# Module MA2321: Analysis in Several Real Variables
# Michaelmas Term 2017
# Part III (Sections 7 to 10)

D. R. Wilkins



## Contents

# 7 Differentiation of Functions of One Real Variable

## 7.1 Interior Points and Open Sets in the Real Line

**Definition** Let $D$ be a subset of the set $\mathbb{R}$ of real numbers, and let $s$ be a real number belonging to $D$. We say that $s$ is an *interior point* of $D$ if there exists some strictly positive number $\delta$ such that $x \in D$ for all real numbers $x$ satisfying $s - \delta < x < s + \delta$. The *interior* of $D$ is then the subset of $D$ consisting of all real numbers belonging to $D$ that are interior points of $D$.

It follows from the definition of open sets in Euclidean spaces that a subset $D$ of the set $\mathbb{R}$ of real numbers is an open set in $\mathbb{R}$ if and only if every point of $D$ is an interior point of $D$.

Let $s$ be a real number. We say that a function $f \colon D \to \mathbb{R}$ is defined *around* $s$ if the real number $s$ is an interior point of the domain $D$ of the function $f$. It follows that the function $f$ is defined around $s$ if and only if there exists some strictly positive real number $\delta$ such that $f(x)$ is defined for all real numbers $x$ satisfying $s - \delta < x < s + \delta$.

## 7.2 Differentiable Functions of a Single Real Variable

We recall basic results of the theory of differentiable functions.

**Definition** Let $s$ be some real number, and let $f$ be a real-valued function defined around $s$. The function $f$ is said to be *differentiable* at $s$, with *derivative* $f'(s)$, if and only if the limit

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

is well-defined. We denote by $f'$, or by $\dfrac{df}{dx}$ the function whose value at $s$ is the derivative $f'(s)$ of $f$ at $s$.

Let $s$ be some real number, and let $f$ and $g$ be real-valued functions defined around $s$ that are differentiable at $s$. The basic rules of differential calculus then ensure that the functions $f+g$, $f-g$ and $f \cdot g$ are differentiable at $s$ (where

$$(f+g)(x) = f(x) + g(x), \quad (f-g)(x) = f(x) - g(x)$$

and

$$(f.g)(x) = f(x)g(x)$$

for all real numbers $x$ at which both $f(x)$ and $g(x)$ are defined), and

$$(f+g)'(s) = f'(s) + g'(s), \qquad (f-g)'(s) = f'(s) - g'(s).$$

$$(f \cdot g)'(s) = f'(s)g(s) + f(s)g'(s) \quad (\textit{Product Rule}).$$

If moreover $g(s) \neq 0$ then the function $f/g$ is differentiable at $s$ (where $(f/g)(x) = f(x)/g(x)$ where both $f(x)$ and $g(x)$ are defined), and

$$(f/g)'(s) = \frac{f'(s)g(s) - f(s)g'(s)}{g(s)^2} \quad (\textit{Quotient Rule}).$$

Moreover if $h$ is a real-valued function defined around $f(s)$ which is differentiable at $f(s)$ then the composition function $h \circ f$ is differentiable at $f(s)$ and

$$(h \circ f)'(s) = h'(f(s))f'(s) \quad (\textit{Chain Rule}).$$

Derivatives of some standard functions are as follows:—

$$\frac{d}{dx}(x^m) = mx^{m-1}, \quad \frac{d}{dx}(\sin x) = \cos x, \quad \frac{d}{dx}(\cos x) = -\sin x,$$

$$\frac{d}{dx}(\exp x) = \exp x, \quad \frac{d}{dx}(\log x) = \frac{1}{x} \;\; (x > 0).$$

## 7.3 Rolle's Theorem

**Theorem 7.1 (Rolle's Theorem)** *Let $f\colon [a,b] \to \mathbb{R}$ be a real-valued function defined on some interval $[a,b]$. Suppose that $f$ is continuous on $[a,b]$ and is differentiable on $(a,b)$. Suppose also that $f(a) = f(b)$. Then there exists some real number $s$ satisfying $a < s < b$ which has the property that $f'(s) = 0$.*

**Proof** First we show that if the function $f$ attains its minimum value at $u$, and if $a < u < b$, then $f'(u) = 0$. Now the difference quotient

$$\frac{f(u+h) - f(u)}{h}$$

is non-negative for all sufficiently small positive values of $h$; therefore $f'(u) \geq 0$. On the other hand, this difference quotient is non-positive for all sufficiently small negative values of $h$; therefore $f'(u) \leq 0$. We deduce therefore that $f'(u) = 0$.

Similarly if the function $f$ attains its maximum value at $v$, and if $a < v < b$, then $f'(v) = 0$. (Indeed the result for local maxima can be deduced from the corresponding result for local minima simply by replacing the function $f$ by $-f$.)

Now the function $f$ is continuous on the closed bounded interval $[a, b]$. It therefore follows from the Extreme Value Theorem that there must exist real numbers $u$ and $v$ in the interval $[a, b]$ with the property that $f(u) \leq f(x) \leq f(v)$ for all real numbers $x$ satisfying $a \leq x \leq b$ (see Theorem 4.21). If $a < u < b$ then $f'(u) = 0$ and we can take $s = u$. Similarly if $a < v < b$ then $f'(v) = 0$ and we can take $s = v$. The only remaining case to consider is when both $u$ and $v$ are endpoints of the interval $[a, b]$. In that case the function $f$ is constant on $[a, b]$, since $f(a) = f(b)$, and we can choose $s$ to be any real number satisfying $a < s < b$. ∎

## 7.4 The Mean Value Theorem

Rolle's Theorem can be generalized to yield the following important theorem.

**Theorem 7.2 (The Mean Value Theorem)** *Let $f\colon [a, b] \to \mathbb{R}$ be a real-valued function defined on some interval $[a, b]$. Suppose that $f$ is continuous on $[a, b]$ and is differentiable on $(a, b)$. Then there exists some real number $s$ satisfying $a < s < b$ which has the property that*

$$f(b) - f(a) = f'(s)(b - a).$$

**Proof** Let $g\colon [a, b] \to \mathbb{R}$ be the real-valued function on the closed interval $[a, b]$ defined by

$$g(x) = f(x) - \frac{b - x}{b - a} f(a) - \frac{x - a}{b - a} f(b).$$

Then the function $g$ is continuous on $[a, b]$ and differentiable on $(a, b)$. Moreover $g(a) = 0$ and $g(b) = 0$. It follows from Rolle's Theorem (Theorem 7.1) that $g'(s) = 0$ for some real number $s$ satisfying $a < s < b$. But

$$g'(s) = f'(s) - \frac{f(b) - f(a)}{b - a}.$$

Therefore $f(b) - f(a) = f'(s)(b - a)$, as required. ∎

## 7.5 Concavity and the Second Derivative

**Proposition 7.3** *Let $s$ and $h$ be real numbers, and let $f$ be a twice differentiable real-valued function defined on some open interval containing $s$ and $s+h$. Then there exists a real number $\theta$ satisfying $0 < \theta < 1$ for which*

$$f(s+h) = f(s) + hf'(s) + \tfrac{1}{2}h^2 f''(s+\theta h).$$

**Proof** Let $I$ be an open interval, containing the real numbers 0 and 1, chosen to ensure that $f(s+th)$ is defined for all $t \in I$, and let $q\colon I \to \mathbb{R}$ be defined so that

$$q(t) = f(s+th) - f(s) - thf'(s) - t^2(f(s+h) - f(s) - hf'(s)).$$

for all $t \in I$. Differentiating, we find that

$$q'(t) = hf'(s+th) - hf'(s) - 2t(f(s+h) - f(s) - hf'(s))$$

and

$$q''(t) = h^2 f''(s+th) - 2(f(s+h) - f(s) - hf'(s)).$$

Now $q(0) = q(1) = 0$. It follows from Rolle's Theorem, applied to the function $q$ on the interval $[0,1]$, that there exists some real number $\varphi$ satisfying $0 < \varphi < 1$ for which $q'(\varphi) = 0$.

Then $q'(0) = q'(\varphi) = 0$, and therefore Rolle's Theorem can be applied to the function $q'$ on the interval $[0,\varphi]$ to prove the existence of some real number $\theta$ satisfying $0 < \theta < \varphi$ for which $q''(\theta) = 0$. Then

$$0 = q''(\theta) = h^2 f''(s+\theta h) - 2(f(s+h) - f(s) - hf'(s)).$$

Rearranging, we find that

$$f(s+h) = f(s) + hf'(s) + \tfrac{1}{2}h^2 f''(s+\theta h),$$

as required. ∎

**Corollary 7.4** *Let $f\colon (s-\delta_0, s+\delta_0)$ be a twice-differentiable function throughout some open interval $(s-\delta_0, s+\delta_0)$ centred on a real number $s$. Suppose that $f''(s+h) > 0$ for all real numbers $h$ satisfying $|h| < \delta_0$. Then*

$$f(s+h) \geq f(s) + hf'(s)$$

*for all real numbers $h$ satisfying $|h| < \delta_0$.*

It follows from Corollary 7.4 that if a twice-differentiable function has positive second derivative throughout some open interval, then it is concave upwards throughout that interval. In particular the function has a local minimum at any point of that open interval where the first derivative is zero and the second derivative is positive.

**Corollary 7.5** *Let $f\colon D \to \mathbb{R}$ be a twice-differentiable function defined over a subset $D$ of $\mathbb{R}$, and let $s$ be a point in the interior of $D$. Suppose that $f'(s) = 0$ and that $f''(x) > 0$ for all real numbers $x$ belonging to some sufficiently small neighbourhood of $x$. Then $s$ is a local minimum for the function $f$.*

## 7.6 Taylor's Theorem

The result obtained in Proposition 7.3 is a special case of a more general result. That more general result is a version of Taylor's Theorem with remainder. The proof of this theorem will make use of the following lemma.

**Lemma 7.6** *Let $s$ and $h$ be real numbers, let $f$ be a $k$ times differentiable real-valued function defined on some open interval containing $s$ and $s + h$, let $c_0, c_1, \ldots, c_{k-1}$ be real numbers, and let*

$$p(t) = f(s + th) - \sum_{n=0}^{k-1} c_n t^n.$$

*for all real numbers $t$ belonging to some open interval $D$ for which $0 \in D$ and $1 \in D$. Then $p^{(n)}(0) = 0$ for all integers $n$ satisfying $0 \leq n < k$ if and only if*

$$c_n = \frac{h^n f^{(n)}(s)}{n!}$$

*for all integers $n$ satisfying $0 \leq n < k$.*

**Proof** On setting $t = 0$, we find that $p(0) = f(s) - c_0$, and thus $p(0) = 0$ if and only if $c_0 = f(s)$.

Let the integer $n$ satisfy $0 < n < k$. On differentiating $p(t)$ $n$ times with respect to $t$, we find that

$$p^{(n)}(t) = h^n f^{(n)}(s + th) - \sum_{j=n}^{k-1} \frac{j!}{(j-n)!} c_j t^{j-n}.$$

Then, on setting $t = 0$, we find that only the term with $j = n$ contributes to the value of the sum on the right hand side of the above identity, and therefore

$$p^{(n)}(0) = h^n f^{(n)}(s) - n! c_n.$$

The result follows. ∎

**Theorem 7.7** (Taylor's Theorem) *Let $s$ and $h$ be real numbers, and let $f$ be a $k$ times differentiable real-valued function defined on some open interval containing $s$ and $s + h$. Then*

$$f(s+h) = f(s) + \sum_{n=1}^{k-1} \frac{h^n}{n!} f^{(n)}(s) + \frac{h^k}{k!} f^{(k)}(s + \theta h)$$

*for some real number $\theta$ satisfying $0 < \theta < 1$.*

**Proof** Let $D$ be an open interval, containing the real numbers 0 and 1, chosen to ensure that $f(s + th)$ is defined for all $t \in D$, and let $p\colon D \to \mathbb{R}$ be defined so that

$$p(t) = f(s+th) - f(s) - \sum_{n=1}^{k-1} \frac{t^n h^n}{n!} f^{(n)}(s)$$

for all $t \in D$. A straightforward calculation shows that $p^{(n)}(0) = 0$ for $n = 0, 1, \ldots, k-1$ (see Lemma 7.6). Thus if $q(t) = p(t) - p(1)t^k$ for all $s \in [0, 1]$ then $q^{(n)}(0) = 0$ for $n = 0, 1, \ldots, k-1$, and $q(1) = 0$. We can therefore apply Rolle's Theorem (Theorem 7.1) to the function $q$ on the interval $[0, 1]$ to deduce the existence of some real number $t_1$ satisfying $0 < t_1 < 1$ for which $q'(t_1) = 0$. We can then apply Rolle's Theorem to the function $q'$ on the interval $[0, t_1]$ to deduce the existence of some real number $t_2$ satisfying $0 < t_2 < t_1$ for which $q''(t_2) = 0$. By continuing in this fashion, applying Rolle's Theorem in turn to the functions $q'', q''', \ldots, q^{(k-1)}$, we deduce the existence of real numbers $t_1, t_2, \ldots, t_k$ satisfying $0 < t_k < t_{k-1} < \cdots < t_1 < 1$ with the property that $q^{(n)}(t_n) = 0$ for $n = 1, 2, \ldots, k$. Let $\theta = t_k$. Then $0 < \theta < 1$ and

$$0 = \frac{1}{k!} q^{(k)}(\theta) = \frac{1}{k!} p^{(k)}(\theta) - p(1) = \frac{h^k}{k!} f^{(k)}(s + \theta h) - p(1),$$

hence

$$\begin{aligned} f(s+h) &= f(s) + \sum_{n=1}^{k-1} \frac{h^n}{n!} f^{(n)}(s) + p(1) \\ &= f(s) + \sum_{n=1}^{k-1} \frac{h^n}{n!} f^{(n)}(s) + \frac{h^k}{k!} f^{(k)}(s + \theta h), \end{aligned}$$

as required. ∎

**Corollary 7.8** *Let $f\colon D \to \mathbb{R}$ be a $k$-times continuously differentiable function defined over an open subset $D$ of $\mathbb{R}$ and let $s \in \mathbb{R}$. Then given any strictly positive real number $\varepsilon$, there exists some strictly positive real number $\delta$ such that*

$$\left| f(s+h) - f(s) - \sum_{n=1}^{k} \frac{h^n}{n!} f^{(n)}(s) \right| < \varepsilon |h|^k$$

*whenever $|h| < \delta$.*

**Proof** The function $f$ is $k$-times continuously differentiable, and therefore its $k$th derivative $f^{(k)}$ is continuous. Let some strictly positive real number $\varepsilon$ be given. Then there exists some strictly positive real number $\delta$ that is small enough to ensure that $s + h \in D$ and $|f^{(k)}(s+h) - f^{(k)}(s)| < k!\varepsilon$ whenever $|h| < \delta$. If $h$ is an real number satisfying $|h| < \delta$, and if $\theta$ is a real number satisfying $0 < \theta < 1$, then $s+\theta h \in D$ and $|f^{(k)}(s+\theta h) - f^{(k)}(s)| < k!\varepsilon$. Now it follows from Taylor's Theorem (Theorem 7.7) that, given any real number $h$ satisfying $|h| < \delta$ there exists some real number $\theta$ satisfying $0 < \theta < 1$ for which

$$f(s+h) = f(s) + \sum_{n=1}^{k-1} \frac{h^n}{n!} f^{(n)}(s) + \frac{h^k}{k!} f^{(k)}(s + \theta h).$$

Then

$$\begin{aligned} \left| f(s+h) - f(s) - \sum_{n=1}^{k} \frac{h^n}{n!} f^{(n)}(s) \right| &= \frac{|h|^k}{k!} |f^{(k)}(s+\theta h) - f^{(k)}(s)| \\ &< \varepsilon |h|^k, \end{aligned}$$

as required. ∎

Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function on a closed interval $[a,b]$. We say that $f$ is *continuously differentiable* on $[a,b]$ if the derivative $f'(x)$ of $f$ exists for all $x$ satisfying $a < x < b$, the one-sided derivatives

$$\begin{aligned} f'(a) &= \lim_{h \to 0^+} \frac{f(a+h) - f(a)}{h}, \\ f'(b) &= \lim_{h \to 0^-} \frac{f(b+h) - f(b)}{h} \end{aligned}$$

exist at the endpoints of $[a,b]$, and the function $f'$ is continuous on $[a,b]$.

If $f\colon [a,b] \to \mathbb{R}$ is continuous, and if $F(x) = \int_a^x f(t)\,dt$ for all $x \in [a,b]$ then the one-sided derivatives of $F$ at the endpoints of $[a,b]$ exist, and

$$\lim_{h \to 0^+} \frac{F(a+h) - F(a)}{h} = f(a), \qquad \lim_{h \to 0^-} \frac{F(b+h) - F(b)}{h} = f(b).$$

One can verify these results by adapting the proof of the Fundamental Theorem of Calculus.

**Proposition 7.9** *Let $f$ be a continuously differentiable real-valued function on the interval $[a,b]$. Then*

$$\int_a^b \frac{df(x)}{dx}\,dx = f(b) - f(a)$$

**Proof** Define $g\colon [a,b] \to \mathbb{R}$ by

$$g(x) = f(x) - f(a) - \int_a^x \frac{df(t)}{dt}\,dt.$$

Then $g(a) = 0$, and

$$\frac{dg(x)}{dx} = \frac{df(x)}{dx} - \frac{d}{dx}\left(\int_a^x \frac{df(t)}{dt}\,dt\right) = 0$$

for all $x$ satisfying $a < x < b$, by the Fundamental Theorem of Calculus. Now it follows from the Mean Value Theorem (Theorem 7.2) that there exists some $s$ satisfying $a < s < b$ for which $g(b) - g(a) = (b-a)g'(s)$. We deduce therefore that $g(b) = 0$, which yields the required result. ∎

**Corollary 7.10 (Integration by Parts)** *Let $f$ and $g$ be continuously differentiable real-valued functions on the interval $[a,b]$. Then*

$$\int_a^b f(x)\,\frac{dg(x)}{dx}\,dx = f(b)g(b) - f(a)g(a) - \int_a^b g(x)\,\frac{df(x)}{dx}\,dx.$$

**Proof** This result follows from Proposition 7.9 on integrating the identity

$$f(x)\frac{dg(x)}{dx} = \frac{d}{dx}\left(f(x)g(x)\right) - g(x)\,\frac{df(x)}{dx}. \quad ∎$$

**Corollary 7.11 (Integration by Substitution)** *Let $u\colon [a,b] \to \mathbb{R}$ be a continuously differentiable monotonically increasing function on the interval $[a,b]$, and let $c = u(a)$ and $d = u(b)$. Then*

$$\int_c^d f(x)\,dx = \int_a^b f(u(t))\frac{du(t)}{dt}\,dt.$$

*for all continuous real-valued functions $f$ on $[c,d]$.*

**Proof** Let $F$ and $G$ be the functions on $[a, b]$ defined by

$$F(x) = \int_c^{u(x)} f(y)dy, \qquad G(x) = \int_a^x f(u(t))\frac{du(t)}{dt}dt.$$

Then $F(a) = 0 = G(a)$. Moreover $F(x) = H(u(x))$, where

$$H(s) = \int_c^s f(y)\, dy,$$

and $H'(s) = f(s)$ for all $s \in [a, b]$. Using the Chain Rule and the Fundamental Theorem of Calculus, we deduce that

$$F'(x) = H'(u(x))u'(x) = f(u(x))u'(x) = G'(x)$$

for all $x \in (a, b)$. On applying the Mean Value Theorem (Theorem 7.2) to the function $F-G$ on the interval $[a, b]$, we see that $F(b)-G(b) = F(a)-G(a) = 0$. Thus $H(d) = F(b) = G(b)$, which yields the required identity. ∎

**Proposition 7.12** *Let $s$ and $h$ be real numbers, and let $f$ be a function whose first $k$ derivatives are continuous on an interval containing $s$ and $s+h$. Then*

$$f(s+h) \quad = \quad f(s) + \sum_{n=1}^{k-1} \frac{h^n}{n!} f^{(n)}(s) + \frac{h^k}{(k-1)!} \int_0^1 (1-t)^{k-1} f^{(k)}(s+th)\, dt.$$

**Proof** Let

$$r_m(s, h) = \frac{h^m}{(m-1)!} \int_0^1 (1-t)^{m-1} f^{(m)}(s+th)\, dt$$

for $m = 1, 2, \ldots, k-1$. Then

$$r_1(s, h) = h \int_0^1 f'(s+th)\, dt = \int_0^1 \frac{d}{dt} f(s+th)\, dt = f(s+h) - f(s).$$

Let $m$ be an integer between 1 and $k-2$. It follows from the rule for Integration by Parts (Corollary 7.10) that

$$\begin{aligned} r_{m+1}(s, h) &= \frac{h^{m+1}}{m!} \int_0^1 (1-t)^m f^{(m+1)}(s+th)\, dt \\ &= \frac{h^m}{m!} \int_0^1 (1-t)^m \frac{d}{dt} \left( f^{(m)}(s+th) \right)\, dt \\ &= \frac{h^m}{m!} \left[ (1-t)^m f^{(m)}(s+th) \right]_0^1 \end{aligned}$$

$$
\begin{aligned}
& -\frac{h^m}{m!}\int_0^1 \frac{d}{dt}\left((1-t)^m\right) f^{(m)}(s+th)\,dt \\
&= -\frac{h^m}{m!}f^{(m)}(s) + \frac{h^m}{(m-1)!}\int_0^1 (1-t)^{m-1} f^{(m)}(s+th)\,dt \\
&= r_m(s,h) - \frac{h^m}{m!}f^{(m)}(s).
\end{aligned}
$$

Thus

$$
r_m(s,h) = \frac{h^m}{m!}f^{(m)}(s) + r_{m+1}(s,h)
$$

for $m = 1, 2, \ldots, k-1$. It follows by induction on $k$ that

$$
\begin{aligned}
f(s+h) &= f(s) + \sum_{n=1}^{k-1} \frac{h^n}{n!} f^{(n)}(s) + r_k(s,h) \\
&= f(s) + \sum_{n=1}^{k-1} \frac{h^n}{n!} f^{(n)}(s) + \frac{h^k}{(k-1)!}\int_0^1 (1-t)^{k-1} f^{(k)}(s+th)\,dt,
\end{aligned}
$$

as required. ∎

# 8 Norms of Linear Transformations

## 8.1 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$T_{i,1}x_1 + T_{i,2}x_2 + \cdots + T_{i,m}x_m.$$

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

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

then

$$\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} T_{1,1} & T_{1,2} & \ldots & T_{1,m} \\ T_{2,1} & T_{2,2} & \ldots & T_{2,m} \\ \vdots & \vdots & & \vdots \\ T_{n,1} & T_{n,2} & \ldots & T_{m,n} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}.$$

## 8.2 The Operator Norm of a Linear Transformation

**Definition** Given $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. The *operator norm* $\|T\|_{\mathrm{op}}$ of $T$ is defined such that

$$\|T\|_{\mathrm{op}} = \sup\{|T\mathbf{x}| : \mathbf{x} \in \mathbb{R}^m \quad \text{and} \quad |\mathbf{x}| = 1\}.$$

**Lemma 8.1** *Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ and $U\colon \mathbb{R}^m \to \mathbb{R}^n$ be linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let $\lambda$ be a real number. Then $\|T\|_{\mathrm{op}}$ is the smallest non-negative real number with the property that $|T\mathbf{x}| \leq \|T\|_{\mathrm{op}}|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$. Moreover*

$$\|\lambda T\|_{\mathrm{op}} = |\lambda|\, \|T\|_{\mathrm{op}} \quad \textit{and} \quad \|T + U\|_{\mathrm{op}} \leq \|T\|_{\mathrm{op}} + \|U\|_{\mathrm{op}}.$$

**Proof** Let $\mathbf{x}$ be an element of $\mathbb{R}^m$. Then we can express $\mathbf{x}$ in the form $\mathbf{x} = \mu\mathbf{z}$, where $\mu = |\mathbf{x}|$ and $\mathbf{z} \in \mathbb{R}^m$ satisfies $|\mathbf{z}| = 1$. Then

$$|T\mathbf{x}| = |T(\mu\mathbf{z})| = |\mu T\mathbf{z}| = |\mu|\, |T\mathbf{z}| = |\mathbf{x}|\, |T\mathbf{z}| \leq \|T\|_{\mathrm{op}}|\mathbf{x}|.$$

Next let $C$ be a non-negative real number with the property that $|T\mathbf{x}| \leq C|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$. Then $C$ is an upper bound for the set

$$\{|T\mathbf{x}| : \mathbf{x} \in \mathbb{R}^m \quad \text{and} \quad |\mathbf{x}| = 1\},$$

and thus $\|T\|_{\mathrm{op}} \leq C$. Thus $\|T\|_{\mathrm{op}}$ is the smallest non-negative real number $C$ with the property that $|T\mathbf{x}| \leq C|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$.

Next we note that

$$\begin{aligned}
\|\lambda T\|_{\mathrm{op}} &= \sup\{|\lambda T\mathbf{x}| : \mathbf{x} \in \mathbb{R}^m \quad \text{and} \quad |\mathbf{x}| = 1\} \\
&= \sup\{|\lambda|\,|T\mathbf{x}| : \mathbf{x} \in \mathbb{R}^m \quad \text{and} \quad |\mathbf{x}| = 1\} \\
&= |\lambda| \sup\{|T\mathbf{x}| : \mathbf{x} \in \mathbb{R}^m \quad \text{and} \quad |\mathbf{x}| = 1\} \\
&= |\lambda|\,\|T\|_{\mathrm{op}}.
\end{aligned}$$

Let $\mathbf{x} \in \mathbb{R}^m$. Then

$$\begin{aligned}
|(T+U)\mathbf{x}| &\leq |T\mathbf{x}| + |U\mathbf{x}| \leq \|T\|_{\mathrm{op}}|\mathbf{x}| + \|U\|_{\mathrm{op}}|\mathbf{x}| \\
&\leq (\|T\|_{\mathrm{op}} + \|U\|_{\mathrm{op}})|\mathbf{x}|
\end{aligned}$$

It follows that

$$\|(T+U)\|_{\mathrm{op}} \leq \|T\|_{\mathrm{op}} + \|U\|_{\mathrm{op}}.$$

This completes the proof. ∎

## 8.3 The Hilbert-Schmidt Norm of a Linear Transformation

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

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

Note that the Hilbert-Schmidt norm is just the Euclidean norm on the real vector space of dimension $mn$ whose elements are $n \times m$ matrices representing linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$ 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 2.2) that

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

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

**Lemma 8.2** *Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. 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 $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$, where $\|T\|_{HS}$ is the Hilbert-Schmidt norm of the linear transformation $T$.*

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

$$v_i = T_{i,1}(x_1 - y_1) + T_{i,2}(x_2 - y_2) + \cdots + T_{i,m}(x_m - y_m)$$

for all integers $i$ between 1 and $n$. It follows from Schwarz's Inequality (Lemma 2.1) that

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

Hence

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

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 $\varepsilon$ be given. We can then choose $\delta$ 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 $T\colon \mathbb{R}^m \to \mathbb{R}^n$ is uniformly continuous on $\mathbb{R}^m$, as required. ∎

**Lemma 8.3** *Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ and let $S\colon \mathbb{R}^n \to \mathbb{R}^p$ be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^p$. 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 $k$th row and the $j$th column of the $p \times m$ matrix representing the linear transformation $ST$ satisfies

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

where $S_{k,i}$ and $T_{i,j}$ denote the components in the relevant rows and columns of the matrices representing the linear transformations $S$ and $T$ respectively. It follows from Schwarz's Inequality (Lemma 2.1) that

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

Summing over $k$, we find that

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

Then summing over $j$, we find that

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

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

**Proposition 8.4** *Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. Then the operator norm and the Hilbert-Schmidt norm of the linear operator $T$ satisfies the inequalities linear operators $T$ and $S$ satisfies the inequality*

$$\|T\|_{\mathrm{op}} \leq \|T\|_{\mathrm{HS}} \leq \sqrt{n}\,\|T\|_{\mathrm{op}}.$$

**Proof** It follows from Lemma 8.2 that that $|T\mathbf{x}| \leq \|T\|_{\mathrm{HS}}|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$. Now the operator norm $\|T\|_{\mathrm{op}}$ of $T$ is by definition the smallest non-negative real number with the property that $|T\mathbf{x}| \leq \|T\|_{\mathrm{op}}|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$. It follows that $\|T\|_{\mathrm{op}} \leq \|T\|_{\mathrm{HS}}$.

We denote by $T_{i,j}$ the coefficient in the $i$th row and $j$th column of the matrix representing the linear transformation $T$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

Let $i$ be an integer between 1 and $n$, and let

$$\mathbf{x} = (T_{i,1}, T_{i,2}, \ldots, T_{i,m}).$$

Then the $i$th component $(T\mathbf{x})_i$ of the vector $T\mathbf{x}$ satisfies the equation

$$(T\mathbf{x})_i = \sum_{j=1}^{m} T_{i,j}^2.$$

It follows that

$$\sum_{j=1}^{m} T_{i,j}^2 \leq |T\mathbf{x}| \leq \|T\|_{\mathrm{op}} |\mathbf{x}| = \|T\|_{\mathrm{op}} \sqrt{\sum_{j=1}^{m} T_{i,j}^2}.$$

Therefore

$$\sum_{j=1}^{m} T_{i,j}^2 \leq \|T\|_{\mathrm{op}}^2,$$

and hence

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

The result follows. ∎

The Hilbert-Schmidt norm on the real vector space $L(\mathbb{R}^m, \mathbb{R}^n)$ of linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$ is just the Euclidean norm on that vector space obtained on identifying that vector space with $\mathbb{R}^{mn}$ by means of its natural basis. The definition of convergence in a Euclidean space then ensures that an infinite sequence $T_1, T_2, T_3, \ldots$, of linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$ converges to some linear transformation $T$ if and only if, given any positive real number $\eta$, there exists some positive integer $N$ such that $\|T_j - T\|_{\mathrm{HS}} < \eta$ whenever $j \geq N$. Similarly a subset $V$ of $L(\mathbb{R}^m, \mathbb{R}^n)$ is open in $L(\mathbb{R}^m, \mathbb{R}^n)$ if and only if, given any linear transformation $S\colon \mathbb{R}^m \to \mathbb{R}^n$ that belongs to $V$, there exists some strictly positive real number $\eta$ such that

$$\{T \in L(\mathbb{R}^m, \mathbb{R}^n) : \|T - S\|_{\mathrm{HS}} < \eta\} \subset V.$$

**Lemma 8.5** *Let $T_1, T_2, T_3, \ldots$ be an infinite sequence of linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let $T$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$. Then the infinite sequence $T_1, T_2, T_3, \ldots$ of linear transformations converges to the linear transformation $T$ if and only if, given any positive real number $\varepsilon$, there exists some positive integer $N$ such that positive integer $N$ such that $\|T_j - T\|_{\mathrm{op}} < \varepsilon$ whenever $j \geq N$.*

**Proof** First suppose that the infinite sequence $T_1, T_2, T_3, \ldots$ converges to $T$. Then, given any positive real number $\varepsilon$, there exists some positive integer $N$ such that $\|T_j - T\|_{\mathrm{HS}} < \varepsilon$ whenever $j \geq N$. It then follows from Proposition 8.4 that

$$\|T_j - T\|_{\mathrm{op}} \leq \|T_j - T\|_{\mathrm{HS}} < \varepsilon$$

whenever $j \geq N$.

Conversely, suppose that, given any positive real number $\eta$, there exists some positive integer $N$ such that $\|T_j - T\|_{\mathrm{op}} < \eta$ whenever $j \geq N$. Let some positive real number $\varepsilon$ be given. Then there exists some positive integer $N$ such that $\|T_j - \|_{\mathrm{op}} < \varepsilon/\sqrt{n}$ whenever $j \geq N$. It then follows from Proposition 8.4 that

$$\|T_j - T\|_{\mathrm{HS}} \leq \sqrt{n}\, \|T_j - T\|_{\mathrm{op}} < \varepsilon$$

whenever $j \geq N$, and thus $T_j$ converges to $T$ as $j \to +\infty$. ∎

**Lemma 8.6** *Let $V$ be a subset of the set $L(\mathbb{R}^m, \mathbb{R}^n)$ of linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$. Then $V$ is open in $L(\mathbb{R}^m, \mathbb{R}^n)$ if and only if, given any element $S$ of $V$, there exists some positive real number $\varepsilon$ such that*

$$\{T \in L(\mathbb{R}^m, \mathbb{R}^n) : \|T - S\|_{\mathrm{op}} < \varepsilon\} \subset V$$

**Proof** The set $V$ is open in $L(\mathbb{R}^m, \mathbb{R}^n)$ if and only if, given any element $S$ of $V$, there exists some positive real number $\eta$ such that

$$\{T \in L(\mathbb{R}^m, \mathbb{R}^n) : \|T - S\|_{\mathrm{HS}} < \eta\} \subset V.$$

Let some positive real number $\varepsilon$ be given. It follows from Proposition 8.4 that if the set $V$ contains all linear transformations $T$ from $\mathbb{R}^m$ to $\mathbb{R}^n$ that satisfy $\|T - S\|_{\mathrm{op}} < \varepsilon$ then it contains all linear transformations $T$ that satisfy $\|T - S\|_{\mathrm{HS}} < \varepsilon$, because $\|T - S\|_{\mathrm{op}} \leq \|T - S\|_{\mathrm{HS}}$. In the other direction, if the set $V$ contains all linear transformations that satisfy $\|T - S\|_{\mathrm{HS}} < \varepsilon$ then it contains all linear transformations $T$ that satisfy $\|T - S\|_{\mathrm{op}} < \varepsilon/\sqrt{n}$, because $\|T - S\|_{\mathrm{HS}} \leq \sqrt{n}\, \|T - S\|_{\mathrm{op}}$. The result follows. ∎

# 9 Differentiation of Functions of Several Real Variables

## 9.1 Observations on the Concept of Differentiability

Let us consider the definition of differentiability for functions of a single real variable. Let $D$ be a subset of the set $\mathbb{R}$ of real numbers, and let $p$ be a value in the interior of $D$. The function $f$ is then said to be differentiable at $p$, with derivative $f'(p)$, if and only if the limit of the *difference quotient*

$$\frac{f(p+h) - f(p)}{h}$$

as $h \to 0$ exists and is equal to $f'(p)$.

We wish to extend the definition of differentiability to real-valued functions $f\colon D \to \mathbb{R}$ defined on subsets $D$ of $m$-dimensional Euclidean space $\mathbb{R}^m$. Let $\mathbf{p}$ be a point in the interior of $D$. Then $\mathbf{p} + \mathbf{h} \in D$ for all $m$-dimensional vectors $\mathbf{h}$ that are sufficiently close to the zero vector in $\mathbb{R}^m$. Then given any non-zero $m$-dimensional vector $\mathbf{h}$ for which $\mathbf{p} + \mathbf{h} \in D$, the difference $f(\mathbf{p} + \mathbf{h}) - f(\mathbf{p})$ is a real number. *And, in dimensions greater than two, there is no algebraic operation for dividing real numbers by non-zero vectors that is compatible with the usual "laws" of algebra* such as the Commutative Laws, Associative Laws and Distributive Law satisfied by the operations of addition, subtraction, multiplication and division within the fields of real and complex numbers).

Thus the *derivative of a function of several variables cannot be defined as a limit of difference quotients.*

Next we consider partial derivatives. Let $f\colon D \to \mathbb{R}$ be a function defined on a subset $D$ of $\mathbb{R}^m$. On might be tempted to define differentiability by saying that the function $f$ is differentiable at a point ı in the interior of $D$ if and only if all the partial derivatives of $f$ exist at the point $\mathbf{p}$. However *mathematicians do not define differentiability for functions of several variables in this fashion*, and the following example demonstrates why they do not do so.

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

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

The usual propositions and rules of calculus ensure that

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

are defined for all points $(x, y)$ of $\mathbb{R}^2$ that are distinct from the origin $(0, 0)$. Also

$$\left.\frac{\partial f(x,y)}{\partial x}\right|_{(x,y)=(0,0)} = \lim_{h\to 0}\frac{f(h,0)-f(0,0)}{h} = 0,$$

$$\left.\frac{\partial f(x,y)}{\partial y}\right|_{(x,y)=(0,0)} = \lim_{h\to 0}\frac{f(0,h)-f(0,0)}{h} = 0,$$

because $f(h, 0) = 0$ and $f(0, h) = 0$ for all non-zero real numbers $h$. Thus the partial derivatives of the function $f$ exist at every point of $\mathbb{R}^2$.

But now let us consider the behaviour of the function $f$ along the line $x = y$. Now

$$f(t,t) = \begin{cases} \dfrac{1}{4t^2} & \text{if } t \neq 0; \\ 0 & \text{if } t = 0. \end{cases}$$

It follows that $f(t, t)$ increases without limit at $t \to 0$, and therefore the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ is not continuous at the origin $(0, 0)$.

This example demonstrates that, were mathematicians to take the existence of well-defined partial derivatives as the criterion for differentiability, then many functions would be differentiable that were not continuous. The following lemma however provides a characterization of differentiability for functions of a single real variable that can be generalized directly so as to apply to functions of several real variables.

**Lemma 9.1** *Let $f\colon D \to \mathbb{R}$ be a real-valued function defined on some subset $D$ of the set of real numbers. Let $s$ be a real number in the interior of $D$. The function $f$ is differentiable at $s$ with derivative $f'(s)$ (where $f'(s)$ is some real number) if and only if*

$$\lim_{x\to s}\frac{1}{|x-s|}\left(f(x) - f(s) - f'(s)(x-s)\right) = 0.$$

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

$$\lim_{x\to s}\frac{f(x)-f(s)}{x-s} = f'(s)$$

if and only if

$$\lim_{x\to s}\left|\frac{f(x)-f(s)}{x-s} - f'(s)\right| = 0.$$

But

$$\left|\frac{f(x)-f(s)}{x-s} - f'(s)\right| = \frac{1}{|x-s|}\left|f(x) - f(s) - f'(s)(x-s)\right|.$$

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

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

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 $f'(s)$ of $f$ at $s$ as representing a linear transformation $h \mapsto f'(s)h$, characterized by the property that the map

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

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

$$\lim_{x \to s} \frac{e(x)}{|x-s|} = 0,$$

where

$$e(x) = f(x) - f(s) - f'(s)(x-s)$$

(i.e., $e(x)$ measures the difference between $f(x)$ and the value $f(s)+f'(s)(x-s)$ of the approximation at $x$, and thus provides a measure of the error of this approximation).

We shall generalize the notion of differentiability to functions $\varphi$ from $\mathbb{R}^m$ to $\mathbb{R}^n$ by defining the derivative $(D\varphi)_{\mathbf{p}}$ of $\varphi$ at $\mathbf{p}$ to be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ characterized by the property that the map

$$\mathbf{x} \mapsto \varphi(\mathbf{p}) + (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})$$

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

## 9.2 Derivatives of Functions of Several Variables

**Definition** Let $X$ be an open subset of $\mathbb{R}^m$ and let $\varphi: X \to \mathbb{R}^n$ be a map from $X$ into $\mathbb{R}^n$. Let $\mathbf{p}$ be a point of $X$. The function $\varphi$ is said to be *differentiable* at $\mathbf{p}$, with *derivative* $(D\varphi)_{\mathbf{p}}: \mathbb{R}^m \to \mathbb{R}^n$, where $(D\varphi)_{\mathbf{p}}$ is a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, if and only if

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

The derivative of a map $\varphi: X \to \mathbb{R}^n$ defined on a open subset $X$ of $\mathbb{R}^m$ at a point $\mathbf{p}$ of $X$ is usually denoted either by $(D\varphi)_\mathbf{p}$ or else by $\varphi'(\mathbf{p})$.

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

**Lemma 9.2** *Let $T: \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ into $\mathbb{R}^n$. Then $T$ is differentiable at each point $\mathbf{p}$ of $\mathbb{R}^m$, and $(DT)_\mathbf{p} = T$.*

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

**Lemma 9.3** *Let $\varphi: X \to \mathbb{R}^n$ be a function, let $T: \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, and let $\Omega: X \to \mathbb{R}^n$ be defined so that*

$$\Omega(\mathbf{x}) = \begin{cases} \dfrac{1}{|\mathbf{x} - \mathbf{p}|} \left(\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p})\right) & \textit{if } \mathbf{x} \neq \mathbf{p}; \\ \mathbf{0} & \textit{if } \mathbf{x} = \mathbf{p}. \end{cases}$$

*Then $\varphi: X \to \mathbb{R}^n$ is differentiable at $\mathbf{p}$ with derivative $T: \mathbb{R}^m \to \mathbb{R}^n$ if and only if $\lim_{\mathbf{x} \to \mathbf{p}} \Omega(\mathbf{x}) = \mathbf{0} = \Omega(\mathbf{p})$. Thus the function $\varphi: X \to \mathbb{R}^n$ is differentiable at $\mathbf{p}$, with derivative $T$, if and only if the associated function $\Omega: X \to \mathbb{R}^n$ is continuous at $\mathbf{p}$*

**Proof** It follows from the definition of differentiability that the function $\varphi$ is differentiable, with derivative $T: \mathbb{R}^m \to \mathbb{R}^n$, if and only if $\lim_{\mathbf{x} \to \mathbf{p}} \Omega(\mathbf{x}) = \mathbf{0}$. But $\Omega(\mathbf{p}) = \mathbf{0}$. It follows that $\lim_{\mathbf{x} \to \mathbf{p}} \Omega(\mathbf{x}) = \mathbf{0}$ if and only if the function $\Omega$ is continuous at $\mathbf{p}$ (see Proposition 4.17). The result follows. ∎

**Example** Let $\varphi: \mathbb{R}^2 \to \mathbb{R}^2$ 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 $p$ and $q$ be fixed real numbers. Then

$$\begin{aligned} &\varphi\left(\begin{pmatrix} x \\ y \end{pmatrix}\right) - \varphi\left(\begin{pmatrix} p \\ q \end{pmatrix}\right) \\ &\quad = \begin{pmatrix} x^2 - y^2 \\ 2xy \end{pmatrix} - \begin{pmatrix} p^2 - q^2 \\ 2pq \end{pmatrix} \\ &\quad = \begin{pmatrix} (x+p)(x-p) - (y+q)(y-q) \\ 2q(x-p) + 2p(y-q) + 2(x-p)(y-q) \end{pmatrix} \\ &\quad = \begin{pmatrix} 2p(x-p) - 2q(y-q) + (x-p)^2 - (y-q)^2 \\ 2q(x-p) + 2p(y-q) + 2(x-p)(y-q) \end{pmatrix} \\ &\quad = \begin{pmatrix} 2p & -2q \\ 2q & 2p \end{pmatrix} \begin{pmatrix} x-p \\ y-q \end{pmatrix} + \begin{pmatrix} (x-p)^2 - (y-q)^2 \\ 2(x-p)(y-q) \end{pmatrix}. \end{aligned}$$

Now, given $(x, y) \in \mathbb{R}^2$, let $r = \sqrt{(x-p)^2 + (y-q)^2}$. Then $|x - p| < r$ and $|y - q| < r$, and therefore

$$|(x-p)^2 - (y-q)^2| \leq |x-p|^2 + |y-q|^2 < 2r^2$$

and $2(x-p)(y-q) < 2r^2$, and thus

$$\frac{(x-p)^2 - (y-q)^2}{\sqrt{(x-p)^2 + (y-q)^2}} < 2r \quad \text{and} \quad \frac{2(x-p)(y-q)}{\sqrt{(x-p)^2 + (y-q)^2}} < 2r.$$

Thus, given any positive real number $\varepsilon$, let $\delta = \frac{1}{2}\varepsilon$. Then

$$\left| \frac{(x-p)^2 - (y-q)^2}{\sqrt{(x-p)^2 + (y-q)^2}} \right| < \varepsilon \quad \text{and} \quad \left| \frac{2(x-p)(y-q)}{\sqrt{(x-p)^2 + (y-q)^2}} \right| < \varepsilon$$

whenever $0 < |(x, y) - (p, q)| < \delta$. It follows therefore that

$$\lim_{(x,y)\to(0,0)} \frac{1}{\sqrt{(x-p)^2 + (y-q)^2}} \begin{pmatrix} (x-p)^2 - (y-q)^2 \\ 2(x-p)(y-q) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Thus the function $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ is differentiable at $(p, q)$, and the derivative of this function at $(p, q)$ is the linear transformation represented by the matrix

$$\begin{pmatrix} 2p & -2q \\ 2q & 2p \end{pmatrix}.$$

## 9.3 Properties of Differentiable Functions of Several Real Variables

**Lemma 9.4** *Let $\varphi\colon X \to \mathbb{R}^n$ be a function which maps an open subset $X$ of $\mathbb{R}^m$ into $\mathbb{R}^n$ which is differentiable at some point $\mathbf{p}$ of $X$. Then $\varphi$ is continuous at $\mathbf{p}$.*

**Proof** Let $\Omega\colon X \to \mathbb{R}^n$ be defined so that $\Omega(\mathbf{p}) = 0$ and

$$\Omega(\mathbf{x}) = \frac{1}{|\mathbf{x} - \mathbf{p}|} \left( \varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}} (\mathbf{x} - \mathbf{p}) \right)$$

for all points $\mathbf{x}$ of $X$ satisfying $\mathbf{x} \neq \mathbf{p}$. If $\varphi\colon X \to \mathbb{R}^n$ is differentiable at $\mathbf{p}$ then $\Omega\colon X \to \mathbb{R}^n$ is continuous at $\mathbf{p}$ (see Lemma 9.3). Moreover

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

for all $\mathbf{x} \in X$. It follows that $\varphi\colon X \to \mathbb{R}^n$ is continuous at $\mathbf{p}$, as required. ∎

**Lemma 9.5** *Let $\varphi: X \to \mathbb{R}^n$ be a function which maps an open subset $X$ of $\mathbb{R}^m$ into $\mathbb{R}^n$ which is differentiable at some point $\mathbf{p}$ of $X$. Let $(D\varphi)_{\mathbf{p}}: \mathbb{R}^m \to \mathbb{R}^n$ be the derivative of $\varphi$ at $\mathbf{p}$. Let $\mathbf{u}$ be an element of $\mathbb{R}^m$. 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 $\varphi$ at $\mathbf{p}$ is uniquely determined by the map $\varphi$.*

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

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

In particular, if we set $(\mathbf{x} - \mathbf{p}) = t\mathbf{u}$, and $(\mathbf{x} - \mathbf{p}) = -t\mathbf{u}$, 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

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

as required. ∎

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

**Proposition 9.6** *Let $X$ be an open set in $\mathbb{R}^m$, and let $f: X \to \mathbb{R}$ and $g: X \to \mathbb{R}$ be functions mapping $X$ into $\mathbb{R}$. Let $\mathbf{p}$ be a point of $X$. Suppose that $f$ and $g$ are differentiable at $\mathbf{p}$. Then the functions $f+g$ and $f-g$ are differentiable at $\mathbf{p}$, and*

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

*and*

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

**Proof** The limit of a sum of functions is the sum of the limits of those functions, provided that these limits exist. Applying the definition of differentiability, it therefore follows that

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

Therefore

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

Also the function $-g$ is differentiable, with derivative $-(Dg)_{\mathbf{p}}$. It follows that $f-g$ is differentiable, with derivative $(Df)_{\mathbf{p}} - (Dg)_{\mathbf{p}}$. This completes the proof. ∎

## 9.4 The Multidimensional Product Rule

**Proposition 9.7 (Product Rule)** *Let $X$ be an open set in $\mathbb{R}^m$, and let $f\colon X\to\mathbb{R}$ and $g\colon X\to\mathbb{R}$ be functions mapping $X$ into $\mathbb{R}$. Let $\mathbf{p}$ be a point of $X$. Suppose that $f$ and $g$ are differentiable at $\mathbf{p}$. Then the function $f\cdot g$ is differentiable at $\mathbf{p}$, and*

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

**Proof** The functions $f$ and $g$ are differentiable at $\mathbf{p}$, and therefore there are well-defined functions $Q_1\colon X\to\mathbb{R}$ and $Q_2\colon X\to\mathbb{R}$, where

$$\lim_{\mathbf{x}\to\mathbf{p}} Q_1(\mathbf{x}) = 0 = Q_1(\mathbf{p}) \quad\text{and}\quad \lim_{\mathbf{x}\to\mathbf{p}} Q_2(\mathbf{x}) = 0 = Q_2(\mathbf{p}),$$

that are defined throughout $X$ so as to ensure that

$$f(\mathbf{x}) = f(\mathbf{p}) + (Df)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) + |\mathbf{x}-\mathbf{p}|\, Q_1(\mathbf{x})$$

and

$$g(\mathbf{x}) = g(\mathbf{p}) + (Dg)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) + |\mathbf{x}-\mathbf{p}|\, Q_2(\mathbf{x})$$

for all $\mathbf{x}\in X$ (see Lemma 9.3).

Then

$$\begin{aligned} f(\mathbf{x})g(\mathbf{x}) &= f(\mathbf{p})g(\mathbf{p}) \\ &\qquad + \Big(g(\mathbf{p})\,(Df)_{\mathbf{p}} + f(\mathbf{p})\,(Dg)_{\mathbf{p}}\Big)(\mathbf{x}-\mathbf{p}) \\ &\qquad + |\mathbf{x}-\mathbf{p}|\,Q(\mathbf{x}) \end{aligned}$$

where

$$\begin{aligned} Q(\mathbf{x}) &= \frac{1}{|\mathbf{x}-\mathbf{p}|}(Df)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) \times (Dg)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) \\ &\qquad + (g(\mathbf{p}) + (Dg)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}))Q_1(\mathbf{x}) \\ &\qquad + (f(\mathbf{p}) + (Df)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}))Q_2(\mathbf{x}) \\ &\qquad + |\mathbf{x}-\mathbf{p}|\,Q_1(\mathbf{x})Q_2(\mathbf{x}). \end{aligned}$$

Now

$$|(Df)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})| \leq \|(Df)_{\mathbf{p}}\|_{\mathrm{op}}|\mathbf{x}-\mathbf{p}|$$

where $\|(Df)_{\mathbf{p}}\|_{\mathrm{op}}$ denotes the operator norm of $(Df)_{\mathbf{p}}$ (see Lemma 8.1) Similarly

$$|(Dg)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})| \leq \|(Dg)_{\mathbf{p}}\|_{\mathrm{op}}|\mathbf{x}-\mathbf{p}|.$$

It follows that

$$\begin{aligned} &\left|\frac{1}{|\mathbf{x}-\mathbf{p}|}(Df)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) \times (Dg)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})\right| \\ &\qquad\qquad \leq \|(Df)_{\mathbf{p}}\|_{\mathrm{op}}\|(Dg)_{\mathbf{p}}\|_{\mathrm{op}}|\mathbf{x}-\mathbf{p}|, \end{aligned}$$

and therefore

$$\lim_{\mathbf{x}\to\mathbf{p}} \left(\frac{1}{|\mathbf{x}-\mathbf{p}|}(Df)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}) \times (Dg)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})\right) = 0.$$

Next we note that

$$\begin{aligned} &\lim_{\mathbf{x}\to\mathbf{p}}\Big((g(\mathbf{p}) + (Dg)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}))Q_1(\mathbf{x})\Big) \\ &\qquad = \lim_{\mathbf{x}\to\mathbf{p}}(g(\mathbf{p}) + (Dg)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})) \times \lim_{\mathbf{x}\to\mathbf{p}} Q_1(\mathbf{x}) = 0, \end{aligned}$$

because $\lim_{\mathbf{x}\to\mathbf{p}} Q_1(\mathbf{x}) = 0$.

Similarly

$$\begin{aligned} &\lim_{\mathbf{x}\to\mathbf{p}}\Big((f(\mathbf{p}) + (Df)_{\mathbf{p}}(\mathbf{x}-\mathbf{p}))Q_2(\mathbf{x})\Big) \\ &\qquad = \lim_{\mathbf{x}\to\mathbf{p}}(f(\mathbf{p}) + (Df)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})) \times \lim_{\mathbf{x}\to\mathbf{p}} Q_2(\mathbf{x}) = 0, \end{aligned}$$

because $\lim_{\mathbf{x}\to\mathbf{p}} Q_2(\mathbf{x}) = 0$.

The quantities $Q_1(\mathbf{x})$ and $Q_2(\mathbf{x})$ converge to zero and therefore remain bounded as $\mathbf{x}$ tends to $\mathbf{p}$. It follows that

$$\lim_{\mathbf{x}\to\mathbf{p}} |\mathbf{x}-\mathbf{p}|\, Q_1(\mathbf{x})Q_2(\mathbf{x}) = 0.$$

Putting these results together, we see that

$$\lim_{\mathbf{x}\to\mathbf{p}} Q(\mathbf{x}) = 0.$$

It follows from this that the function $f \cdot g$ is differentiable at $\mathbf{p}$, and

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

(see Lemma 9.3). This completes the proof. ∎

## 9.5 The Multidimensional Chain Rule

**Proposition 9.8 (Chain Rule)** *Let $X$ be an open set in $\mathbb{R}^m$, and let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping $X$ into $\mathbb{R}^n$. Let $Y$ be an open set in $\mathbb{R}^n$ which contains $\varphi(X)$, and let $\psi\colon Y \to \mathbb{R}^k$ be a function mapping $Y$ into $\mathbb{R}^k$. Let $\mathbf{p}$ be a point of $X$. Suppose that $\varphi$ is differentiable at $\mathbf{p}$ and that $\psi$ is differentiable at $\varphi(\mathbf{p})$. Then the composition $\psi\circ\varphi\colon \mathbb{R}^m \to \mathbb{R}^k$ (i.e., $\varphi$ followed by $\psi$) 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 $\mathbf{q} = \varphi(\mathbf{p})$. The functions $\varphi\colon X \to \mathbb{R}^n$ and $\psi\colon Y \to \mathbb{R}^k$ are differentiable at $\mathbf{p}$ and $\mathbf{q}$ respectively, and therefore there are well-defined functions $\Omega_1\colon X \to \mathbb{R}^n$ and $\Omega_2\colon Y \to \mathbb{R}^k$ that are defined throughout $X$ and $Y$ respectively so as to ensure that

$$\lim_{\mathbf{x}\to\mathbf{p}} \Omega_1(\mathbf{x}) = \mathbf{0} = \Omega_1(\mathbf{p}), \quad \lim_{\mathbf{y}\to\mathbf{q}} \Omega_2(\mathbf{y}) = \mathbf{0} = \Omega_2(\mathbf{q})$$

for all $\mathbf{x} \in X$, and

$$\varphi(\mathbf{x}) = \varphi(\mathbf{p}) + (D\varphi)_{\mathbf{p}}\,(\mathbf{x}-\mathbf{p}) + |\mathbf{x}-\mathbf{p}|\,\Omega_1(\mathbf{x})$$

and

$$\psi(\mathbf{y}) = \psi(\mathbf{q}) + (D\psi)_{\mathbf{q}}\,(\mathbf{y} - \mathbf{q}) + |\mathbf{y} - \mathbf{q}|\,\Omega_2(\mathbf{y})$$

for all $\mathbf{y} \in Y$ (see Lemma 9.3).

Substituting $\varphi(\mathbf{x})$ and $\varphi(\mathbf{p})$ for $\mathbf{y}$ and $\mathbf{q}$ respectively, we find that

$$\begin{aligned}\psi(\varphi(\mathbf{x})) &= \psi(\varphi(\mathbf{p})) + (D\psi)_{\mathbf{q}}(\varphi(\mathbf{x}) - \varphi(\mathbf{p})) \\ &\qquad + |\varphi(\mathbf{x}) - \varphi(\mathbf{p})|\,\Omega_2(\varphi(\mathbf{x})) \\ &= \psi(\varphi(\mathbf{p})) + (D\psi)_{\varphi(\mathbf{p})}((D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})) \\ &\qquad + |\mathbf{x} - \mathbf{p}|\,\Omega(\mathbf{x}),\end{aligned}$$

where

$$\begin{aligned}\Omega(\mathbf{x}) &= (D\psi)_{\varphi(\mathbf{p})}(\Omega_1(\mathbf{x})) \\ &\qquad + \left|\frac{1}{|\mathbf{x} - \mathbf{p}|}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) + \Omega_1(\mathbf{x})\right|\,\Omega_2(\varphi(\mathbf{x})).\end{aligned}$$

Let

$$M(\mathbf{x}) = \left|\frac{1}{|\mathbf{x} - \mathbf{p}|}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p}) + \Omega_1(\mathbf{x})\right|$$

for all $\mathbf{x} \in X$ satisfying $\mathbf{x} \neq \mathbf{p}$. Then

$$0 \leq M(\mathbf{x}) \leq \frac{|(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})|}{|\mathbf{x} - \mathbf{p}|} + |\Omega_1(\mathbf{x})|$$

for all $\mathbf{x} \in X$ satisfying $\mathbf{x} \neq \mathbf{p}$. Moreover

$$|(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| \leq \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}|\mathbf{x} - \mathbf{p}|,$$

where $\|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$ denotes the operator norm of the linear operator $(D\varphi)_{\mathbf{p}}$ (see Lemma 8.1). It follows that

$$0 \leq M(x) \leq \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}} + |\Omega_1(\mathbf{x})|$$

for all $\mathbf{x} \in X$ satisfying $\mathbf{x} \neq \mathbf{p}$. It follows from the continuity of the function $\Omega_1$ at $\mathbf{p}$ that $M(\mathbf{x})$ remains bounded as $\mathbf{x}$ tends to $\mathbf{p}$ in $X$. Now

$$\Omega(\mathbf{x}) = (D\psi)_{\varphi(\mathbf{p})}(\Omega_1(\mathbf{x})) + M(\mathbf{x})\Omega_2(\varphi(\mathbf{x}))$$

Also the function $\varphi\colon X \to \mathbb{R}^n$ is continuous at $\mathbf{p}$ and the function $\Omega_2\colon Y \to \mathbb{R}^k$ is continuous at $\varphi(\mathbf{p})$. It follows that the composition function $\Omega_2 \circ \varphi$ is continuous at $\mathbf{p}$ (see Lemma 4.1), and therefore

$$\lim_{\mathbf{x}\to\mathbf{p}} \Omega_2(\varphi(\mathbf{x})) = \Omega_2(\varphi(\mathbf{p})) = \mathbf{0}.$$

We have already shown that $M(\mathbf{x})$ remains bounded as $\mathbf{x}$ tends to $\mathbf{p}$ in $X$. It follows that

$$\lim_{\mathbf{x}\to\mathbf{p}} (M(\mathbf{x})\Omega_2(\varphi(\mathbf{x})) = \mathbf{0}$$

(see Proposition 4.9).

Linear operators on finite-dimensional vector spaces are continuous. It follows that

$$\lim_{\mathbf{x}\to\mathbf{p}} (D\psi)_{\varphi(\mathbf{p})}(\Omega_1(\mathbf{x})) = (D\psi)_{\varphi(\mathbf{p})}\left(\lim_{\mathbf{x}\to\mathbf{p}} \Omega_1(\mathbf{x})\right) = \mathbf{0}.$$

It follows that

$$\begin{aligned}\lim_{\mathbf{x}\to\mathbf{p}} \Omega(\mathbf{x}) &= \lim_{\mathbf{x}\to\mathbf{p}} (D\psi)_{\varphi(\mathbf{p})}(\Omega_1(\mathbf{x})) + \lim_{\mathbf{x}\to\mathbf{p}} (M(\mathbf{x})\Omega_2(\varphi(\mathbf{x}))) \\ &= \mathbf{0} = \Omega(\mathbf{p}).\end{aligned}$$

This result ensures that the composition function $\psi \circ \varphi$ is differentiable at $\mathbf{p}$, and that

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

(see Lemma 9.3). The result follows. ∎

**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 9.2). Now $\varphi(x, y) = y^3 h(x)$. Using Proposition 9.6 and Proposition 9.8, we conclude that $\varphi$ is differentiable everywhere on $\mathbb{R}^2$.

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

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

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

$$\varphi(\mathbf{x}) = (f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_n(\mathbf{x}))$$

for all $\mathbf{x} \in X$. The $j$th partial derivative of $f_i$ at $\mathbf{p} \in X$ 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 $\varphi$ is differentiable at $\mathbf{p}$ then

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

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

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

This matrix is known as the *Jacobian matrix* of $\varphi$ 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 9.4 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 $g: \mathbb{R}^2 \to \mathbb{R}$ 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 $u_{b,c}: \mathbb{R} \to \mathbb{R}$ 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 $t \in \mathbb{R}$. If $b \neq 0$ and $c \neq 0$ then

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

We now show that the function $u_{b,c}: \mathbb{R} \to \mathbb{R}$ has derivatives of all orders. This is obvious when $b = 0$, and when $c = 0$. If $b$ and $c$ are both non-zero, and if the function $u_{b,c}$ has a derivative $u_{b,c}^{(k)}(t)$ of order $k$ that can be represented in the form

$$u_{b,c}^{(k)}(t) = p_k(t)(b^2 + c^2t^2)^{-k-1},$$

where $p_k(t)$ is a polynomial of degree at most $k+1$, then it follows from standard single-variable calculus that the function $u_{b,c}$ has a derivative $u_{b,c}^{(k+1)}(t)$ of order $k+1$ that can be represented in the form

$$u_{b,c}^{(k+1)}(t) = p_{k+1}(t)(b^2 + c^2t^2)^{-k-2},$$

where $p_{k+1}(t)$ is the polynomial of degree at most $k+2$ determined by the formula

$$p_{k+1}(t) = p_k'(t)(b^2 + c^2t^2) - 2(k+1)c^2tp_k(t).$$

Thus the function $u_{b,c}: \mathbb{R} \to \mathbb{R}$ has derivatives of all orders.

Moreover the first derivative $u_{b,c}'(0)$ of $u_{b,c}(t)$ at $t = 0$ is given by the formula

$$u_{b,c}'(0) = \begin{cases} \dfrac{c^2}{b} & \text{if } b \neq 0; \\ 0 & \text{if } b = 0. \end{cases}$$

We have shown that the restriction of the function $g: \mathbb{R}^2 \to \mathbb{R}$ to any line passing through the origin determines a function that may be differentiated any number of times with respect to distance along the line. Analogous arguments show that the restriction of the function $g$ to any other line in the plane also determines a function that may be differentiated any number of times with respect to distance along the line.

Now $g(x, y) = \frac{1}{2}$ 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. Therefore the function $g\colon \mathbb{R}^2 \to \mathbb{R}$ 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 we shall show 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.

## 9.6 Partial Derivatives and Continuous Differentiability

Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m$ denote the standard basis of $\mathbb{R}^m$, defined so that

$$(z_1, z_2, \ldots, z_m) = \sum_{j=1}^{m} z_j \mathbf{e}_j$$

for all $(z_1, z_2, \ldots, z_m) \in \mathbb{R}^m$. Similarly let $\overline{\mathbf{e}}_1, \overline{\mathbf{e}}_2, \ldots, \overline{\mathbf{e}}_n$ denote the standard basis of $\mathbb{R}^n$, defined so that

$$(w_1, w_2, \ldots, w_n) = \sum_{j=1}^{n} w_j \overline{\mathbf{e}}_j$$

for all $(w_1, w_2, \ldots, w_n) \in \mathbb{R}^n$.

Let $X$ be an open set in $\mathbb{R}^m$, and let $\varphi\colon X \to \mathbb{R}^n$ be a function from $X$ to $\mathbb{R}^n$ which is differentiable at some point $\mathbf{p}$ of $X$. Then the partial derivative of the $i$th component $f_i$ of the function $\varphi$ with respect to the $j$th coordinate function $x_j$ at a point $\mathbf{p}$ of $X$ is determined by the formula

$$\left.\frac{\partial f_i}{\partial x_j}\right|_{\mathbf{p}} = \overline{\mathbf{e}}_i . (D\varphi)_{\mathbf{p}} \mathbf{e}_j.$$

**Definition** Let $X$ be an open set in $\mathbb{R}^m$. A function $\varphi: X \to \mathbb{R}^n$ is *continuously differentiable* if the function sending each point $\mathbf{x}$ of $X$ to the derivative $(D\varphi)$ of $\varphi$ at the point $\mathbf{x}$ is a continuous function from $X$ to the vector space $L(\mathbb{R}^m, \mathbb{R}^n)$ of linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$.

A function of several real variables is said to be "$C^1$" if and only if it is continuously differentiable.

**Lemma 9.9** *Let $X$ be an open set in $\mathbb{R}^m$. and let $\varphi: X \to \mathbb{R}^n$ be a continuously differentiable function on $X$. Then the first order partial derivatives of the components of $\varphi$ exist and are continuous throughout $X$.*

**Proof** Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m$ be the basis vectors that determine the standard basis of $\mathbb{R}^m$ and let $\overline{\mathbf{e}}_1, \overline{\mathbf{e}}_2, \ldots, \overline{\mathbf{e}}_n$ be the basis vectors that determine the standard basis of $\mathbb{R}_n$. Then the partial derivative of the $i$th component $f_i$ of the function $\varphi$ with respect to the $j$th coordinate function $x_j$ at a point $\mathbf{p}$ of $X$ is determined by the formula

$$\left.\frac{\partial f_i}{\partial x_j}\right|_{\mathbf{p}} = \overline{\mathbf{e}}_i.(D\varphi)_{\mathbf{p}}\mathbf{e}_j.$$

It follows that if $(D\varphi)_{\mathbf{p}}$ is a continuous function of $\mathbf{p}$ then so are the partial derivatives of $\varphi$. ∎

## 9.7 Functions with Continuous Partial Derivatives

**Proposition 9.10** *Let $X$ be an open set in $\mathbb{R}^m$, let $\varphi: X \to \mathbb{R}^n$ be a function mapping $X$ into $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of $X$, where $\mathbf{p} = (p_1, p_2, \ldots, p_m)$. Suppose that the partial derivatives of the components of $\varphi$ with respect to the Cartesian coordinates exist and are continuous throughout $X$. Suppose also that the partial derivatives of the components of $\varphi$ are all equal to zero at the point $\mathbf{p}$. Then, given any positive real number $\varepsilon$, there exists a positive real number $\delta$ such that, for all points $\mathbf{u}$ and $\mathbf{v}$ of $H(\mathbf{p}, \delta)$, $\mathbf{u} \in X$, $\mathbf{v} \in X$ and*

$$|f(\mathbf{u}) - f(\mathbf{v})| \leq \varepsilon|\mathbf{u} - \mathbf{v}|,$$

*where*

$$H(\mathbf{p}, \delta) = \{(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m : |x_j - p_j| < \delta \text{ for } j = 1, 2, \ldots, m\}.$$

**Proof** Let us denote the $j$th partial derivative $\dfrac{\partial f_i}{\partial x_j}$ of the $i$th component $f_i$ of $\varphi$ by $\partial_j f_i$ for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. Then $\partial_j f_i$ is a continuous function on $f$.

Let some positive real number $\varepsilon$ be given. Then there exists a positive real number $\delta$ that is small enough to ensure that $\mathbf{x} \in X$ and

$$|(\partial_j f_i)(x_1, x_2, \ldots, x_m)| \leq \varepsilon/\sqrt{mn}$$

for all points $\mathbf{x}$ of $H(\mathbf{p}, \delta)$.

Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m$ denote the standard basis of $\mathbb{R}^m$, defined so that

$$(z_1, z_2, \ldots, z_m) = \sum_{j=1}^{m} z_j \mathbf{e}_j$$

for all $(z_1, z_2, \ldots, z_m) \in \mathbb{R}^m$. Let $\mathbf{u}$ and $\mathbf{v}$ be points of $H(\mathbf{p}, \delta)$, and let points $\mathbf{q}_j$ be defined for $j = 0, 1, 2, \ldots, m$ so that $\mathbf{q}_0 = \mathbf{v}$ and

$$\mathbf{q}_j = \mathbf{q}_{j-1} + (u_j - v_j)\mathbf{e}_j$$

for $j = 1, 2, \ldots, n$. Then $\mathbf{q}_m = \mathbf{u}$ and $\mathbf{q}_j \in H(\mathbf{p}, \delta)$ for $j = 1, 2, \ldots, m$. Now, for each integer $j$ between 1 and $m$, the points $\mathbf{q}_j$ and $\mathbf{q}_{j-1}$ differ only in the $j$th coordinate. Applying the Mean Value Theorem of single-variable calculus (Theorem 7.2), we find that, given any pair of integers $i$ and $j$, where $1 \leq i \leq n$ and $1 \leq j \leq m$, there exists some real number $\theta$ satisfying $0 < \theta < 1$ such that

$$f_i(\mathbf{q}_j) - f_i(\mathbf{q}_{j-1}) = (u_j - v_j)(\partial_j f_i)\Big((1-\theta)\mathbf{q}_{j-1} + \theta\mathbf{q}_j\Big).$$

It follows that

$$|f_i(\mathbf{q}_j) - f_i(\mathbf{q}_{j-1})| \leq \frac{\varepsilon}{\sqrt{mn}}|u_j - v_j|$$

for $j = 1, 2, \ldots, m$. It then follows that

$$|f_i(\mathbf{u}) - f_i(\mathbf{v})| \leq \sum_{j=1}^{m} |f_i(\mathbf{q}_j) - f_i(\mathbf{q}_{j-1})| \leq \frac{\varepsilon}{\sqrt{mn}} \sum_{j=1}^{m} |u_j - v_j|.$$

On applying Schwarz's Inequality (Lemma 2.1), we find that

$$\left(\sum_{j=1}^{m} |u_j - v_j|\right)^2 \leq m \sum_{j=1}^{m} (u_j - v_j)^2 = m\,|\mathbf{u} - \mathbf{v}|^2.$$

It follows that

$$\sum_{j=1}^{m} |u_j - v_j| \leq \sqrt{m}\,|\mathbf{u} - \mathbf{v}|$$

and therefore

$$|f_i(\mathbf{u}) - f_i(\mathbf{v})| \leq \frac{\varepsilon}{\sqrt{n}}|\mathbf{u} - \mathbf{v}|.$$

It follows that

$$|\varphi(\mathbf{u}) - \varphi(\mathbf{v})|^2 = \sum_{i=1}^{n} (f_i(\mathbf{u}) - f_i(\mathbf{v}))^2 \leq \varepsilon^2 |\mathbf{u} - \mathbf{v}|^2$$

for all points $\mathbf{u}$ and $\mathbf{v}$ of $H(\mathbf{p}, \delta)$. The result follows. ∎

**Remark** The essential strategy underlying the proof of Proposition 9.10 can be presented, in the two-dimensional case, as follows. Consider a city laid out on a gridiron pattern, where all streets run either from north to south, or from east to west. To get from one street intersection to another, it is always possible to find a route that does not involve both northward and southward legs, and does not involve both eastward and westward legs. (Thus to get from one street intersection to another that lies to the northeast, one can choose a route that involves only travelling northwards or travelling eastwards along city streets.) Suppose that all streets have a maximum gradient equal to $m$. Then the height difference between any two intersections is bounded above by $\sqrt{2}md$, where $d$ is the direct distance between those street intersections.

**Corollary 9.11** *Let $\varphi: X \to \mathbb{R}^n$ be a continuously differentiable function defined over an open set $X$ in $\mathbb{R}^m$, and let $\mathbf{p}$ be a point of $X$. Let $M$ be a positive real number satisfying $M > \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$, where $\|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$ denotes the operator norm of the derivative of $\varphi$ at $\mathbf{p}$. Then there exists a positive real number $\delta$ such that*

$$|\varphi(\mathbf{u}) - \varphi(\mathbf{v})| \leq M|\mathbf{u} - \mathbf{v}|$$

*for all points $\mathbf{u}$ and $\mathbf{v}$ of $X$ that satisfy $|\mathbf{u} - \mathbf{p}| < \delta$ and $|\mathbf{v} - \mathbf{p}| < \delta$.*

**Proof** Let $M_0 = \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$ and let $\varepsilon = M - M_0$. Let $\psi: X \to \mathbb{R}^n$ be defined such that

$$\psi(\mathbf{u}) = \varphi(\mathbf{u}) - (D\varphi)_{\mathbf{p}}\mathbf{u}$$

for all $\mathbf{u} \in X$. Then $(D\psi)_{\mathbf{p}} = (D\varphi)_{\mathbf{p}} - (D\varphi)_{\mathbf{p}} = 0$. It follows from Proposition 9.10 that there exists a positive real number $\delta$ such that

$$|\psi(\mathbf{u}) - \psi(\mathbf{v})| \leq \varepsilon|\mathbf{u} - \mathbf{v}|$$

for all points $\mathbf{u}$ and $\mathbf{v}$ of $X$ that satisfy $|\mathbf{u}-\mathbf{p}| < \delta$ and $|\mathbf{v}-\mathbf{p}| < \delta$. Then

$$\begin{aligned}
|\varphi(\mathbf{u}) - \varphi(\mathbf{v})| &= |\psi(\mathbf{u}) - \psi(\mathbf{v}) + (D\varphi)_{\mathbf{p}}(\mathbf{u}-\mathbf{v})| \\
&\leq |\psi(\mathbf{u}) - \psi(\mathbf{v})| + |(D\varphi)_{\mathbf{p}}(\mathbf{u}-\mathbf{v})| \\
&\leq \varepsilon|\mathbf{u}-\mathbf{v}| + M_0|\mathbf{u}-\mathbf{v}| = M|\mathbf{u}-\mathbf{v}|
\end{aligned}$$

for all points $\mathbf{u}$ and $\mathbf{v}$ of $X$ that satisfy $|\mathbf{u}-\mathbf{p}| < \delta$ and $|\mathbf{v}-\mathbf{p}| < \delta$, as required. ∎

Corollary 9.11 ensures that continuously differentiable functions of several real variables are *locally Lipschitz continuous*. This means that they satisfy a Lipschitz condition in some sufficiently small neighbourhood of any given point. This in turn ensures that standard theorems concerning the existence and uniqueness of ordinary differential equations can be applied to systems of ordinary differential equations specified in terms of continuously differentiable functions.

**Theorem 9.12** *Let $X$ be an open subset of $\mathbb{R}^m$ and let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping $X$ into $\mathbb{R}^n$. Suppose that the Jacobian matrix*

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

*exists at every point of $X$, where $f_i$ denotes the $i$th component of $\varphi$ for $i = 1, 2, \ldots, n$. Suppose also that the coefficients of this Jacobian matrix are continuous functions on $X$. Then $\varphi$ is differentiable at every point of $X$, and the derivative of $\varphi$ at each point is represented by the Jacobian matrix.*

**Proof** Let $\mathbf{p} \in X$, and, for each integer $i$ between 1 and $n$, let $g_i\colon X \to \mathbb{R}$ be defined such that

$$g_i(\mathbf{x}) = f_i(\mathbf{x}) - \sum_{j=1}^{m} J_{i,j}(x_j - p_j)$$

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

$$J_{i,j} = (\partial_j f_i)(\mathbf{p}) = \left.\frac{\partial f_i}{\partial x_j}\right|_{\mathbf{p}}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. The partial derivatives $\partial_j g_i$ of the function $g_i$ are then determined by those of $f_i$ so that

$$(\partial_j g_i)(\mathbf{x}) = (\partial_j f_i)(\mathbf{x}) - J_{i,j}$$

for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$. It follows that $(\partial_j g_i)(\mathbf{p}) = 0$ for $j = 1, 2, \ldots, m$.

Let $\psi: X \to \mathbb{R}^n$ be defined so that $\psi(\mathbf{x}) = (g_1(\mathbf{x}), g_2(\mathbf{x}), \ldots, g_n(\mathbf{x}))$ for all $\mathbf{x} \in X$. Then the partial derivatives of the function $\psi: X \to \mathbb{R}^n$ are all equal to zero at the point $\mathbf{p}$.

Let some positive real number $\varepsilon$ be given. It follows from Proposition 9.10 that there exists some positive real number $\delta$ such that

$$|\psi(\mathbf{x}) - \psi(\mathbf{p})| \leq \varepsilon|\mathbf{x} - \mathbf{p}|$$

for all $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$. But then

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - J(\mathbf{x} - \mathbf{p})| \leq \varepsilon|\mathbf{x} - \mathbf{p}|$$

for all $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$, where $J$ denotes the Jacobian matrix of $\varphi$ at the point $\mathbf{p}$ (i.e., the matrix whose coefficient in the $i$th row and $j$th column of the matrix is equal to the value of the partial derivative

$$\frac{\partial f_i}{\partial x_j}$$

at the point $\mathbf{p}$). It follows from this that

$$\lim_{\mathbf{x} \to \mathbf{p}} \frac{1}{|\mathbf{x} - \mathbf{p}|} |\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - J(\mathbf{x} - \mathbf{p})| = 0,$$

and thus the function $\varphi$ is differentiable at $\mathbf{p}$. Moreover the matrix representing the derivative $(D\varphi)_\mathbf{p}$ of $\varphi$ at the point $\mathbf{p}$ is the Jacobian matrix at that point, as required. ∎

**Corollary 9.13** *Let $X$ be an open set in $\mathbb{R}^m$. A function $\varphi: X \to \mathbb{R}^n$ is continuously differentiable if and only if the first order partial derivatives of the components of $\varphi$ exist and are continuous throughout $X$.*

**Proof** The result follows directly on combining the results of Lemma 9.9 and Theorem 9.12. ∎

## 9.8 Summary of Differentiability Results

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

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

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

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

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

(iii) There exist functions $\varphi: X \to \mathbb{R}^n$ whose first order partial derivatives are well-defined at a particular point of $X$ 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 $\varphi: X \to \mathbb{R}^n$ 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 $\varphi: X \to \mathbb{R}^n$ is differentiable if and only if its components are differentiable functions on $X$ (where $X$ is an open set in $\mathbb{R}^m$).

(vii) Given two differentiable functions from $X$ to $\mathbb{R}$, where $X$ is an open set in $\mathbb{R}^m$, 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.

# 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 $f\colon X \to \mathbb{R}$ 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

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

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

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

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

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

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

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

Similarly

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

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

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

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 $f\colon X \to \mathbb{R}$ be a real-valued function on $X$. Suppose that the partial derivatives*

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

*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 7.2) 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 $q\colon (b - L, b + L) \to \mathbb{R}$ 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 7.2) 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 $\varepsilon$ be given. The function $f_{xy}$ is continuous. Therefore there exists some real number $\delta$ 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

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

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 $\dfrac{f_x(a,b+k)-f_x(a,b)}{k}$ 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

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

as required. ∎

**Corollary 10.2** *Let $X$ be an open set in $\mathbb{R}^n$ and let $f\colon X \to \mathbb{R}$ 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

Let $f\colon X \to \mathbb{R}$ 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 $\mathbf{p} = (p_1, p_2, \ldots, p_n)$. Now for each integer $i$ between 1 and $n$ the map

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

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 $\theta$ satisfying $0 < \theta < 1$ for which*

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

*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 9.8) 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

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

and

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

Now

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

for some real number $\theta$ satisfying $0 < \theta < 1$. (see Proposition 7.3). It follows that

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

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 $j = 1, 2, \ldots, n$, 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 $\theta$ 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 $\displaystyle\sum_{i=1}^{n} \sum_{j=1}^{n} c_{i,j} h_i h_j \geq 0$ 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 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

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

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 4.21 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 $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. 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 $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. 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 $f\colon X \to \mathbb{R}$ 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

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

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 $f: X \to \mathbb{R}$ 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*

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

*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 $\theta$ satisfying $0 < \theta < 1$ (where $\theta$ 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

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

The result follows. ∎

Let $f: X \to \mathbb{R}$ 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 $f$ 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 10.6** *Let $f\colon X \to \mathbb{R}$ 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*

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

*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 $\theta$ satisfying $0 < \theta < 1$ (where $\theta$ 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 $\lambda_1$ and $\lambda_2$ are the eigenvalues of a symmetric $2 \times 2$ matrix $C$, then

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

Thus a symmetric $2 \times 2$ 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 10.6 that the function $f$ has a local minimum at $(0, 0)$.