to \mathbb{R}$ defined by

$$\lambda(\mathbf{x}) = \varphi(\mathbf{p}) + (D\varphi)_{\mathbf{p}}\,(\mathbf{x}-\mathbf{p})$$

provides a good approximation to the function around $\mathbf{p}$. The difference between $\varphi(\mathbf{x})$ and $\lambda(\mathbf{x})$ is equal to $e(\mathbf{p},\mathbf{x}-\mathbf{p})$, and this quantity tends to $\mathbf{0}$ faster than $|\mathbf{x}-\mathbf{p}|$ as $\mathbf{x}$ tends to $\mathbf{p}$.

Example Let $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ be defined so that

$$\varphi\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) = \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix}$$

for all real numbers x and y. Let p, q, h and k be real numbers. Then

$$\begin{aligned}
\varphi\left(\begin{pmatrix} p+h \\ q+k \end{pmatrix}\right) &= \begin{pmatrix} (p+h)^2-(q+k)^2 \\ 2(p+h)(q+k) \end{pmatrix} \\
&= \begin{pmatrix} p^2-q^2+2(ph-qk)+h^2-k^2 \\ 2pq+2(qh+pk)+2hk \end{pmatrix} \\
&= \begin{pmatrix} p^2-q^2 \\ 2pq \end{pmatrix} + \begin{pmatrix} 2(ph-qk) \\ 2(qh+pk) \end{pmatrix} + \begin{pmatrix} h^2-k^2 \\ 2hk \end{pmatrix} \\
&= \varphi\left(\begin{pmatrix} p \\ q \end{pmatrix}\right) + \begin{pmatrix} 2p & -2q \\ 2q & 2p \end{pmatrix}\begin{pmatrix} h \\ k \end{pmatrix} + \begin{pmatrix} h^2-k^2 \\ 2hk \end{pmatrix}.
\end{aligned}$$

Now $|(h,k)| = \sqrt{h^2+k^2}$, and

$$\frac{1}{h^2+k^2}\left|\begin{pmatrix} h^2-k^2 \\ 2hk \end{pmatrix}\right|^2 = \frac{(h^2-k^2)^2+4h^2k^2}{h^2+k^2}$$

for $(h,k) \neq (0,0)$. Note that if h and k are both multiplied by some positive real number t then the right hand side of the above equality is multiplied by t^2. It follows that if K is the maximum value of the right hand side of this equality on the circle $\{(h,k) : h^2+k^2 = 1\}$ then

$$\frac{1}{h^2+k^2}\left|\begin{pmatrix} h^2-k^2 \\ 2hk \end{pmatrix}\right|^2 \leq K(h^2+k^2).$$

Therefore

$$\frac{1}{\sqrt{h^2+k^2}}\left|\begin{pmatrix} h^2-k^2 \\ 2hk \end{pmatrix}\right| \to 0 \text{ as } (h,k) \to (0,0).$$

It follows that the function $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ is differentiable, and the derivative $(D\varphi)_{(p,q)}$ of this function at the point (p,q) is the linear transformation represented as a matrix with respect to the standard bases as follows:

$$(D\varphi)_{(p,q)} = \begin{pmatrix} 2p & -2q \\ 2q & 2p \end{pmatrix}.$$

Example Let $M_n(\mathbb{R})$ denote the real vector space consisting of all $n \times n$ matrices with real coefficients. $M_n(\mathbb{R})$ may be regarded as a Euclidean space, where the Euclidean distance between two $n \times n$ matrices A and B is the Hilbert-Schmidt norm of $\|A-B\|_{\mathrm{HS}}$ of $A-B$, defined such that

$$\|A-B\|_{\mathrm{HS}} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(A_{i,j}-B_{i,j})^2}.$$

Let $\mathrm{GL}_(n, \mathbb{R})$ denote the set of invertible $n \times n$ matrices with real coefficients. Then

$$\mathrm{GL}(n, \mathbb{R}) = \{A \in M_n(\mathbb{R}) : \det A \neq 0\}.$$

Now the determinant $\det A$ of a square $n \times n$ matrix A is a continuous function of the coefficients of the matrix. It follows from this that $\mathrm{GL}(n, \mathbb{R})$ is an open subset of $M_n(\mathbb{R})$. We denote the identity $n \times n$ matrix by I. Then $\|I\|_{\mathrm{HS}} = \sqrt{n}$, because the square of the Hilbert-Schmidt norm $\|I\|_{\mathrm{HS}}$ is the sum of the squares of the components of the identity matrix, and is therefore equal to n.

Let $\varphi\colon \mathrm{GL}(n, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R})$ be the function defined so that $\varphi(A) = A^{-1}$ for all invertible $n \times n$ matrices A. We show that this function $\varphi\colon \mathrm{GL}(n, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R})$ is differentiable.

Let A be an invertible $n \times n$ matrix. Then for all $n \times n$ matrices H. Now the matrix $I + A^{-1}H$ is invertible if and only if $\det(I + A^{-1}H) \neq 0$. Moreover this determinant is a continuous function of the coefficients of the matrix H. It follows that there exists some positive number δ_0 such that $I + A^{-1}H$ is invertible whenever $\|H\|_{\mathrm{HS}} < \delta_0$. Moreover the function mapping the matrix H to $\|(I + A^{-1}H)^{-1}\|_{\mathrm{HS}}$ is continuous and takes the value $\sqrt{n}$ when H is the zero matrix. We can therefore choose a positive number δ_0 small enough to ensure that $I + A^{-1}H$ is invertible and $\|(I + A^{-1}H)^{-1}\|_{\mathrm{HS}} < 2\sqrt{n}$ whenever $\|H\|_{\mathrm{HS}} < \delta_0$.

Let the $n \times n$ matrix H satisfy $\|H\|_{\mathrm{HS}} < \delta_0$. Then

$$(I - A^{-1}H)(I + A^{-1}H) = I - A^{-1}HA^{-1}H,$$

and therefore

$$I = (I - A^{-1}H)(I + A^{-1}H) + A^{-1}HA^{-1}H.$$

Multiplying this identity on the right by the matrix $(I + A^{-1}H)^{-1}$, we find that

$$(I + A^{-1}H)^{-1} = I - A^{-1}H + A^{-1}HA^{-1}H(I + A^{-1}H)^{-1}.$$

It follows that

$$\begin{aligned}(A + H)^{-1} &= (A(I + A^{-1}H))^{-1} = (I + A^{-1}H)^{-1}A^{-1} \\ &= A^{-1} - A^{-1}HA^{-1} + A^{-1}HA^{-1}H(I + A^{-1}H)^{-1}A^{-1}.\end{aligned}$$

The Hilbert-Schmidt norm of a product of $n \times n$ matrices is bounded above by the product of the Hilbert-Schmidt norms of those matrices. Therefore if $\|H\|_{\mathrm{HS}} < \delta_0$ then

$$\|A^{-1}HA^{-1}H(I + A^{-1}H)^{-1}A^{-1}\|_{\mathrm{HS}} \leq \|A^{-1}\|_{\mathrm{HS}}^3 \|(I + A^{-1}H)^{-1}\|_{\mathrm{HS}} \|H\|_{\mathrm{HS}}^2,$$

where $\|(I + A^{-1}H)^{-1}\|_{\mathrm{HS}} < 2\sqrt{n}$, and therefore

$$\left\|(A + H)^{-1} - A^{-1} + A^{-1}HA^{-1}\right\|_{\mathrm{HS}} \leq 2\sqrt{n}\|A^{-1}\|_{\mathrm{HS}}^{3}\|H\|_{\mathrm{HS}}^{2}.$$

It follows that

$$\lim_{H\to 0} \frac{1}{\|H\|_{\mathrm{HS}}} \left\|(A + H)^{-1} - A^{-1} + A^{-1}HA^{-1}\right\|_{\mathrm{HS}} = 0.$$

Therefore the function $\varphi\colon \mathrm{GL}(n,\mathbb{R}) \to \mathrm{GL}(n,\mathbb{R})$ is differentiable, where $\varphi(A) = A^{-1}$ for all invertible $n \times n$ matrices A with real coefficients, and moreover

$$(D\varphi)_A(H) = -A^{-1}HA^{-1}.$$

Lemma 7.6 *Let $\varphi\colon V \to \mathbb{R}^m$ be a function which maps an open subset V of $\mathbb{R}^n$ into $\mathbb{R}^m$ which is differentiable at some point $\mathbf{p}$ of V. Then φ is continuous at $\mathbf{p}$.*

Proof If we define

$$e(\mathbf{p},\mathbf{h}) = \varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}\,\mathbf{h}$$

then

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{e(\mathbf{p},\mathbf{h})}{|\mathbf{h}|} = \mathbf{0}$$

(because φ is differentiable at $\mathbf{p}$), and hence

$$\lim_{\mathbf{h}\to\mathbf{0}} e(\mathbf{p},\mathbf{h}) = \left(\lim_{\mathbf{h}\to\mathbf{0}} |\mathbf{h}|\right)\left(\lim_{\mathbf{h}\to\mathbf{0}} \frac{e(\mathbf{p},\mathbf{h})}{|\mathbf{h}|}\right) = \mathbf{0}.$$

But

$$\lim_{\mathbf{h}\to\mathbf{0}} e(\mathbf{p},\mathbf{h}) = \lim_{\mathbf{h}\to\mathbf{0}} \varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}),$$

since

$$\lim_{\mathbf{h}\to\mathbf{0}} (D\varphi)_{\mathbf{p}}\,\mathbf{h} = (D\varphi)_{\mathbf{p}}\left(\lim_{\mathbf{h}\to\mathbf{0}} \mathbf{h}\right) = \mathbf{0}$$

(on account of the fact that every linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ is continuous). We conclude therefore that

$$\lim_{\mathbf{h}\to\mathbf{0}} \varphi(\mathbf{p} + \mathbf{h}) = \varphi(\mathbf{p}),$$

showing that φ is continuous at $\mathbf{p}$. ∎

Lemma 7.7 *Let $\varphi: V \to \mathbb{R}^m$ be a function which maps an open subset V of $\mathbb{R}^n$ into $\mathbb{R}^m$ which is differentiable at some point $\mathbf{p}$ of V. Let $(D\varphi)_{\mathbf{p}}: \mathbb{R}^n \to \mathbb{R}^m$ be the derivative of φ at $\mathbf{p}$. Let $\mathbf{u}$ be an element of $\mathbb{R}^n$. Then*

$$(D\varphi)_{\mathbf{p}}\mathbf{u} = \lim_{t\to 0} \frac{1}{t}\left(\varphi(\mathbf{p} + t\mathbf{u}) - \varphi(\mathbf{p})\right).$$

Thus the derivative $(D\varphi)_{\mathbf{p}}$ of φ at $\mathbf{p}$ is uniquely determined by the map φ.

Proof It follows from the differentiability of φ at $\mathbf{p}$ that

$$\lim_{\mathbf{h}\to \mathbf{0}} \frac{1}{|\mathbf{h}|}\left(\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}\,\mathbf{h}\right) = \mathbf{0}.$$

In particular, if we set $\mathbf{h} = t\mathbf{u}$, and $\mathbf{h} = -t\mathbf{u}$, where t is a real variable, we can conclude that

$$\lim_{t\to 0^+} \frac{1}{t}\left(\varphi(\mathbf{p} + t\mathbf{u}) - \varphi(\mathbf{p}) - t(D\varphi)_{\mathbf{p}}\mathbf{u}\right) = \mathbf{0},$$

$$\lim_{t\to 0^-} \frac{1}{t}\left(\varphi(\mathbf{p} + t\mathbf{u}) - \varphi(\mathbf{p}) - t(D\varphi)_{\mathbf{p}}\mathbf{u}\right) = \mathbf{0},$$

It follows that

$$\lim_{t\to 0} \frac{1}{t}\left(\varphi(\mathbf{p} + t\mathbf{u}) - \varphi(\mathbf{p}) - t(D\varphi)_{\mathbf{p}}\mathbf{u}\right) = \mathbf{0},$$

as required. ∎

We now show that given two differentiable functions mapping V into $\mathbb{R}$, where V is an open set in $\mathbb{R}^n$, the sum, difference and product of these functions are also differentiable.

Theorem 7.8 *Let V be an open set in $\mathbb{R}^n$, and let $f: V \to \mathbb{R}$ and $g: V \to \mathbb{R}$ be functions mapping V into $\mathbb{R}$. Let $\mathbf{p}$ be a point of V. Suppose that f and g are differentiable at $\mathbf{p}$. Then the functions $f + g$, $f - g$ and $f.g$ are differentiable at $\mathbf{p}$, and*

$$\begin{aligned} (D(f+g)_{\mathbf{p}} &= (Df)_{\mathbf{p}} + (Dg)_{\mathbf{p}}, \\ D(f-g)_{\mathbf{p}} &= (Df)_{\mathbf{p}} - (Dg)_{\mathbf{p}}, \\ D(f.g)_{\mathbf{p}} &= g(\mathbf{p})(Df)_{\mathbf{p}} + f(\mathbf{p})(Dg)_{\mathbf{p}}. \end{aligned}$$

Proof We can write

$$\begin{aligned} f(\mathbf{p}+\mathbf{h}) &= f(\mathbf{p}) + (Df)_{\mathbf{p}}\,\mathbf{h} + e_1(\mathbf{p},\mathbf{h}), \\ g(\mathbf{p}+\mathbf{h}) &= g(\mathbf{p}) + (Dg)_{\mathbf{p}}\,\mathbf{h} + e_2(\mathbf{p},\mathbf{h}), \end{aligned}$$

for all sufficiently small h, where

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{e_1(\mathbf{p},\mathbf{h})}{|\mathbf{h}|} = \mathbf{0}, \qquad \lim_{\mathbf{h}\to\mathbf{0}} \frac{e_2(\mathbf{p},\mathbf{h})}{|\mathbf{h}|} = \mathbf{0},$$

on account of the fact that f and g are differentiable at $\mathbf{p}$. Then

$$\begin{aligned} &\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} |f(\mathbf{p}+\mathbf{h}) + g(\mathbf{p}+\mathbf{h}) - (f(\mathbf{p}) + g(\mathbf{p})) - ((Df)_{\mathbf{p}} + (Dg)_{\mathbf{p}})\,\mathbf{h}| \\ &\qquad = \lim_{\mathbf{h}\to\mathbf{0}} \frac{e_1(\mathbf{p},\mathbf{h}) + e_2(\mathbf{p},\mathbf{h})}{|\mathbf{h}|} = 0, \\ &\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} |f(\mathbf{p}+\mathbf{h}) - g(\mathbf{p}+\mathbf{h}) - (f(\mathbf{p}) - g(\mathbf{p})) - ((Df)_{\mathbf{p}} - (Dg)_{\mathbf{p}})\,\mathbf{h}| \\ &\qquad = \lim_{\mathbf{h}\to\mathbf{0}} \frac{e_1(\mathbf{p},\mathbf{h}) - e_2(\mathbf{p},\mathbf{h})}{|\mathbf{h}|} = 0. \end{aligned}$$

Thus $f+g$ and $f-g$ are differentiable at $\mathbf{p}$. Also

$$f(\mathbf{p}+\mathbf{h})g(\mathbf{p}+\mathbf{h}) = f(\mathbf{p})g(\mathbf{p}) + g(\mathbf{p})(Df)_{\mathbf{p}}\,\mathbf{h} + f(\mathbf{p})(Dg)_{\mathbf{p}}\,\mathbf{h} + e(\mathbf{p},\mathbf{h}),$$

where

$$\begin{aligned} e(\mathbf{p},\mathbf{h}) &= (f(\mathbf{p}) + (Df)_{\mathbf{p}}\,\mathbf{h})e_2(\mathbf{p},\mathbf{h}) + (g(\mathbf{p}) + (Dg)_{\mathbf{p}}\,\mathbf{h})e_1(\mathbf{p},\mathbf{h}) \\ &\quad + ((Df)_{\mathbf{p}}\,\mathbf{h})((Dg)_{\mathbf{p}}\,\mathbf{h}) + e_1(\mathbf{p},\mathbf{h})e_2(\mathbf{p},\mathbf{h}). \end{aligned}$$

It follows from Lemma 7.1 that there exist constants M_1 and M_2 such that

$$|(Df)_{\mathbf{p}}\,\mathbf{h}| \le M_1|\mathbf{h}|, \qquad |(Dg)_{\mathbf{p}}\,\mathbf{h}| \le M_2|\mathbf{h}|.$$

Therefore

$$|((Df)_{\mathbf{p}}\,\mathbf{h})((Dg)_{\mathbf{p}}\,\mathbf{h})| \le M_1 M_2 |\mathbf{h}|^2,$$

so that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|}((Df)_{\mathbf{p}}\,\mathbf{h})((Dg)_{\mathbf{p}}\,\mathbf{h}) = 0.$$

Also

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|}((f(\mathbf{p}) + (Df)_{\mathbf{p}}\,\mathbf{h})e_2(\mathbf{p},\mathbf{h}))$$

$$
\begin{aligned}
&= \lim_{\mathbf{h}\to\mathbf{0}} (f(\mathbf{p}) + (Df)_{\mathbf{p}}\,\mathbf{h}) \lim_{\mathbf{h}\to\mathbf{0}} \frac{e_2(\mathbf{p},\mathbf{h})}{|\mathbf{h}|} = 0, \\
\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|}((g(\mathbf{p}) + (Dg)_{\mathbf{p}}\,\mathbf{h})e_1(\mathbf{p},\mathbf{h})) & \\
&= \lim_{\mathbf{h}\to\mathbf{0}} (g(\mathbf{p}) + (Dg)_{\mathbf{p}}\,\mathbf{h}) \lim_{\mathbf{h}\to\mathbf{0}} \frac{e_1(\mathbf{p},\mathbf{h})}{|\mathbf{h}|} = 0, \\
\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|}(e_1(\mathbf{p},\mathbf{h})e_2(\mathbf{p},\mathbf{h})) &= \lim_{\mathbf{h}\to\mathbf{0}} e_1(\mathbf{p},\mathbf{h}) \lim_{\mathbf{h}\to\mathbf{0}} \frac{e_2(\mathbf{p},\mathbf{h})}{|\mathbf{h}|} = 0.
\end{aligned}
$$

Therefore

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{e(\mathbf{p},\mathbf{h})}{|\mathbf{h}|} = 0,$$

showing that the function $f.g$ is differentiable at $\mathbf{p}$ and that

$$D(f.g)_{\mathbf{p}} = g(\mathbf{p})(Df)_{\mathbf{p}} + f(\mathbf{p})(Dg)_{\mathbf{p}}. \quad \blacksquare$$

Theorem 7.9 (Chain Rule) *Let V be an open set in $\mathbb{R}^n$, and let $\varphi\colon V \to \mathbb{R}^m$ be a function mapping V into $\mathbb{R}^m$. Let W be an open set in $\mathbb{R}^m$ which contains $\varphi(V)$, and let $\psi\colon W \to \mathbb{R}^l$ be a function mapping W into $\mathbb{R}^l$. Let $\mathbf{p}$ be a point of V. Suppose that φ is differentiable at $\mathbf{p}$ and that ψ is differentiable at $\varphi(\mathbf{p})$. Then the composition $\psi \circ \varphi\colon \mathbb{R}^n \to \mathbb{R}^l$ (i.e., φ followed by ψ) is differentiable at $\mathbf{p}$. Moreover*

$$D(\psi \circ \varphi)_{\mathbf{p}} = (D\psi)_{\varphi(\mathbf{p})} \circ (D\varphi)_{\mathbf{p}}.$$

Thus the derivative of the composition $\psi\circ\varphi$ of the functions at the given point is the composition of the derivatives of those functions at the appropriate points.

Proof Let $\mathbf{q} = \varphi(\mathbf{p})$. First we note that there exist positive real numbers L, and M such that $|(D\varphi)_{\mathbf{p}}\mathbf{h}| \leq L|\mathbf{h}|$ for all $\mathbf{h} \in \mathbb{R}^m$ and $|(D\psi)_{\mathbf{q}}\mathbf{k}| \leq M|\mathbf{k}|$ for all $\mathbf{k} \in \mathbb{R}^n$. Indeed it follows from Lemma 7.1 that we can take $L = \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{HS}}$ and $M = \|(D\psi)_{\mathbf{q}}\|_{\mathrm{HS}}$, where $\|(D\varphi)_{\mathbf{p}}\|_{\mathrm{HS}}$ and $M = \|(D\psi)_{\mathbf{q}}\|_{\mathrm{HS}}$ denote the Hilbert-Schmidt norms of the linear transformations $(D\varphi)_{\mathbf{p}}$ and $(D\psi)_{\mathbf{q}}$.

Let some strictly positive number ε be given. The function ψ is differentiable at $\mathbf{q}$, with derivative $(D\psi)_{\mathbf{q}}$, and therefore there exists a strictly positive real number η such that $\mathbf{q} + \mathbf{k} \in W$ and

$$|\psi(\mathbf{q} + \mathbf{k}) - \psi(\mathbf{q}) - (D\psi)_{\mathbf{q}}\,\mathbf{k}| \leq \frac{1}{2(L+1)}\varepsilon|\mathbf{k}|$$

for all $\mathbf{k} \in \mathbb{R}^m$ satisfying $|\mathbf{k}| < \eta$ (see Lemma 7.5). Let ε_0 be a strictly positive number chosen such that $\varepsilon_0 < 1$ and $2M\varepsilon_0 < \varepsilon$. It then follows from the continuity and differentiability of φ at $\mathbf{p}$ that there exists some strictly positive real number δ satisfying $(L+1)\delta < \eta$ with the property that

$$\mathbf{p} + \mathbf{h} \in D \quad \text{and} \quad |\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}\mathbf{h}| \leq \varepsilon_0 |\mathbf{h}|$$

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

Let $\mathbf{h} \in \mathbb{R}^n$ satisfy $|\mathbf{h}| < \delta$, and let

$$\mathbf{k} = \varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) = \varphi(\mathbf{p} + \mathbf{h}) - \mathbf{q}.$$

Then

$$\begin{aligned} &\psi(\varphi(\mathbf{p} + \mathbf{h})) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}\,\mathbf{h} \\ &\qquad = (\psi(\mathbf{q} + \mathbf{k}) - \psi(\mathbf{q}) - (D\psi)_{\mathbf{q}}\,\mathbf{k}) \\ &\qquad\qquad + (D\psi)_{\mathbf{q}}\,(\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}\,\mathbf{h}). \end{aligned}$$

Also, on applying the Triangle Inequality satisfied by the Euclidean norm (see Corollary 6.2), we find that

$$\begin{aligned} |\mathbf{k}| &= |\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p})| \\ &\leq |(D\varphi)_{\mathbf{p}}\mathbf{h}| + |\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}\mathbf{h}| \\ &\leq L|\mathbf{h}| + \varepsilon_0|\mathbf{h}| \\ &\leq (L+1)|\mathbf{h}| < (L+1)\delta < \eta. \end{aligned}$$

It follows that

$$\begin{aligned} &|\psi(\varphi(\mathbf{p} + \mathbf{h})) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}\,\mathbf{h}| \\ &\qquad \leq |\psi(\mathbf{q} + \mathbf{k}) - \psi(\mathbf{q}) - (D\psi)_{\mathbf{q}}\,\mathbf{k}| \\ &\qquad\qquad + |(D\psi)_{\mathbf{q}}\,(\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}\,\mathbf{h})| \\ &\qquad \leq |\psi(\mathbf{q} + \mathbf{k}) - \psi(\mathbf{q}) - (D\psi)_{\mathbf{q}}\,\mathbf{k}| \\ &\qquad\qquad + M\,|\varphi(\mathbf{p} + \mathbf{h}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}\,\mathbf{h}| \\ &\qquad \leq \frac{1}{2(L+1)}\varepsilon|\mathbf{k}| + M\varepsilon_0|\mathbf{h}| \\ &\qquad \leq \tfrac{1}{2}\varepsilon|\mathbf{h}| + \tfrac{1}{2}\varepsilon|\mathbf{h}| = \varepsilon|\mathbf{h}| \end{aligned}$$

whenever $|\mathbf{h}| < \delta$. It follows that the composition function $\psi \circ \varphi\colon \mathbb{R}^n \to \mathbb{R}^l$ is differentiable, and its derivative at the point $\mathbf{p}$ is $(D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}$, as required. ∎

Example Consider the function $\varphi\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$\varphi(x, y) = \begin{cases} x^2 y^3 \sin \dfrac{1}{x} & \text{if } x \neq 0; \\ 0 & \text{if } x = 0. \end{cases}$$

Now one can verify from the definition of differentiability that the function $h\colon \mathbb{R} \to \mathbb{R}$ defined by

$$h(t) = \begin{cases} t^2 \sin \dfrac{1}{t} & \text{if } t \neq 0, \\ 0 & \text{if } t = 0 \end{cases}$$

is differentiable everywhere on $\mathbb{R}$, though its derivative $h'\colon \mathbb{R} \to \mathbb{R}$ is not continuous at 0. Also the functions $(x, y) \mapsto x$ and $(x, y) \mapsto y$ are differentiable everywhere on $\mathbb{R}$ (by Lemma 7.4). Now $\varphi(x, y) = y^3 h(x)$. Using Theorem 7.8 and Theorem 7.9, we conclude that φ is differentiable everywhere on $\mathbb{R}^2$.

Let $(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n)$ denote the standard basis of $\mathbb{R}^n$, where

$$\mathbf{e}_1 = (1, 0, \ldots, 0), \quad \mathbf{e}_2 = (0, 1, \ldots, 0), \ldots, \quad \mathbf{e}_n = (0, 0, \ldots, 1).$$

Let us denote by $f^i\colon V \to \mathbb{R}$ the ith component of the map $\varphi\colon V \to \mathbb{R}^m$, where V is an open subset of $\mathbb{R}^n$. Thus

$$\varphi(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}))$$

for all $\mathbf{x} \in V$. The jth partial derivative of f_i at $\mathbf{p} \in V$ is then given by

$$\left.\frac{\partial f_i}{\partial x_j}\right|_{\mathbf{x}=\mathbf{p}} = \lim_{t \to 0} \frac{f_i(\mathbf{p} + t\mathbf{e}_j) - f_i(\mathbf{p})}{t}.$$

We see therefore that if φ is differentiable at $\mathbf{p}$ then

$$(D\varphi)_{\mathbf{p}} \mathbf{e}_j = \left(\frac{\partial f_1}{\partial x_j}, \frac{\partial f_2}{\partial x_j}, \ldots, \frac{\partial f_m}{\partial x_j} \right).$$

Thus the linear transformation $(D\varphi)_{\mathbf{p}}\colon \mathbb{R}^n \to \mathbb{R}^m$ is represented by the $m \times n$ matrix

$$\begin{pmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} & \cdots & \dfrac{\partial f_1}{\partial x_n} \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} & \cdots & \dfrac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial f_m}{\partial x_1} & \dfrac{\partial f_m}{\partial x_2} & \cdots & \dfrac{\partial f_m}{\partial x_n} \end{pmatrix}$$

This matrix is known as the *Jacobian matrix* of φ at $\mathbf{p}$.

Example Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

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

Note that this function is not continuous at $(0,0)$. (Indeed $f(t,t) = 1/(4t^2)$ if $t \neq 0$ so that $f(t,t) \to +\infty$ as $t \to 0$, yet $f(x,0) = f(0,y) = 0$ for all $x, y \in \mathbb{R}$, thus showing that

$$\lim_{(x,y)\to(0,0)} f(x,y)$$

cannot possibly exist.) Because f is not continuous at $(0,0)$ we conclude from Lemma 7.6 that f cannot be differentiable at $(0,0)$. However it is easy to show that the partial derivatives

$$\frac{\partial f(x,y)}{\partial x} \text{ and } \frac{\partial f(x,y)}{\partial y}$$

exist everywhere on $\mathbb{R}^2$, even at $(0,0)$. Indeed

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

on account of the fact that $f(x,0) = f(0,y) = 0$ for all $x, y \in \mathbb{R}$.

Example Consider the function $g\colon \mathbb{R}^2 \to \mathbb{R}$ defined by

$$g(x,y) = \begin{cases} \dfrac{xy^2}{x^2+y^4} & \text{if } (x,y) \neq (0,0); \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

Given real numbers b and c, let $u_{b,c}\colon \mathbb{R} \to \mathbb{R}$ be defined so that $u_{b,c}(t) = g(bt, ct)$ for all $t \in \mathbb{R}$. If $b = 0$ or $c = 0$ then $u_{b,c}(t) = 0$ for all $t \in \mathbb{R}$, and the function $u_{b,c}$ is thus a smooth function of t. If $b \neq 0$ and $c \neq 0$ then

$$u_{b,c}(t) = \frac{bc^2t^3}{b^2t^2 + c^4t^4} = \frac{bc^2t}{b^2 + c^2t^2},$$

and therefore $u_{b,c}(t)$ is a smooth function of t. Moreover

$$\left.\frac{du_{b,c}(t)}{dt}\right|_{t=0} = \begin{cases} \dfrac{c^2}{b} & \text{if } b \neq 0; \\ 0 & \text{if } b = 0. \end{cases}$$

The restriction of the function g to any line passing through the origin determines a smooth function of distance along the line. The restriction of the function g to any other line in the plane also determines a smooth function of distance. It follows that, when restricted to any straight line in $\mathbb{R}^2$, the value of the function g is a smooth function of distance along that line.

However $g(x, y) = \frac{1}{2}$ for all $(x, y) \in \mathbb{R}^2$ satisfying $x > 0$ and $y = \pm\sqrt{x}$, and similarly $g(x, y) = -\frac{1}{2}$ for all $(x, y) \in \mathbb{R}^2$ satisfying $x < 0$ and $y = \pm\sqrt{-x}$. It follows that every open disk about the origin $(0, 0)$ contains some points at which the function g takes the value $\frac{1}{2}$, and other points at which the function takes the value $-\frac{1}{2}$, and indeed the function g will take on all real values between $-\frac{1}{2}$ and $\frac{1}{2}$ on any open disk about the origin, no matter how small the disk. It follows that the function $g\colon \mathbb{R}^2 \to \mathbf{R}$ is not continuous at zero, even though the partial derivatives of the function g with respect to x and y exist at each point of $\mathbb{R}^2$.

Remark These last two examples exhibits an important point. They show that *even if all the partial derivatives of a function exist at some point, this does not necessarily imply that the function is differentiable at that point.* However Theorem 7.11 below shows that if the first order partial derivatives of the components of a function exist *and are continuous* throughout some neighbourhood of a given point then the function is differentiable at that point.

Proposition 7.10 *Let M and δ_0 be positive real numbers, and let*

$$V = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : -\delta_0 < x_j < \delta_0 \textit{ for } j = 1, 2, \ldots, n\}.$$

let $f\colon V \to \mathbb{R}$ be a real-valued function defined over V. Suppose that the partial derivatives of the function f with respect to $x_1, x_2, \ldots, x_n$ exist throughout V, and satisfy

$$\left|\frac{\partial f(x_1, x_2, \ldots, x_n)}{\partial x_j}\right| \leq M$$

whenever $-\delta_0 < x_j \leq \delta_0$ for $j = 1, 2, \ldots, n$. Then

$$|f(\mathbf{v}) - f(\mathbf{u})| \leq \sqrt{n}\, M|\mathbf{v} - \mathbf{u}|$$

for all $\mathbf{u}, \mathbf{v} \in V$.

Proof Let points $\mathbf{w}_k$ for $k = 0, 1, 2, \ldots, n$ be defined so that

$$\mathbf{w}_k = (w_{k,1}, w_{k,2}, \ldots, w_{k,n}),$$

where

$$w_{k,j} = \begin{cases} u_j & \text{if } j > k; \\ v_j & \text{if } j \leq k. \end{cases}$$

Then $\mathbf{w}_0 = \mathbf{u}$ and $\mathbf{w}_n = \mathbf{v}$. Moreover $\mathbf{w}_k$ and $\mathbf{w}_{k-1}$ differ only in the kth coordinate for $k = 1, 2, \ldots, n$, and indeed $w_{k-1,k} = u_k$, $w_{k,k} = v_k$ and $w_{k,j} = w_{k-1,j}$ for $j \neq k$. Let $q_k \colon [0, 1] \to \mathbb{R}$ be defined such that

$$q_k(t) = f((1 - t)\mathbf{w}_{k-1} + t\mathbf{w}_k)$$

for all $t \in [0, 1]$. Then $q_k(0) = f(\mathbf{w}_{k-1})$ and $q_k(1) = f(\mathbf{w}_k)$, and therefore

$$f(\mathbf{v}) - f(\mathbf{u}) = \sum_{k=1}^{n} (f(\mathbf{w}_k) - f(\mathbf{w}_{k-1})) = \sum_{k=1}^{n} (q_k(1) - q_k(0)).$$

Now

$$q_k'(t) = \frac{dq_k(t)}{dt} = (v_k - u_k)(\partial_k f)((1 - t)\mathbf{w}_{k-1} + t\mathbf{w}_k)$$

for all $t \in [0, 1]$, where $\partial_k f$ denotes the partial derivative of the function f with respect to x_k. Moreover $|(\partial_k f)(\mathbf{x})| \leq M$ for all $\mathbf{x} \in V$. It follows that $|q_k'(t)| \leq M|v_k - u_k|$ for all $t \in [0, 1]$. Applying the Mean Value Function (Theorem 4.6) to the function q on the interval $[0, 1]$, we see that

$$|q_k(1) - q_k(0)| \leq M|v_k - u_k|$$

for $k = 1, 2, \ldots, n$. It follows that

$$|f(\mathbf{v}) - f(\mathbf{u})| \leq \sum_{k=1}^{n} |q_k(1) - q_k(0)| \leq M \sum_{k=1}^{n} |v_k - u_k|.$$

Now

$$\sum_{k=1}^{n} |v_k - u_k| \leq \sqrt{n}|\mathbf{v} - \mathbf{u}|.$$

Indeed let $\mathbf{s} \in \mathbf{R}^n$ be defined such that $\mathbf{s} = (s_1, s_2, \ldots, s_n)$ where $s_j = +1$ if $v_j \geq u_j$ and $s_j = -1$ if $v_j < u_j$. Then

$$\sum_{k=1}^{n} |v_k - u_k| = \mathbf{s} \,.\, (\mathbf{v} - \mathbf{u}) \leq |\mathbf{s}||\mathbf{v} - \mathbf{u}| = \sqrt{n}|\mathbf{v} - \mathbf{u}|.$$

The result follows. ∎

Theorem 7.11 *Let V be an open subset of $\mathbb{R}^m$ and let $f\colon V \to \mathbb{R}$ be a function mapping V into $\mathbb{R}$. Suppose that the first order partial derivatives of the components of f exist and are continuous on V. Then f is differentiable at each point of V, and*

$$(Df)_{\mathbf{p}}\mathbf{h} = \sum_{j=1}^{n} h_j \left.\frac{\partial f}{\partial x_j}\right|_{\mathbf{x}=\mathbf{p}}$$

for all $\mathbf{p} \in V$ and $\mathbf{h} \in \mathbb{R}^n$, where $\mathbf{h} = (h_1, h_2, \ldots, h_n)$.

Proof Let $\mathbf{p} \in V$, and let $g\colon V \to \mathbb{R}$ be defined such that

$$g(\mathbf{x}) = f(\mathbf{x}) - \sum_{j=1}^{n} a_j(x_j - p_j)$$

for all $\mathbf{x} \in V$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and

$$a_j = (\partial_j f)(\mathbf{p}) = \left.\frac{\partial f}{\partial x_j}\right|_{\mathbf{p}}$$

for $j = 1, 2, \ldots, n$. The partial derivatives $\partial_j g$ of the function g are then determined by those of f so that

$$(\partial_j g)(\mathbf{x}) = (\partial_j f)(\mathbf{x}) - a_j$$

for $j = 1, 2, \ldots, n$. It follows that $(\partial_j g)(\mathbf{p}) = 0$ for $j = 1, 2, \ldots, n$. It follows from the continuity of the partial derivatives of f that, given any positive real number ε, there exists some positive real number δ such that $(x_1, x_2, \ldots, x_n) \in V$ and, for each integer k between 1 and n,

$$|(\partial_k g)(x_1, x_2, \ldots, x_n)| \leq \frac{\varepsilon}{\sqrt{n}}$$

whenever $p_j - \delta < x_j < p_j + \delta$ for $j = 1, 2, \ldots, n$. It then follows from Proposition 7.10 that

$$|g(\mathbf{p} + \mathbf{h}) - g(\mathbf{p})| \leq \varepsilon|\mathbf{h}|$$

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

$$\left| f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p}) - \sum_{j=1}^{n} h_j(\partial_j f)(\mathbf{p}) \right| \leq \varepsilon|\mathbf{h}|$$

for all $\mathbf{h} \in \mathbb{R}^n$ satisfying $|\mathbf{h}| < \delta$, where $\mathbf{h} = (h_1, h_2, \ldots, h_n)$. It follows from Lemma 7.5 that the function f is differentiable at $\mathbf{p}$. Moreover the Cartesian components of the derivative of f at $\mathbf{p}$ are equal to the partial derivatives of f at that point, as required. ∎

We can generalize this result immediately to functions $u: V \to \mathbb{R}^m$ which map some open subset V of $\mathbb{R}^n$ into $\mathbb{R}^m$. Let u_i denote the ith component of u for $i = 1, 2, \ldots, m$. One sees easily from the definition of differentiability that u is differentiable at a point of V if and only if each u_i is differentiable at that point. We can therefore deduce immediately the following corollary.

Corollary 7.12 *Let V be an open subset of $\mathbb{R}^n$ and let $u: V \to \mathbb{R}^m$ be a function mapping V into $\mathbb{R}^m$. Suppose that the Jacobian matrix*

$$\begin{pmatrix} \dfrac{\partial u_1}{\partial x_1} & \dfrac{\partial u_1}{\partial x_2} & \cdots & \dfrac{\partial u_1}{\partial x_n} \\ \dfrac{\partial u_2}{\partial x_1} & \dfrac{\partial u_2}{\partial x_2} & \cdots & \dfrac{\partial u_2}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial u_m}{\partial x_1} & \dfrac{\partial u_m}{\partial x_2} & \cdots & \dfrac{\partial u_m}{\partial x_n} \end{pmatrix}$$

exists at every point of V and that the entries of the Jacobian matrix are continuous functions on V. Then φ is differentiable at every point of V, and the derivative of φ at each point is represented by the Jacobian matrix.

We now summarize the main conclusions regarding differentiability of functions of several real variables. They are as follows.

(i) A function $\varphi: V \to \mathbb{R}^m$ defined on an open subset V of $\mathbb{R}^n$ is said to be *differentiable* at a point $\mathbf{p}$ of V if and only if there exists a linear transformation $(D\varphi)_{\mathbf{p}}: \mathbb{R}^n \to \mathbb{R}^m$ with the property that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} \left(\varphi(\mathbf{p}+\mathbf{h}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}\,\mathbf{h}\right) = \mathbf{0}.$$

The linear transformation $(D\varphi)_{\mathbf{p}}$ (if it exists) is unique and is known as the *derivative* (or *total derivative*) of φ at $\mathbf{p}$.

(ii) If the function $\varphi: V \to \mathbb{R}^m$ is differentiable at a point $\mathbf{p}$ of V then the derivative $(D\varphi)_{\mathbf{p}}$ of φ at $\mathbf{p}$ is represented by the Jacobian matrix of the function φ at $\mathbf{p}$ whose entries are the first order partial derivatives of the components of φ.

(iii) There exist functions $\varphi: V \to \mathbb{R}^m$ whose first order partial derivatives are well-defined at a particular point of V but which are not differentiable at that point. Indeed there exist such functions whose first order partial derivatives exist throughout their domain, though the functions themselves are not even continuous. *Thus in order to show that a function is differentiable at a particular point, it is not sufficient to show that the first order partial derivatives of the function exist at that point.*

(iv) However if the first order partial derivatives of the components of a function $\varphi: V \to \mathbb{R}^m$ exist *and are continuous* throughout some neighbourhood of a given point then the function is differentiable at that point. (However the converse does not hold: there exist functions which are differentiable whose first order partial derivatives are not continuous.)

(v) Linear transformations are everywhere differentiable.

(vi) A function $\varphi: V \to \mathbb{R}^m$ is differentiable if and only if its components are differentiable functions on V (where V is an open set in $\mathbb{R}^n$).

(vii) Given two differentiable functions from V to $\mathbb{R}$, where V is an open set in $\mathbb{R}^n$, the sum, difference and product of these functions are also differentiable.

(viii) (The Chain Rule). The composition of two differentiable functions is differentiable, and the derivative of the composition of the functions at any point is the composition of the derivatives of the functions.

7.4 Second Order Partial Derivatives

Let V be an open subset of $\mathbb{R}^n$ and let $f: V \to \mathbb{R}$ be a real-valued function on V. 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

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

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 $f\colon \mathbb{R}^2 \to \mathbb{R}$ be the function defined by

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

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

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

Similarly

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

Thus if $(x,y) \neq (0,0)$ then

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

Note that

$$\lim_{(x,y)\to(0,0)} f_x(x,y) = 0, \qquad \lim_{(x,y)\to(0,0)} f_y(x,y) = 0.$$

Indeed if $(x,y) \neq (0,0)$ then

$$|f_x| \leq \frac{6r^5}{r^4} = 6r,$$

where $r = \sqrt{x^2 + y^2}$, and similarly $|f_y| \leq 6r$. However

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

does not exist. Indeed

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

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

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

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

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

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 $f_{xy}(0,0) \neq f_{yx}(0,0)$.

We now prove that the continuity of the first and second order partial derivatives of a function f of two variables x and y is sufficient to ensure that

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

Theorem 7.13 *Let V be an open set in $\mathbb{R}^2$ and let $f\colon V \to \mathbb{R}$ be a real-valued function on V. Suppose that the partial derivatives*

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

exist and are continuous on V. Then

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

Proof For convenience, we shall denote the values of

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

at a point (x, y) of V by $f_x(x, y)$, $f_y(x, y)$, $f_{xy}(x, y)$ and $f_{yx}(x, y)$ respectively.

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

Let us define a differentiable function u by

$$u(t) = f(t, b + k) - f(t, b)$$

We apply the Mean Value Theorem to the function u on the closed interval $[a, a + h]$ to conclude that there exists θ_1, where $0 < \theta_1 < 1$, such that

$$u(a + h) - u(a) = hu'(a + \theta_1 h).$$

But

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

and

$$u'(a + \theta_1 h) = f_x(a + \theta_1 h, b + k) - f_x(a + \theta_1 h, b).$$

Moreover, on applying the Mean Value Theorem to the function that sends $y \in [b, b + k]$ to $f_x(a + \theta_1 h, y)$, we see that there exists θ_2, where $0 < \theta_2 < 1$, such that

$$f_x(a + \theta_1 h, b + k) - f_x(a + \theta_1 h, b) = k f_{yx}(a + \theta_1 h, b + \theta_2 k)$$

Thus

$$\begin{aligned} & f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b) \\ & \qquad = \quad hk f_{yx}(a + \theta_1 h, b + \theta_2 k) = hk \left. \frac{\partial^2 f}{\partial y \partial x} \right|_{(x,y)=(a+\theta_1 h, b+\theta_2 k)}. \end{aligned}$$

Now let $\varepsilon > 0$ be given. Then there exists some positive real number δ_1, where $\delta_1 \leq R$, such that

$$|f_{yx}(x, y) - f_{yx}(a, b)| < \tfrac{1}{2}\varepsilon$$

whenever $(x - a)^2 + (y - b)^2 < \delta_1^2$, by the continuity of the function f_{yx}. Thus

$$\left| \frac{f(a + h, b + k) - f(a + h, b) - f(a, b + k) + f(a, b)}{hk} - f_{yx}(a, b) \right| < \tfrac{1}{2}\varepsilon$$

for all $(h,k) \in \mathbb{R}^2$ for which $hk \neq 0$ and $\sqrt{h^2+k^2} < \delta_1$.

A corresponding result holds with the roles of x and y interchanged, and therefore there exists some positive real number δ_2, where $\delta_2 \leq R$, such that

$$\left|\frac{f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b)}{hk} - f_{xy}(a,b)\right| < \tfrac{1}{2}\varepsilon$$

for all $(h,k) \in \mathbb{R}^2$ for which $hk \neq 0$ and $\sqrt{h^2+k^2} < \delta_2$.

Take δ to be the minimum of δ_1 and δ_2. If $hk \neq 0$ and $\sqrt{h^2+k^2} < \delta^2$ then

$$\begin{aligned}
\left|\frac{f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b)}{hk} - f_{yx}(a,b)\right| &< \tfrac{1}{2}\varepsilon, \\
\left|\frac{f(a+h,b+k) - f(a+h,b) - f(a,b+k) + f(a,b)}{hk} - f_{xy}(a,b)\right| &< \tfrac{1}{2}\varepsilon.
\end{aligned}$$

Using the triangle inequality we conclude that

$$|f_{yx}(a,b) - f_{xy}(a,b)| < \varepsilon.$$

But this inequality has to hold for all $\varepsilon > 0$. Therefore we must have

$$f_{yx}(a,b) = f_{xy}(a,b).$$

We conclude therefore that

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

at each point (a,b) of V, as required. ∎

Remark It is actually possible to prove a somewhat stronger theorem which states that, if $f\colon V \to \mathbb{R}$ is a real-valued function defined on a open subset V of $\mathbb{R}^2$ and if the partial derivatives

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

exist and are continuous at some point (a,b) of V then

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

exists at (a,b) and

$$\left.\frac{\partial^2 f}{\partial y \partial x}\right|_{(a,b)} = \left.\frac{\partial^2 f}{\partial x \partial y}\right|_{(a,b)}.$$

Corollary 7.14 *Let V be an open set in $\mathbb{R}^n$ and let $f: V \to \mathbb{R}$ be a real-valued function on V. 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 V 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.

7.5 Maxima and Minima

Let $f: V \to \mathbb{R}$ be a real-valued function defined over some open subset V of $\mathbb{R}^n$ whose first and second order partial derivatives exist and are continuous throughout V. Suppose that f has a local minimum at some point $\mathbf{p}$ of V, where $\mathbf{p} = (p_1, p_2, \ldots, a_n)$. Now for each integer i between 1 and n the map

$$t \mapsto f(p_1, \ldots, a_{i-1}, t, a_{i+1}, \ldots, a_n)$$

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

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

The following lemma applies Taylor's Theorem (for functions of a single real variable) the local behaviour of real-valued functions of several real variables that are twice continuously differentiable throughout an open neighbourhood of some given point.

Lemma 7.15 *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*

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

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 7.9

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

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

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

It follows from Taylor's Theorem for functions of a single real variable (Theorem 4.21) that if the function f has continuous partial derivatives of orders one and two then

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

for some real number θ satisfying $0 < \theta < 1$. It follows that

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

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 neigbourhood 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 7.15 that if

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

for $j = 1, 2, \ldots, n$, 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 $H_{i,j}(\mathbf{p}) = Hji(\mathbf{p})$ for all i and j, by Corollary 7.14. 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 $(h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$.

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 $(h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$.

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 $(h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$.

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 $(h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$.

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

Lemma 7.16 *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

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

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 $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. 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 $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. But S^{n-1} is a closed bounded set in $\mathbb{R}^n$, it therefore follows from Theorem 6.21 that there exists some $(k_1, k_2, \ldots, k_n) \in S^{n-1}$ 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 $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. Thus there exists a strictly positive constant $A > 0$ with the property that

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

for all $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. 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 $(h_1, h_2, \ldots, h_n) \in S^{n-1}$, 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 $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. 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 $f\colon V \to \mathbb{R}$ be a real-valued function whose partial derivatives of first and second order exist and are continuous throughout some open set V in $\mathbb{R}^n$. Let $\mathbf{p}$ be a point of V. We have already observed that if the function f has a local maximum or a local minimum at $\mathbf{p}$ then

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

We now apply Taylor's theorem to 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 7.17 *Let $f\colon V \to \mathbb{R}$ be a real-valued function whose partial derivatives of first and second order exist and are continuous throughout some open set V in $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of V at which*

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

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

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 7.15). Suppose that the Hessian matrix $H_{i,j}(\mathbf{p})$ is not positive semi-definite. Then there exists some $\mathbf{k} \in \mathbb{R}^n$, where $|\mathbf{k}| = 1$ with the property that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} k_i k_j H_{i,j}(\mathbf{p}) < 0.$$

It follows from the continuity of the second order partial derivatives of f that there exists some $\delta > 0$ such that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} k_i k_j H_{i,j}(\mathbf{x}) < 0$$

for all $\mathbf{x} \in V$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$. Choose any λ such that $0 < \lambda < \delta$ and set $\mathbf{h} = \lambda\mathbf{k}$. Then

$$\sum_{i=1}^{n}\sum_{j=1}^{n} h_i h_j H_{i,j}(\mathbf{p} + \theta\mathbf{h}) < 0$$

for all $\theta \in (0, 1)$. We conclude from Taylor's theorem that $f(\mathbf{p} + \lambda\mathbf{k}) < f(\mathbf{p})$ for all λ satisfying $0 < \lambda < \delta$ (see Lemma 7.15). We have thus shown that if the Hessian matrix at $\mathbf{p}$ is not positive semi-definite then $\mathbf{p}$ is not a local minimum. Thus the Hessian matrix of f is positive semi-definite at every local minimum of f, as required. ∎

Let $f: V \to \mathbb{R}$ be a real-valued function whose partial derivatives of first and second order exist and are continuous throughout some open set V 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 is has a local minimum at $\mathbf{p}$, as the following example shows.

Example Consider the function $f: \mathbb{R}^2 \to \mathbb{R}$ 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 7.18 *Let $f: V \to \mathbb{R}$ be a real-valued function whose partial derivatives of first and second order exist and are continuous throughout some open set V in $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of V at which*

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

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 7.15). Suppose that the Hessian matrix $(H_{i,j}(\mathbf{p}))$ is positive definite. It follows from Lemma 7.16 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 a symmetric 2×2 matrix C, then

$$\lambda_1 + \lambda_2 = \operatorname{trace} C, \qquad \lambda_1 \lambda_2 = \det C.$$

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 $f \colon \mathbb{R}^2 \to \mathbb{R}$ defined by

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

Now

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

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 7.18 that the function f has a local minimum at $(0, 0)$.