Module MAU23203: Analysis in Several Real Variables
Michaelmas Term 2019
Part III (Sections 8, 9, 10 and 11)

D. R. Wilkins

Contents

8 Differentiation of Functions of One Real Variable

8.1 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 δ 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 δ such that $f(x)$ is defined for all real numbers x satisfying $s - \delta < x < s + \delta$.

8.2 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:—

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

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

8.3 Rolle's Theorem

Theorem 8.1 (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 $a < s < b$. ∎

8.4 The Mean Value Theorem

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

Theorem 8.2 (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 8.1) 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. ∎

A number of basic principles of single variable calculus follow as immediate consequences of the Mean Value Theorem (Theorem 8.2). A number of such consequences are presented in the following corollaries.

Corollary 8.3 *Let $f: [a,b] \to \mathbb{R}$ 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) and that $f'(x) > 0$ for all real numbers x satisfying $a < x < b$. Then $f(b) > f(a)$.*

Corollary 8.4 *Let $f: [a,b] \to \mathbb{R}$ 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) and that $f'(x) = 0$ for all real numbers x satisfying $a < x < b$. Then $f(x) = f(a)$ for all $x \in [a,b]$.*

Corollary 8.5 *Let $f: [a,b] \to \mathbb{R}$ be a real-valued function defined on some interval $[a,b]$, and let M be a real number. Suppose that f is continuous on $[a,b]$ and is differentiable on (a,b) and that $f'(x) \leq M$ for all real numbers x satisfying $a < x < b$. Then $f(x) \leq f(a) + M(x-a)$ for all $x \in [a,b]$.*

Corollary 8.6 *Let $f: [a,b] \to \mathbb{R}$ be a real-valued function defined on some interval $[a,b]$, and let M be a real number. Suppose that f is continuous on $[a,b]$ and is differentiable on (a,b) and that $|f'(x)| \leq M$ for all real numbers x satisfying $a < x < b$. Then $|f(b) - f(a)| \leq M(b-a)$.*

8.5 Concavity and the Second Derivative

Proposition 8.7 *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 θ 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 $q: I \to \mathbb{R}$ 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 φ 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 θ 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 8.8 *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 8.8 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 8.9 *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.*

8.6 Taylor's Theorem

The result obtained in Proposition 8.7 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 8.10 *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 8.11 (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 θ 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 8.10). 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 8.1) 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 8.12 *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 ε, there exists some strictly positive real number δ 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 kth derivative $f^{(k)}$ is continuous. Let some strictly positive real number ε be given. Then there exists some strictly positive real number δ 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 θ 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 8.11) that, given any real number h satisfying $|h| < \delta$ there exists some real number θ 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 8.13 *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 8.2) 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 8.14 (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 8.13 on integrating the identity

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

Corollary 8.15 (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 8.2) 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. $\blacksquare$

Proposition 8.16 (Taylor's Theorem with Integral Remainder) *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*

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

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 8.14) that

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

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. ∎

9 Differentiation of Functions of Several Real Variables

9.1 Functions with First Order Partial Derivatives

If a real-valued function of a single real variable is differentiable, then it is guaranteed to be continuous. However, for a function of two or more real variables, the mere existence of first order partial derivatives throughout the domain of the function is not sufficient to ensure continuity.

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

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

If $(x,y) \neq (0,0)$ then the partial derivatives of f are well-defined at (x,y), and

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

The partial derivatives of the function f at $(0,0)$ are also well-defined, and are equal to zero, because the function f has the value zero along the lines $y = 0$ and $x = 0$. Thus the first order partial derivatives of the function f are well-defined throughout the domain $\mathbb{R}^2$ of the function.

Nevertheless $f(x,y) = 1$ at all points of the line $x = y$ with the exception of the origin $(0,0)$, where the function takes the value zero. It follows from this that the function f is discontinuous at $(0,0)$.

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

$$g(x,y) = \begin{cases} \dfrac{2xy}{x^2+y^2}\, e^{\frac{1}{x^2+y^2}} & \text{if } (x,y) \neq (0,0); \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

This function g also have well-defined first order partial derivatives throughout $\mathbb{R}^2$. But $|g(x,y)|$ increases faster than any negative power of the distance from the origin as the point (x,y) approaches the origin along any straight line other than the lines $y = 0$ and $x = 0$.

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

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

The function h takes the value zero along the lines $y = 0$ and $x = 0$. It therefore has well-defined first order partial derivatives at the origin that have the value zero. It also has well-defined first order partial derivatives at all other points of $\mathbb{R}^2$.

Now if (u, v) is a point of $\mathbb{R}^2$, and if $v \neq 0$ and $t \neq 0$, then

$$h(tu, tv) = \frac{2tu^2 v}{t^2 u^4 + v^2}.$$

It follows that $\lim_{t \to 0} h(tu, tv) = 0$ whenever $v \neq 0$. This limit is also zero when $v = 0$ because the function takes the value zero along the line $y = 0$. Nevertheless $h(t, t^2) = 1$ for all non-zero real numbers t. The point (t, t^2) approaches the origin $(0, 0)$ as t tends to zero, and $h(0, 0) = 0$. It follows that the function h is not continuous at the origin.

9.2 Growth of Functions with Bounded Partial Derivatives

An open set X in $\mathbb{R}^m$ is a product of open intervals $J_1, J_2, \ldots, J_m$ if

$$\begin{aligned} X &= J_1 \times J_2 \times \cdots \times J_m \\ &= \{(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m : x_i \in J_i \text{ for } i = 1, 2, \ldots, m\}. \end{aligned}$$

Suppose that $\mathbf{u}$ and $\mathbf{v}$ are points of X, where X is an open set in $\mathbb{R}^m$ that is a product of open intervals. Then there exist real numbers a_i and b_i for $i = 1, 2, \ldots, m$ such that $a_i < u_i < b_i$ and $a_i < v_i < b_i$ for $i = 1, 2, \ldots, m$. Then $H \subset X$, where H is the closed subset of $\mathbb{R}^m$ consisting of those points $(x_1, x_2, \ldots, x_m)$ whose ith coordinate x_i satisfies $\min(u_i, v_i) \leq x_i \leq \max(u_i, v_i)$ for $i = 1, 2, \ldots, m$.

Lemma 9.1 *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be points of* $\mathbb{R}^m$. *Then*

$$\sum_{i=1}^{m} |u_i - v_i| \leq \sqrt{m}\, |\mathbf{u} - \mathbf{v}|.$$

Proof Consider the scalar product $(\mathbf{u} - \mathbf{v}) . \mathbf{s}$ of the m-dimensional vector $\mathbf{u} - \mathbf{v}$ with the vector $\mathbf{s}$ whose ith component s_i is determined for $i = 1, 2, \ldots, m$ so that $s_i = 1$ if $u_i \geq v_i$ and $s_i = -1$ if $u_i < v_i$. Now $(\mathbf{u} - \mathbf{v}) . \mathbf{s} = \sum_{i=1}^{m} |u_i - v_i|$ and $|\mathbf{s}| = \sqrt{m}$. Schwarz's Inequality ensures that $(\mathbf{u} - \mathbf{v}) . \mathbf{s} \leq |\mathbf{u} - \mathbf{v}|\, |\mathbf{s}|$. The required inequality follows immediately. ∎

Proposition 9.2 *Let X be an open set in $\mathbb{R}^m$ that is a product of open intervals, let $f\colon X \to \mathbb{R}$ be a real-valued function on X, and let M be a positive constant. Suppose that*

$$\left|\frac{\partial f}{\partial x_i}\right| \leq M$$

throughout the open set X for $i = 1, 2, \ldots, m$. Then

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

for all points $\mathbf{u}$ and $\mathbf{v}$ of X.

Proof Let $\mathbf{u} = (u_1, u_2, \ldots, u_m)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_m)$. The fact that X is a product of open intervals guarantees that there exist real numbers a_i and b_i for $i = 1, 2, \ldots, m$ such that $a_i < u_i < b_i$ and $a_i < v_i < b_i$ for $i = 1, 2, \ldots, m$. For each integer k between 0 and m, let

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

where

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

Then $a_i < w_{k,i} < b_i$ for $k = 0, 1, 2, \ldots, m$ and $i = 1, 2, \ldots, m$. Moreover, for each integer k between 1 and m, the points $\mathbf{w}_{k-1}$ and and $\mathbf{w}_k$ differ only in the kth coordinate, and the line segment joining these points is wholly contained in the open set X. It follows that

$$\frac{d}{dt}\left(f((1-t)\mathbf{w}_{k-1} + t\mathbf{w}_k)\right) = (v_k - u_k)\left.\frac{\partial f}{\partial x_k}\right|_{(1-t)\mathbf{w}_{k-1}+t\mathbf{w}_k},$$

It follows that

$$\left|\frac{d}{dt}\left(f((1-t)\mathbf{w}_{k-1} + t\mathbf{w}_k)\right)\right| \leq M\,|u_k - v_k|,$$

and therefore

$$|f(\mathbf{w}_{k-1}) - f(\mathbf{w}_k)| \leq M\,|u_k - v_k|$$

for $i = 1, 2, \ldots, m$ (see Corollary 8.6). Applying the inequality stated in Lemma 9.1, we conclude that

$$\begin{aligned} |f(\mathbf{u}) - f(\mathbf{v})| &\leq \sum_{k=1}^{m} |f(\mathbf{w}_{k-1}) - f(\mathbf{w}_k)| \leq M \sum_{k=1}^{n} |u_k - v_k| \\ &\leq \sqrt{m}\, M\, |\mathbf{u} - \mathbf{v}|, \end{aligned}$$

as required. ∎

Example Let $f: \mathbb{R}^m \to \mathbb{R}$ be the function defined so that $f(x_1, x_2, \ldots, x_m) = \sum_{i=1}^m x_i$ for all $(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m$. Now $\dfrac{\partial f}{\partial x_i} = 1$ throughout $\mathbb{R}^m$ for $i = 1, 2, \ldots, m$. Let $\mathbf{u} = (0, 0, \ldots, 0)$ and $\mathbf{v} = (1, 1, \ldots, 1)$. Then $|\mathbf{u} - \mathbf{v}| = \sqrt{m}$ and $|f(\mathbf{u}) - f(\mathbf{v})| = m$, and thus $|f(\mathbf{u}) - f(\mathbf{v})| = \sqrt{m}\, |\mathbf{u} - \mathbf{v}|$. This shows that the inequality proved in Proposition 9.2 is sharp, i.e., there exist instances where, with an appropriate choice of the function f and the points $\mathbf{u}$ and $\mathbf{v}$, the stated upper bound on $|f(\mathbf{u}) - f(\mathbf{v})|$ is attained.

Corollary 9.3 *Let X be an open set in $\mathbb{R}^m$ that is a product of open intervals, let $\varphi: X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, and let M be a positive constant. Suppose that*

$$\left| \frac{\partial f_j}{\partial x_i} \right| \leq M$$

throughout the open set X for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, where $f_j: X \to \mathbb{R}$ is the jth component of the map φ. Then

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

for all points $\mathbf{u}$ and $\mathbf{v}$ of X.

Proof Let $\mathbf{u}$ and $\mathbf{v}$ be points of X. On applying the inequality stated in Proposition 9.2, we find that

$$|\varphi(\mathbf{u}) - \varphi(\mathbf{v})|^2 = \sum_{j=1}^n (f_j(\mathbf{u}) - f_j(\mathbf{v}))^2 \leq mnM^2\, |\mathbf{u} - \mathbf{v}|^2.$$

The result follows. ∎

Corollary 9.4 *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 M be a positive constant. Suppose that*

$$\left| \frac{\partial f_j}{\partial x_i} \right| \leq M$$

throughout the open set X for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, where $f_j: X \to \mathbb{R}$ is the jth component of the map φ. Then the function φ is continuous on X.

Proof Let $\mathbf{p}$ be a point of X. The set X is open in $\mathbb{R}^m$, and therefore there exists an open set V that is a product of open intervals such that $\mathbf{p} \in V$ and $V \subset X$. It then follows from Corollary 9.3 that

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

for all points $\mathbf{u}$ and $\mathbf{v}$ of V. This inequality ensures that the function φ is continuous around the point $\mathbf{p}$. The result follows. ∎

9.3 Functions with Continuous Partial Derivatives

We now investigate the behaviour of functions of several real variables whose first order partial derivatives are continuous.

Definition Let X be an open set in $\mathbb{R}^m$, and let $f: X \to \mathbb{R}^n$ be real-valued function on X, and let $\mathbf{p}$ be a point of X. Suppose that the first order partial derivatives of f are defined at the point $\mathbf{p}$. The *gradient* $(\nabla f)_{\mathbf{p}}$ of f at the point $\mathbf{p}$ is the m-dimensional vector whose components are the partial derivatives of the function f at the point $\mathbf{p}$. Thus

$$(\nabla f)_{\mathbf{p}} = \left(\left.\frac{\partial f}{\partial x_1}\right|_{\mathbf{x}=\mathbf{p}}, \left.\frac{\partial f}{\partial x_2}\right|_{\mathbf{x}=\mathbf{p}}, \ldots \left.\frac{\partial f}{\partial x_m}\right|_{\mathbf{x}=\mathbf{p}} \right).$$

Examples previously considered demonstrate that the mere existence of partial derivatives of a real-valued function around a given point is not sufficient to enable the gradient of that function to provide a reasonable approximation to the function around that point. On the other hand, as we shall see, if the partial derivatives are not only defined around that point but are also continuous there, then the gradient of the function does determine a "first order" approximation to the function around that point.

Proposition 9.5 *Let X be an open set in $\mathbb{R}^m$, let $f: X \to \mathbb{R}$ be a real-valued function on X, and let $\mathbf{p}$ be a point of X. Suppose that the first order partial derivatives of the function f are defined throughout the set X and are continuous at the point $\mathbf{p}$. Then, given any positive real number ε, there exists some positive real number δ such that*

$$|f(\mathbf{u}) - f(\mathbf{v}) - (\nabla f)_{\mathbf{p}} \,.\, (\mathbf{u} - \mathbf{v})| \leq \varepsilon |\mathbf{u} - \mathbf{v}|$$

for all points $\mathbf{u}$ and $\mathbf{v}$ of X satisfying $|\mathbf{u} - \mathbf{p}| < \delta$ and $|\mathbf{v} - \mathbf{p}| < \delta$, where $(\nabla f)_{\mathbf{p}}$ denotes the gradient of the function f at the point $\mathbf{p}$.

Proof Let $\mathbf{p} = (p_1, p_2, \ldots, p_m)$, and let $g: X \to \mathbb{R}$ be the real-valued function on X defined so that

$$g(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{p}) - (\nabla f)_{\mathbf{p}} \,.\, (\mathbf{x} - \mathbf{p})$$

for all $\mathbf{x} \in X$. Then the function g has first order partial derivatives, defined throughout the open set X, which are continuous at the point $\mathbf{p}$. Moreover $g(\mathbf{p}) = 0$ and

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

for $i = 1, 2, \ldots, n$. Moreover

$$g(\mathbf{u}) - g(\mathbf{v}) = f(\mathbf{u}) - f(\mathbf{v}) - (\nabla f)_{\mathbf{p}} \,.\, (\mathbf{u} - \mathbf{v})$$

for all points $\mathbf{u}$ and $\mathbf{v}$ of X.

Let some positive real number ε be given. for $i = 1, 2, \ldots, n$. Now the domain X of the functions f and g is an open subset of $\mathbb{R}^m$. This, together with the continuity of the first order partial derivatives of the function g at the point $\mathbf{p}$, ensures that some positive real number δ can then be chosen small enough to ensure both that

$$\{(x_1, x_2, \ldots, x_m) \in \mathbb{R}^m : |p_i - x_i| \leq \delta \text{ for } i = 1, 2, \ldots, m\} \subset X$$

and also that

$$\left|\frac{\partial g}{\partial x_i}\right| \leq \frac{\varepsilon}{\sqrt{m}}$$

for $i = 1, 2, \ldots, m$ at all points $(x_1, x_2, \ldots, x_m)$ of $\mathbb{R}^m$ that satisfy $|x_i - p_i| < \delta$ for $i = 1, 2, \ldots, m$. Now $g(\mathbf{p}) = 0$. It therefore follows from Proposition 9.2 that

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

at all points $\mathbf{x}$ of $\mathbb{R}^m$ whose components $x_1, x_2, \ldots, x_m$ satisfy $|x_i - p_i| < \delta$ for $i = 1, 2, \ldots, m$. The result follows. ∎

Corollary 9.6 *Let X be an open set in $\mathbb{R}^m$, let $f\colon X \to \mathbb{R}^n$ be a real-valued function on X, and let $\mathbf{p}$ be a point of X. Suppose that the first order partial derivatives of the function f are defined throughout the set X and are continuous at the point $\mathbf{p}$. Then*

$$\lim_{\mathbf{x} \to \mathbf{p}} \frac{1}{|\mathbf{x} - \mathbf{p}|}\, |f(\mathbf{x}) - f(\mathbf{p}) - (\nabla f)_{\mathbf{p}} \,.\, (\mathbf{x} - \mathbf{p}))| = 0,$$

where $(\nabla f)_{\mathbf{p}}$ denotes the gradient of the function f at the point $\mathbf{p}$.

Proof Proposition 9.5 ensures that, given any positive real number ε, there exists a positive real number δ such that

$$\frac{1}{|\mathbf{x} - \mathbf{p}|}\, |f(\mathbf{x}) - f(\mathbf{p}) - (\nabla f)_{\mathbf{p}} \,.\, (\mathbf{x} - \mathbf{p})| \leq \varepsilon$$

for all points $\mathbf{x}$ of X satisfying $0 < |\mathbf{x} - \mathbf{p}| < \delta$. The result therefore follows directly from the formal definition of limits of functions of several real variables. ∎

Corollary 9.7 *Let X be an open set in $\mathbb{R}^m$, let $\varphi\colon X \to \mathbb{R}^n$ be a function on X taking values in $\mathbb{R}^n$, let $f_1, f_2, \ldots, f_n$ be the components of the map φ, and let $\mathbf{p}$ be a point of X. Suppose that the first order partial derivatives of the components of the map φ are defined throughout the set X and are continuous at the point $\mathbf{p}$. Then, given any positive real number ε, there exists some positive real number δ such that*

$$|\varphi(\mathbf{u}) - \varphi(\mathbf{v}) - (D\varphi)_{\mathbf{p}}\,(\mathbf{u} - \mathbf{v})| \leq \varepsilon|\mathbf{u} - \mathbf{v}|$$

for all points $\mathbf{u}$ and $\mathbf{v}$ of X satisfying $|\mathbf{u} - \mathbf{p}| < \delta$ and $|\mathbf{v} - \mathbf{p}| < \delta$, where

$$(D\varphi)_{\mathbf{p}}\,\mathbf{w} = ((\nabla f_1)_{\mathbf{p}} \,.\, \mathbf{w}, (\nabla f_2)_{\mathbf{p}} \,.\, \mathbf{w}, \ldots, (\nabla f_n)_{\mathbf{p}} \,.\, \mathbf{w})$$

for all $\mathbf{w} \in \mathbb{R}^m$.

Proof It follows from Proposition 9.5 that, given any positive real number ε, there exists some positive real number δ such that

$$|f_j(\mathbf{u}) - f_j(\mathbf{v}) - (\nabla f_j)_{\mathbf{p}} \,.\, (\mathbf{u} - \mathbf{v})| \leq \frac{\varepsilon}{\sqrt{n}}|\mathbf{u} - \mathbf{v}|$$

for $j = 1, 2, \ldots, n$, and for all points $\mathbf{u}$ and $\mathbf{v}$ of X satisfying $|\mathbf{u} - \mathbf{p}| < \delta$ and $|\mathbf{v} - \mathbf{p}| < \delta$. Then

$$\begin{aligned} &|\varphi(\mathbf{u}) - \varphi(\mathbf{v}) - (D\varphi)_{\mathbf{p}}\,(\mathbf{u} - \mathbf{v})|^2 \\ &\qquad = \sum_{j=1}^{n} (f_j(\mathbf{u}) - f_j(\mathbf{v}) - (\nabla f_j)_{\mathbf{p}} \,.\, (\mathbf{u} - \mathbf{v}))^2 \\ &\qquad \leq \varepsilon^2 |\mathbf{u} - \mathbf{v}|^2 \end{aligned}$$

for all points $\mathbf{u}$ and $\mathbf{v}$ of X satisfying $|\mathbf{u} - \mathbf{p}| < \delta$ and $|\mathbf{v} - \mathbf{p}| < \delta$. The result follows. ∎

Corollary 9.8 *Let X be an open set in $\mathbb{R}^m$, let $\varphi\colon X \to \mathbb{R}^n$ be a function on X taking values in $\mathbb{R}^n$, let $f_1, f_2, \ldots, f_n$ be the components of the map φ, and let $\mathbf{p}$ be a point of X. Suppose that the first order partial derivatives of the components of the map φ are defined throughout the set X and are continuous at the point $\mathbf{p}$. Then*

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

where

$$(D\varphi)_{\mathbf{p}}\,\mathbf{w} = ((\nabla f_1)_{\mathbf{p}} \,.\, \mathbf{w}, (\nabla f_2)_{\mathbf{p}} \,.\, \mathbf{w}, \ldots, (\nabla f_n)_{\mathbf{p}} \,.\, \mathbf{w})$$

for all $\mathbf{w} \in \mathbb{R}^m$.

Proof Proposition 9.7 ensures that, given any positive real number ε, there exists a positive real number δ such that

$$\frac{1}{|\mathbf{x}-\mathbf{p}|}\,|\varphi(\mathbf{x})-\varphi(\mathbf{p})-(Df)_{\mathbf{p}}\,(\mathbf{x}-\mathbf{p})|\leq\varepsilon$$

for all points $\mathbf{x}$ of X satisfying $0<|\mathbf{x}-\mathbf{p}|<\delta$. The result therefore follows directly from the formal definition of limits of functions of several real variables. ∎

9.4 Derivatives of Functions of Several Variables

Definition Let X be an open subset of $\mathbb{R}^m$ let $\varphi\colon X\to\mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, let $T\colon\mathbb{R}^m\to\mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. The function φ is said to be *differentiable* at $\mathbf{p}$, with *derivative* $T\colon\mathbb{R}^m\to\mathbb{R}^n$ if and only if

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

Henceforth we shall usually denote the derivative of a differentiable map $\varphi\colon X\to\mathbb{R}^n$ at a point $\mathbf{p}$ of its domain X by $(D\varphi)_{\mathbf{p}}$.

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

Lemma 9.9 *Let $T\colon\mathbb{R}^m\to\mathbb{R}^n$ 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 definition of differentiability, given that $T\mathbf{x}-T\mathbf{p}-T(\mathbf{x}-\mathbf{p})=\mathbf{0}$ for all $\mathbf{x}\in\mathbb{R}^m$. ∎

Lemma 9.10 *Let X be an open subset of $\mathbb{R}^m$ let $\varphi\colon X\to\mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, let $T\colon\mathbb{R}^m\to\mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. Then the function φ is differentiable at $\mathbf{p}$, with derivative T, if and only if, given any positive real number ε, there exists some positive real number δ such that*

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

at all points $\mathbf{x}$ of X that satisfy $|\mathbf{x}-\mathbf{p}|<\delta$.

Proof First suppose that the function $\varphi: X \to \mathbb{R}^n$ has the property that, given any positive real number ε_0, there exists some positive real number δ such that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p})| \leq \varepsilon_0 |\mathbf{x} - \mathbf{p}|$$

at all points $\mathbf{x}$ of X that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$. Let some positive number ε be given, and let ε_0 be chosen so that $0 < \varepsilon_0 < \varepsilon$. Then there exists some positive real number δ such that the above inequality holds at all points $\mathbf{x}$ of X that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$. But then

$$\frac{1}{|\mathbf{x} - \mathbf{p}|} |\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p})| < \varepsilon$$

at all points $\mathbf{x}$ of X that satisfy $0 < |\mathbf{x} - \mathbf{p}| < \delta$, and therefore

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

Thus the function φ is differentiable at the point $\mathbf{p}$.

Conversely suppose that the function φ is differentiable at the point $\mathbf{p}$. Let some positive real number ε be given. Then there exists some positive real number δ such that

$$\frac{1}{|\mathbf{x} - \mathbf{p}|} |\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - T(\mathbf{x} - \mathbf{p})| < \varepsilon$$

at all points $\mathbf{x}$ of X that satisfy $0 < |\mathbf{x} - \mathbf{p}| < \delta$. Considering separately the cases when $\mathbf{x} = \mathbf{p}$ and when $0 < |\mathbf{x} - \mathbf{p}| < \delta$, it then follows that

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

at all points $\mathbf{x}$ of X that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$. The result follows. ∎

Lemma 9.11 *Let X be an open subset of $\mathbb{R}^m$ 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. Suppose that the function φ is differentiable at the point $\mathbf{p}$. Then φ is continuous at $\mathbf{p}$.*

Proof Suppose that the function φ is differentiable at $\mathbf{p}$ with derivative $(D\varphi)_{\mathbf{p}}$. It then follows from the definition of differentiability that

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

It then follows from basic properties of limits that

$$\begin{aligned}
&\lim_{\mathbf{x}\to\mathbf{p}} |\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})| \\
&\quad = \left(\lim_{\mathbf{x}\to\mathbf{p}} |\mathbf{x}-\mathbf{p}|\right)\left(\lim_{\mathbf{x}\to\mathbf{p}} \frac{1}{|\mathbf{x}-\mathbf{p}|} |\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})|\right) \\
&\quad = 0,
\end{aligned}$$

and therefore

$$\lim_{\mathbf{x}\to\mathbf{p}} (\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x}-\mathbf{p})) = \mathbf{0}.$$

But then

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

and thus the function φ is continuous at $\mathbf{p}$. The result follows. ∎

Proposition 9.12 *Let X be an open set in $\mathbb{R}^m$, let $\varphi\colon X \to \mathbb{R}^n$ be a function on X taking values in $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. Suppose that the first order partial derivatives of the components of the map φ are defined throughout the set X and are continuous at the point $\mathbf{p}$. Then the function φ is differentiable at the point $\mathbf{p}$.*

Proof Let $f_1, f_2, \ldots, f_n$ be the components of the map φ. It follows from Corollary 9.8 that

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

where

$$(D\varphi)_{\mathbf{p}}\,\mathbf{w} = ((\nabla f_1)_{\mathbf{p}} \,.\, \mathbf{w}, (\nabla f_2)_{\mathbf{p}} \,.\, \mathbf{w}, \ldots, (\nabla f_n)_{\mathbf{p}} \,.\, \mathbf{w})$$

for all $\mathbf{w} \in \mathbb{R}^m$. The function φ therefore satisfies the definition of differentiability at $\mathbf{p}$, as required. ∎

We now summarize some basic facts concerning differentiability of functions of several real variables. 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 $\mathbb{R}^n$. Now the mere existence of first order partial derivatives of the function φ at a point $\mathbf{p}$ of X is not sufficient to ensure that the function φ is differentiable at the point $\mathbf{p}$. Indeed examples of real-valued functions of two real variables discussed at the beginning of this section demonstrate that the existence of first order partial derivatives of the map φ throughout the domain X of φ is not even sufficient to ensure the continuity of the function φ. However if the first order partial derivatives of the components of the function φ exist and are bounded throughout the open set X then the function φ is at least continuous on X (see Corollary 9.4). And if these first order partial derivatives are continuous throughout the open set X then the function φ is differentiable on X (see Proposition 9.12).

9.5 The Jacobian Matrix of a Differentiable Function

Proposition 9.13 *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 at which the function φ is differentiable. Let $\mathbf{w}$ be an element of $\mathbb{R}^m$. Then*

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

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

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

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

In particular, if we set $(\mathbf{x} - \mathbf{p}) = t\mathbf{w}$, and $(\mathbf{x} - \mathbf{p}) = -t\mathbf{w}$, where t is a real variable, we can conclude that

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

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

It follows that

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

as required. ∎

Corollary 9.14 *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 at which the function φ is differentiable. Then the derivative $(D\varphi)_{\mathbf{p}}$ of φ at the point $\mathbf{p}$ is uniquely determined by the map φ.*

Proof The result of Proposition 9.13 shows that, for all $\mathbf{w} \in \mathbb{R}^m$, the value of $(D\varphi)_{\mathbf{p}}\mathbf{w}$ is expressible as a limit involving the function φ itself and is thus uniquely determined by the function φ itself. Thus there cannot be more than one linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ that can represent the derivative of the function φ at the point $\mathbf{p}$. ∎

Corollary 9.15 *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 at which the function φ is differentiable. Let $f_1, f_2, \ldots, f_n$ denote the components of the function let $\varphi: X \to \mathbb{R}^n$. Then the first order partial derivatives of the components of φ*

are all defined at the point $\mathbf{p}$*, and the derivative* $(D\varphi)_{\mathbf{p}}$ *of the map* φ *at the point* $\mathbf{p}$ *is represented by the* $n \times m$ *matrix whose coefficient in the* i*th row and* j*th column is equal to the value at* $\mathbf{p}$ *of the partial derivative* $\dfrac{\partial f_i}{\partial x_j}$ *of* f_i *with respect to the* j*th coordinate function* x_j *on* X*.*

Proof Let $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_m$ denote the standard basis of $\mathbb{R}^m$, where the ith component of the vector $\mathbf{e}_j$ is equal to 1 when $i = j$, but is equal to zero otherwise. Basic linear algebra ensures that the linear transformation $(D\varphi)_{\mathbf{p}}: \mathbb{R}^m \to \mathbb{R}^n$ is represented by the matrix $J(\mathbf{p})$ whose coefficient $J_{i,j}(\mathbf{p})$ in the ith row and jth column is equal to the ith component of the vector $(D\varphi)_{\mathbf{p}}\, \mathbf{e}_j$. It then follows from Proposition 9.14 that

$$J_{i,j}(\mathbf{p}) = \lim_{t \to 0} \frac{1}{t} \left(f_i(\mathbf{p} + t\mathbf{e}_j) - f_i(\mathbf{e}_j) \right) = \left. \frac{\partial f_i}{\partial x_j} \right|_{\mathbf{x} = \mathbf{p}},$$

as required. ∎

Let X be an open set in $\mathbb{R}^m$, let $\varphi: X \to \mathbb{R}^n$ be a differentiable function mapping X into $\mathbb{R}^n$. Corollary 9.15 ensures that the derivative of φ at any point $\mathbf{p}$ of X is the linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$ that sends $\mathbf{w} \in \mathbb{R}^m$ to $J(\mathbf{p})\mathbf{w}$, where J is the $n \times m$ matrix

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

of functions on X whose coefficients are the first order partial derivatives of the components $f_1, f_2, \ldots, f_n$ of the map φ. This matrix of partial derivatives is known as the *Jacobian matrix* of the map φ.

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

$$\varphi\left(\begin{pmatrix} x \\ y \end{pmatrix} \right) - \varphi\left(\begin{pmatrix} p \\ q \end{pmatrix} \right)$$

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

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

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

9.6 Sums, Differences and Multiples of Differentiable Functions

Proposition 9.16 *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

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

Moreover, given any real number c, the function cf is differentiable at $\mathbf{p}$ and

$$D(cf)_{\mathbf{p}} = c(Df)_{\mathbf{p}}.$$

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 $(Df)_{\mathbf{p}} - (Dg)_{\mathbf{p}}$.

Let c be a real number. Then

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

It follows that the function cf is differentiable at p, and $D(cf)_{\mathbf{p}} = c(Df)_{\mathbf{p}}$, as required. ∎

9.7 An Inequality limiting the Growth of a Differentiable Function

We shall derive an inequality bounding the growth of a function of of several real variables around a point where it is differentiable. The statement of the result makes reference to the *operator norm* of a linear transformation. Accordingly we now give the definition of operator norms.

Definition Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation. The *operator norm* $\|T\|_{\mathrm{op}}$ of T is the smallest non-negative real number with the property that $|T\mathbf{w}| \leq \|T\|_{\mathrm{op}}\,|\mathbf{w}|$ for all $\mathbf{w} \in \mathbb{R}^m$.

The operator norm $\|T\|_{\mathrm{op}}$ of a linear transformation $T\colon \mathbb{R}^m \to \mathbb{R}^n$ may be characterized as the maximum value attained by $|T\mathbf{w}|$ as $\mathbf{w}$ ranges over all vectors in $\mathbb{R}^m$ that satisfy $|\mathbf{w}| = 1$.

Lemma 9.17 *Let $T\colon \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^n$, and let $\|T\|_{\mathrm{op}}$ denote the operator norm of T. Also let $A_{i,j} = T\mathbf{e}_i\,.\,T\mathbf{e}_j$ for all integers i and j between 1 and m, let A denote the matrix whose coefficient in the ith row and jth column is $A_{i,j}$, and let $\lambda_{\max}$ denote the maximum eigenvalue of the real symmetric matrix A. Then $\|T\|_{\mathrm{op}} = \sqrt{\lambda_{\max}}$.*

Proof The matrix A is a real symmetric matrix, and, given any such matrix, there exists an orthogonal matrix R, with transpose R^T for which RAR^T is diagonal. The inverse of the matrix R is then equal to its transpose R^T. Let λ_i denote the coefficient in the ith row and column of the diagonal matrix RAR^T for $i = 1, 2, \ldots, m$. Then $\lambda_1, \lambda_2, \ldots, \lambda_m$ are the eigenvalues of the matrix A.

Let $\mathbf{w} \in \mathbb{R}^m$, and let $\mathbf{w} = (w_1, w_2, \ldots, w_m)$. Then

$$|T\mathbf{w}|^2 = \left(\sum_{i=1}^m w_i \mathbf{e}_i\right) . \left(\sum_{j=1}^m w_j \mathbf{e}_j\right) = \sum_{i=1}^m \sum_{j=1}^m A_{i,j} w_i w_j.$$

Thus if we represent $\mathbf{w}$ in matrix algebra as a column vector with coefficients $w_1, w_2, \ldots, w_n$ then $|T\mathbf{w}|^2 = \mathbf{w}^T A\mathbf{w}$, where $\mathbf{w}^T$ denotes the row vector that is the transpose of the column vector $\mathbf{w}$.

Let $\mathbf{u} = R\mathbf{w}$, so that $u_k = \sum\limits_{j=1}^m R_{k,j} w_j$ for $k = 1, 2, \ldots, m$. Then $\mathbf{w} = R^T\mathbf{u}$, and therefore

$$|T\mathbf{w}|^2 = \mathbf{w}^T A\mathbf{w} = \mathbf{u}^T RAR^T\mathbf{u} = \sum_{k=1}^m \lambda_k u_k^2 \leq \lambda_{\max} \sum_{k=1}^m u_k^2.$$

Moreover

$$\sum_{k=1}^n u_k^2 = \mathbf{u}^T\mathbf{u} = \mathbf{w}^T R^T R\mathbf{w} = \mathbf{w}^T\mathbf{w} = |\mathbf{w}|^2.$$

We conclude therefore that

$$|T\mathbf{w}|^2 \leq \lambda_{\max}|\mathbf{w}|^2$$

for all $\mathbf{w} \in \mathbb{R}^m$. Moreover if $\mathbf{w} = R^T\mathbf{e}_k$, where k is an integer between 1 and m for which $\lambda_k = \lambda_{\max}$, then

$$|T\mathbf{w}|^2 = \lambda_{\max}|\mathbf{w}|^2.$$

The result follows. ∎

Proposition 9.18 *Let X be an open set in $\mathbb{R}^m$, let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, let $\mathbf{p}$ be a point of X at which the function φ is differentiable, and let M be a real number satisfying $M > \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$, where $\|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$ denotes the operator norm of the derivative $(D\varphi)_{\mathbf{p}}$ of φ at $\mathbf{p}$. Then there exists some positive real number δ such that*

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \leq M\,|\mathbf{x} - \mathbf{p}|$$

for all points $\mathbf{x}$ of X satisfying $|\mathbf{x} - \mathbf{p}| < \delta$.

Proof Let $\varepsilon = M - \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$. Then $\varepsilon > 0$. Now

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

for all $\mathbf{v} \in \mathbb{R}^m$. The definition of differentiability ensures that

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

and therefore there exists some positive real number δ such that

$$\frac{1}{|\mathbf{x} - \mathbf{p}|}\,|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| < \varepsilon$$

for all $\mathbf{x} \in X$ satisfying $0 < |\mathbf{x} - \mathbf{p}| < \delta$. Then

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \leq |(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| + \varepsilon|\mathbf{x} - \mathbf{p}|$$

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

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

for all $\mathbf{x} \in X$. It follows that

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

for all $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$, as required. ∎

9.8 The Product Rule for Functions of Several Variables

Let $f\colon X \to \mathbb{R}$ and $g\colon X \to \mathbb{R}$ be real-valued functions defined over a subset X of $\mathbb{R}^m$. We denote by $f \cdot g$ the product of the functions f and g, defined so that $(f \cdot g)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$ for all $\mathbf{x} \in X$.

Proposition 9.19 (Product Rule) *Let X be an open set in $\mathbb{R}^m$, let $f\colon X \to \mathbb{R}$ and $g\colon X \to \mathbb{R}$ be real-valued functions on X, and 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 differentiability of the functions f and g at $\mathbf{p}$ ensures the existence of positive real numbers ensures that there exist positive real numbers M, N and δ such that $\mathbf{x} \in X$,

$$|f(\mathbf{x}) - f(\mathbf{p})| \leq M\,|\mathbf{x} - \mathbf{p}|$$

and

$$|g(\mathbf{x}) - g(\mathbf{p})| \leq N\,|\mathbf{x} - \mathbf{p}|$$

at all points $\mathbf{x}$ of $\mathbb{R}^m$ that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$. (This follows from Proposition 9.18.) Let $h\colon X \to \mathbb{R}$ be the real-valued function on X defined so that

$$\begin{aligned} h(\mathbf{x}) &= (f(\mathbf{x}) - f(\mathbf{p}))(g(\mathbf{x}) - g(\mathbf{p})) \\ &= f(\mathbf{x})g(\mathbf{x}) + f(\mathbf{p})g(\mathbf{p}) - f(\mathbf{p})g(\mathbf{x}) - f(\mathbf{x})g(\mathbf{p}) \end{aligned}$$

for all $\mathbf{x} \in X$. Then $h(\mathbf{p}) = 0$ and

$$\begin{aligned} \frac{h(\mathbf{x})}{|\mathbf{x} - \mathbf{p}|} &= \frac{1}{|\mathbf{x} - \mathbf{p}|}|f(\mathbf{x}) - f(\mathbf{p})|\,|g(\mathbf{x}) - g(\mathbf{p})| \\ &\leq MN\,|\mathbf{x} - \mathbf{p}| \end{aligned}$$

at all points $\mathbf{x}$ of $\mathbb{R}^m$ that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$. It follows that

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

It then follows from the definition of differentiability that the function $h\colon X \to \mathbb{R}$ is differentiable at the point $\mathbf{p}$, and $(Dh)_{\mathbf{p}} = 0$.

Now

$$f(\mathbf{x})g(\mathbf{x}) = f(\mathbf{p})g(\mathbf{x}) + g(\mathbf{p})f(\mathbf{x}) - f(\mathbf{p})g(\mathbf{p}) + h(\mathbf{x})$$

for all $\mathbf{x} \in X$. Differentiating, and using the fact that $(Dh)_{\mathbf{p}} = 0$, we find that $f \cdot g$ is differentiable at $\mathbf{p}$, and

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

as required. ∎

9.9 The Chain Rule for Functions of Several Variables

Proposition 9.20 (Chain Rule) *Let X and Y be open sets in $\mathbb{R}^m$ and $\mathbb{R}^n$ respectively, let $\varphi: X \to \mathbb{R}^n$ and $\psi: Y \to \mathbb{R}^k$ be functions mapping X and Y into $\mathbb{R}^n$ and $\mathbb{R}^k$ respectively, where $\varphi(X) \subset Y$, and let $\mathbf{p}$ be a point of X. Suppose that φ is differentiable at $\mathbf{p}$ and that ψ is differentiable at $\varphi(\mathbf{p})$. Then the composition $\psi \circ \varphi: X \to \mathbb{R}^k$ is differentiable at $\mathbf{p}$, and*

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

Thus the derivative of the composition $\psi \circ \varphi$ of the functions at the point $\mathbf{p}$ is the composition of the derivatives of the functions φ and ψ at $\mathbf{p}$ and $\varphi(\mathbf{p})$ respectively.

Proof The differentiability of the functions φ and ψ at $\mathbf{p}$ and $\varphi(\mathbf{p})$ respectively ensures that there exist positive real numbers M, N, δ_1 and η_1 such that the following conditions hold: $\mathbf{x} \in X$ and $|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \leq M|\mathbf{x} - \mathbf{p}|$ for all $\mathbf{x} \in \mathbb{R}^m$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta_1$; $\mathbf{y} \in Y$ and $|\psi(\mathbf{y}) - \psi(\varphi(\mathbf{p}))| \leq N\,|\mathbf{y} - \varphi(\mathbf{p})|$ for all $\mathbf{y} \in \mathbb{R}^n$ satisfying $|\mathbf{y} - \varphi(\mathbf{p})| < \eta_1$; $|(D\psi)_{\varphi(\mathbf{p})}\mathbf{w}| \leq N\,|\mathbf{w}|$ for all $\mathbf{w} \in \mathbb{R}^n$. (This follows from Proposition 9.18.)

Let some positive real number ε be given. It follows from the differentiability of ψ at $\varphi(\mathbf{p})$ that there exists some real number η_2, where $0 < \eta_2 \leq \eta_1$, such that

$$\left|\psi(\mathbf{y}) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(\mathbf{y} - \varphi(\mathbf{p}))\right| \leq \frac{\varepsilon}{2M}|\mathbf{y} - \varphi(\mathbf{p})|$$

for all $\mathbf{y} \in Y$ satisfying $|\mathbf{y} - \varphi(\mathbf{p})| < \eta_2$. (This follows from a direct application of Lemma 9.10.) Let some real number δ_2 be chosen so that $0 < \delta_2 \leq \delta_1$ and $M\delta_2 \leq \eta_2$. If $\mathbf{x} \in \mathbb{R}^m$ satisfies $|\mathbf{x} - \mathbf{p}| < \delta_2$ then $\mathbf{x} \in X$ and $|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \leq M|\mathbf{x} - \mathbf{p}| < \eta_2$, and therefore

$$\begin{aligned}\left|\psi(\varphi(\mathbf{x})) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(\varphi(\mathbf{x}) - \varphi(\mathbf{p}))\right| &\leq \frac{\varepsilon}{2M}|\varphi(\mathbf{x}) - \varphi(\mathbf{p})| \\ &\leq \tfrac{1}{2}\,\varepsilon|\mathbf{x} - \mathbf{p}|.\end{aligned}$$

Now it follows from the differentiability of φ at $\mathbf{p}$ that there exists some real number δ satisfying the inequalities $0 < \delta \leq \delta_2$ that is small enough to ensure that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| \leq \frac{\varepsilon}{2N} |\mathbf{x} - \mathbf{p}|$$

for all $\mathbf{x} \in \mathbb{R}^m$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$. Now $|(D\psi)_{\varphi(\mathbf{p})}\mathbf{w}| \leq N\,|\mathbf{w}|$ for all $\mathbf{w} \in \mathbb{R}^n$. It follows that

$$\begin{aligned} &\bigl|(D\psi)_{\varphi(\mathbf{p})}(\varphi(\mathbf{x}) - \varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})\bigr| \\ &\qquad \leq N\,|\varphi(\mathbf{x}) - \varphi(\mathbf{p}) - (D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})| \\ &\qquad \leq \tfrac{1}{2}\varepsilon|\mathbf{x} - \mathbf{p}| \end{aligned}$$

for all $\mathbf{x} \in \mathbb{R}^m$ satisfying $|\mathbf{x} - \mathbf{p}| < \delta$.

The inequalities obtained above ensure that $\mathbf{x} \in X$ and

$$\begin{aligned} &\bigl|\psi(\varphi(\mathbf{x})) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})\bigr| \\ &\quad \leq \bigl|\psi(\varphi(\mathbf{x})) - \psi(\varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(\varphi(\mathbf{x}) - \varphi(\mathbf{p}))\bigr| \\ &\qquad + \bigl|(D\psi)_{\varphi(\mathbf{p})}(\varphi(\mathbf{x}) - \varphi(\mathbf{p})) - (D\psi)_{\varphi(\mathbf{p})}(D\varphi)_{\mathbf{p}}(\mathbf{x} - \mathbf{p})\bigr| \\ &\quad \leq \varepsilon|\mathbf{x} - \mathbf{p}| \end{aligned}$$

at all points $\mathbf{x}$ of $\mathbb{R}^m$ that satisfy $|\mathbf{x} - \mathbf{p}| < \delta$. It follows from this that the composition function $\psi \circ \varphi$ is differentiable at $\mathbf{p}$, and that $(D(\psi \circ \varphi))_{\mathbf{p}} = (D\psi)_{\varphi(\mathbf{p})} \circ (D\varphi)_{\mathbf{p}}$, as required. ∎

10 The Inverse and Implicit Function Theorems

10.1 Contraction Mappings on Closed Subsets of Euclidean Spaces

Theorem 10.1 *Let F be a closed subset of $\mathbb{R}^n$, let r be a real number satisfying $0 < r < 1$, and let $\varphi: F \to F$ be a continuous map from F to itself with the property that*

$$|\varphi(\mathbf{x}') - \varphi(\mathbf{x}'')| \leq r|\mathbf{x}' - \mathbf{x}''|$$

for all $\mathbf{x}', \mathbf{x}'' \in F$. Then there exists a unique point $\mathbf{x}^*$ of F for which $\varphi(\mathbf{x}^*) = \mathbf{x}^*$.

Proof Choose $\mathbf{x}_0 \in F$, and let $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ be the infinite sequence of points of F defined such that $\mathbf{x}_j = \varphi(\mathbf{x}_{j-1})$ for all positive integers j. Then

$$|\mathbf{x}_{j+1} - \mathbf{x}_j| \leq r|\mathbf{x}_j - \mathbf{x}_{j-1}|$$

for all positive integers j. It follows that

$$|\mathbf{x}_{j+1} - \mathbf{x}_j| \leq r^j|\mathbf{x}_1 - \mathbf{x}_0|$$

for all positive integers j, and therefore

$$|\mathbf{x}_k - \mathbf{x}_j| \leq \frac{r^j - r^k}{1 - r}|\mathbf{x}_1 - \mathbf{x}_0| \leq \frac{r^j}{1 - r}|\mathbf{x}_1 - \mathbf{x}_0|$$

for all positive integers j and k satisfying $j < k$.

Now the inequality $r < 1$ ensures that, given any positive real number ε, there exists a positive integer N large enough to ensure that $r^j|\mathbf{x}_1 - \mathbf{x}_0| < (1-r)\varepsilon$ for all integers j satisfying $j \geq N$. Then $|\mathbf{x}_k - \mathbf{x}_j| < \varepsilon$ for all positive integers j and k satisfying $k > j \geq N$. The infinite sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ is thus a Cauchy sequence of points of $\mathbb{R}^n$. Now all Cauchy sequences in $\mathbb{R}^n$ are convergent (see Theorem 2.8). We conclude therefore that the infinite sequence $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots$ is convergent. Let $\mathbf{x}^* = \lim\limits_{j \to +\infty} \mathbf{x}_j$. Then $\mathbf{x}^* \in F$, because F is closed in $\mathbb{R}^n$. Moreover

$$\mathbf{x}^* = \lim_{j \to +\infty} \mathbf{x}_{j+1} = \lim_{j \to +\infty} \varphi(\mathbf{x}_j) = \varphi\left(\lim_{j \to +\infty} \mathbf{x}_j\right) = \varphi(\mathbf{x}^*).$$

We have thus proved the existence of a point $\mathbf{x}^*$ of F for which $\varphi(\mathbf{x}^*) = \mathbf{x}^*$.

If $\tilde{\mathbf{x}}$ belongs to F, and if $\varphi(\tilde{\mathbf{x}}) = \tilde{\mathbf{x}}$ then

$$|\tilde{\mathbf{x}} - \mathbf{x}^*| = |\varphi(\tilde{\mathbf{x}}) - \varphi(\mathbf{x}^*)| \leq r|\tilde{\mathbf{x}} - \mathbf{x}^*|.$$

But $r < 1$. It follows that the Euclidean distance $|\tilde{\mathbf{x}} - \mathbf{x}^*|$ from $\tilde{\mathbf{x}}$ to $\mathbf{x}^*$ cannot be strictly positive, and therefore $\tilde{\mathbf{x}} = \mathbf{x}^*$. We conclude therefore that $\mathbf{x}^*$ is the unique point of F for which $\varphi(\mathbf{x}^*) = \mathbf{x}^*$, as required. ∎

10.2 The Inverse Function Theorem

Lemma 10.2 *Let X be an open set in $\mathbb{R}^m$, let $\varphi\colon X \to \mathbb{R}^n$ be a differentiable function mapping X into $\mathbb{R}^n$, let $\mathbf{p}$ be a point of X, and let c be a positive real number. Suppose that $|\mathbf{x} - \mathbf{p}| \leq c|\varphi(\mathbf{x}) - \varphi(\mathbf{p})|$ for all points $\mathbf{x}$ of X. Then $|\mathbf{v}| \leq c|(D\varphi)_{\mathbf{p}}\mathbf{v}|$ for all $\mathbf{v} \in \mathbb{R}^m$.*

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

$$t|\mathbf{v}| = |(\mathbf{p} + t\mathbf{v}) - \mathbf{p}| \leq c|\varphi(\mathbf{p} + t\mathbf{v}) - \varphi(\mathbf{p})|$$

for all positive real numbers t small enough to ensure that $\mathbf{p} + t\mathbf{v} \in X$. Now

$$(D\varphi)_{\mathbf{p}}\mathbf{v} = \lim_{t \to 0^+} \frac{\varphi(\mathbf{p} + t\mathbf{v}) - \varphi(\mathbf{p})}{t}$$

(see Proposition 9.13). It follows that

$$\begin{aligned} |\mathbf{v}| &\leq \lim_{t \to 0^+} c\left|\frac{\varphi(\mathbf{p} + t\mathbf{v}) - \varphi(\mathbf{p})}{t}\right| = c\left|\lim_{t \to 0^+} \frac{\varphi(\mathbf{p} + t\mathbf{v}) - \varphi(\mathbf{p})}{t}\right| \\ &= c|(D\varphi)_{\mathbf{p}}\mathbf{v}|, \end{aligned}$$

as required. ∎

Proposition 10.3 *Let X be an open set in $\mathbb{R}^n$, let $\varphi\colon X \to \mathbb{R}^n$ be a differentiable function on X, and let $\mathbf{p}$ be a point of X at which the derivative of φ is both invertible and continuous. Then there exist positive real numbers r, s and c such that the following properties hold:*

(i) if $\mathbf{x} \in \mathbb{R}^n$ satisfies $|\mathbf{x} - \mathbf{p}| \leq r$ then $x \in X$;

(ii) if $\mathbf{y} \in \mathbb{R}^n$ satisfies $|\mathbf{y} - \varphi(\mathbf{p})| < s$ then there exists $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < r$ for which $\varphi(\mathbf{x}) = \mathbf{y}$;

(iii) $|\mathbf{x}' - \mathbf{x}''| \leq c|\varphi(\mathbf{x}') - \varphi(\mathbf{x}'')|$ for all points $\mathbf{x}'$ and $\mathbf{x}''$ of X for which $|\mathbf{x}' - \mathbf{p}| \leq r$ and $|\mathbf{x}'' - \mathbf{p}| \leq r$.

Proof The derivative $(D\varphi)_{\mathbf{p}}: \mathbb{R}^n \to \mathbb{R}^n$ of φ at the point $\mathbf{p}$ is an invertible linear transformation, by assumption. Let $T = (D\varphi)_{\mathbf{p}}^{-1}$, let a positive real number c be chosen such that $2|T\mathbf{x}| \leq c$ for all $\mathbf{x} \in \mathbb{R}^n$ satisfying $|\mathbf{x}| = 1$, and let $\psi: X \to \mathbb{R}^n$ be defined such that

$$\psi(\mathbf{x}) = \mathbf{x} - T(\varphi(\mathbf{x}) - \mathbf{q})$$

for all $\mathbf{x} \in X$, where $\mathbf{q} = \varphi(\mathbf{p})$.

Now the derivative of any linear transformation at any point is equal to that linear transformation (see Lemma 9.9). It follows from the Chain Rule (Proposition 9.20) that the derivative of the composition function $T \circ \varphi$ at any point $\mathbf{x}$ of X is equal to $T(D\varphi)_{\mathbf{x}}$. It follows that $(D\psi)_{\mathbf{x}} = I - T(D\varphi)_{\mathbf{x}}$ for all $\mathbf{x} \in X$, where I denotes the identity operator on $\mathbb{R}^n$. In particular $(D\psi)_{\mathbf{p}} = I - T(D\varphi)_{\mathbf{p}} = 0$. Moreover $\psi(\mathbf{p}) = \mathbf{p}$. It then follows from a straightforward application of Corollary 9.7 that there exists a positive real number r small enough to ensure both that $\mathbf{x} \in X$ for all elements $\mathbf{x}$ of $\mathbb{R}^n$ satisfying $|\mathbf{x} - \mathbf{p}| \leq r$ and also that

$$|\psi(\mathbf{x}') - \psi(\mathbf{x}'')| \leq \tfrac{1}{2}|\mathbf{x}' - \mathbf{x}''|$$

for all points $\mathbf{x}'$ and $\mathbf{x}''$ of X for which $|\mathbf{x}' - \mathbf{p}| \leq r$ and $|\mathbf{x}'' - \mathbf{p}| \leq r$.

Let $\mathbf{x}'$ and $\mathbf{x}''$ be points of X for which $|\mathbf{x}' - \mathbf{p}| \leq r$ and $|\mathbf{x}'' - \mathbf{p}| \leq r$. Then

$$\psi(\mathbf{x}') - \psi(\mathbf{x}'') = \mathbf{x}' - \mathbf{x}'' - T(\varphi(\mathbf{x}') - \varphi(\mathbf{x}'')),$$

because T is a linear transformation, and therefore

$$\begin{array}{rcl} |\mathbf{x}' - \mathbf{x}''| & = & |\psi(\mathbf{x}') - \psi(\mathbf{x}'') + T(\varphi(\mathbf{x}') - \varphi(\mathbf{x}''))| \\ & \leq & |\psi(\mathbf{x}') - \psi(\mathbf{x}'')| + |T(\varphi(\mathbf{x}') - \varphi(\mathbf{x}''))| \\ & \leq & \frac{1}{2}|\mathbf{x}' - \mathbf{x}''| + |T(\varphi(\mathbf{x}') - \varphi(\mathbf{x}''))|\,. \end{array}$$

Subtracting $\frac{1}{2}|\mathbf{x}' - \mathbf{x}''|$ from both sides of this inequality, and multiplying by 2, we deduce that

$$|\mathbf{x}' - \mathbf{x}''| \leq 2\,|T(\varphi(\mathbf{x}') - \varphi(\mathbf{x}''))| \leq c|\varphi(\mathbf{x}') - \varphi(\mathbf{x}'')|,$$

for all points $\mathbf{x}'$ and $\mathbf{x}''$ of X satisfying $|\mathbf{x}' - \mathbf{p}| \leq r$ and $|\mathbf{x}'' - \mathbf{p}| \leq r$.

Now let

$$F = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{p}| \leq r\}.$$

Then F is a closed subset of $\mathbb{R}^n$, and $F \subset X$. Moreover $|\psi(\mathbf{x}') - \psi(\mathbf{x}'')| \leq \frac{1}{2}|\mathbf{x}' - \mathbf{x}''|$ for all $\mathbf{x}' \in F$ and $\mathbf{x}'' \in F$.

Let $\mathbf{y} \in \mathbb{R}^n$ satisfy $|\mathbf{y} - \mathbf{q}| < s$, where $s = r/c$, let $\mathbf{z} = \mathbf{p} + T(\mathbf{y} - \mathbf{q})$, and let

$$\theta(\mathbf{x}) = \psi(\mathbf{x}) + \mathbf{z} - \mathbf{p}$$

for all $\mathbf{x} \in X$. Now the choice of c then ensures that

$$|\mathbf{z} - \mathbf{p}| \leq \tfrac{1}{2}c|\mathbf{y} - \mathbf{q}| \leq \tfrac{1}{2}cs = \tfrac{1}{2}r.$$

If $\mathbf{x} \in X$ satisfies $|\mathbf{x} - \mathbf{p}| \leq r$, and if

$$\mathbf{x}' = \psi(\mathbf{x}) + \mathbf{z} - \mathbf{p},$$

then

$$|\mathbf{x}' - \mathbf{z}| = |\psi(\mathbf{x}) - \mathbf{p}| = |\psi(\mathbf{x}) - \psi(\mathbf{p})| \leq \tfrac{1}{2}|\mathbf{x} - \mathbf{p}| \leq \tfrac{1}{2}r,$$

and therefore

$$|\mathbf{x}' - \mathbf{p}| \leq |\mathbf{x}' - \mathbf{z}| + |\mathbf{z} - \mathbf{p}| < r.$$

We conclude therefore that θ maps the closed set F into itself, where

$$F = \{\mathbf{x} \in X : |\mathbf{x} - \mathbf{p}| \leq r\}.$$

Moreover $|\theta(\mathbf{x})| < r$ for all $\mathbf{x} \in F$ and

$$|\theta(\mathbf{x}') - \theta(\mathbf{x}'')| = |\psi(\mathbf{x}') - \psi(\mathbf{x}'')| \leq \tfrac{1}{2}|\mathbf{x}' - \mathbf{x}''|$$

for all $\mathbf{x}' \in F$ and $\mathbf{x}'' \in F$. It then follows from Theorem 10.1 that there exists a point $\mathbf{x}$ of F for which $\theta(\mathbf{x}) = \mathbf{x}$. Then $|\mathbf{x} - \mathbf{p}| < r$. Also

$$\mathbf{x} = \theta(\mathbf{x}) = \psi(\mathbf{x}) + \mathbf{z} - \mathbf{p} = \mathbf{x} - T(\varphi(\mathbf{x}) - \mathbf{q}) + \mathbf{z} - \mathbf{p},$$

where $\mathbf{q} = \varphi(\mathbf{p})$, and thus $\mathbf{z} - \mathbf{p} = T(\varphi(\mathbf{x}) - \mathbf{q})$. But $\mathbf{z} - \mathbf{p} = T(\mathbf{y} - \mathbf{q})$. It follows that $T\mathbf{y} = T(\varphi(\mathbf{x}))$, and therefore

$$\mathbf{y} = (D\varphi)_{\mathbf{p}}(T\mathbf{y}) = (D\varphi)_{\mathbf{p}}(T(\varphi(\mathbf{x})) = \varphi(\mathbf{x}).$$

We have thus shown that, given any element $\mathbf{y}$ of $\mathbb{R}^n$ satisfying $|\mathbf{y} - \mathbf{q}| < s$, there exists $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < r$ for which $\varphi(\mathbf{x}) = \mathbf{y}$. This completes the proof. ∎

Theorem 10.4 (Inverse Function Theorem) *Let $\varphi: X \to \mathbb{R}^n$ be a continuously differentiable function defined over an open set X in n-dimensional Euclidean space $\mathbb{R}^n$ and mapping X into $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X. Suppose that the derivative $(D\varphi)_{\mathbf{p}}: \mathbb{R}^n \to \mathbb{R}^n$ of the map φ at the point $\mathbf{p}$ is an invertible linear transformation. Then there exists an open set W in $\mathbb{R}^n$ and a continuously differentiable function $\mu: W \to X$ that satisfies the following conditions:—*

(i) $\mu(W)$ is an open set in $\mathbb{R}^n$ contained in X, and $\mathbf{p} \in \mu(W)$;

(ii) $\varphi(\mu(\mathbf{y})) = \mathbf{y}$ for all $\mathbf{y} \in W$.

Proof It follows from Proposition 10.3 that there exist positive real numbers r, s and c such that the following properties hold: if $\mathbf{x} \in \mathbb{R}^n$ satisfies $|\mathbf{x}-\mathbf{p}| \leq r$ then $x \in X$; if $\mathbf{y} \in \mathbb{R}^n$ satisfies $|\mathbf{y} - \varphi(\mathbf{p})| < s$ then there exists $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{p}| < r$ for which $\varphi(\mathbf{x}) = \mathbf{y}$; $|\mathbf{x}' - \mathbf{x}''| \leq c|\varphi(\mathbf{x}') - \varphi(\mathbf{x}'')|$ for all points $\mathbf{x}'$ and $\mathbf{x}''$ of X for which $|\mathbf{x}' - \mathbf{p}| \leq r$ and $|\mathbf{x}'' - \mathbf{p}| \leq r$. It then follows from Lemma 10.2 that $|(D\varphi)_{\mathbf{u}}\mathbf{v}| \geq c|\mathbf{v}|$ for all $\mathbf{u} \in X$ satisfying $|\mathbf{u} - \mathbf{p}| < r$ and for all $\mathbf{v} \in \mathbb{R}^n$.

Let

$$W = \{\mathbf{y} \in \mathbb{R}^n : |\mathbf{y} - \varphi(\mathbf{p})| < s\}.$$

If $\mathbf{y}$ is a point of W, there exists a point $\mathbf{x}$ of X such that $|\mathbf{x} - \mathbf{p}| < r$ and $\varphi(\mathbf{x}) = \mathbf{y}$. There cannot exist more than one point of X with this property because if $\mathbf{x}'$ is a point of X distinct from $\mathbf{x}$, and if $|\mathbf{x}' - \mathbf{p}| < r$, then

$$|\varphi(\mathbf{x}') - \mathbf{y}| \geq c|\mathbf{x}' - \mathbf{x}| > 0.$$

Therefore there is a well-defined function $\mu\colon W \to \mathbb{R}^n$ characterized by the property that, for each $\mathbf{y} \in W$, $\mu(\mathbf{y})$ is the unique point of X for which $|\mu(\mathbf{y}) - \mathbf{p}| < r$ and $\varphi(\mu(\mathbf{y})) = \mathbf{y}$.

We next show that $\mu(W)$ is an open subset of $\mathbb{R}^n$. Let $\mathbf{u} \in \mu(W)$. Then $|\mathbf{u}-\mathbf{p}| < r$, and there exists $\mathbf{w} \in W$ for which $\mu(\mathbf{w}) = \mathbf{u}$. But then $\varphi(\mathbf{u}) = \mathbf{w}$, and thus $\mathbf{u} \in \varphi^{-1}(W)$. We conclude that

$$\mu(W) \subset \varphi^{-1}(W) \cap \{\mathbf{x} \in X : |\mathbf{x} - \mathbf{p}| < r\}.$$

Conversely let $\mathbf{u}$ be a point of $\varphi^{-1}(W)$ satisfying $|\mathbf{u} - \mathbf{p}| < r$, and let $\mathbf{w} = \varphi(\mathbf{u})$. Then $\mathbf{w} \in W$ and $\mu(\mathbf{w}) = \mathbf{u}$, and therefore $\mathbf{u} \in \mu(W)$. We conclude from this that

$$\mu(W) = \varphi^{-1}(W) \cap \{\mathbf{x} \in X : |\mathbf{x} - \mathbf{p}| < r\}.$$

It follows that $\mu(W)$ is the intersection of two open subsets of X, and must therefore itself be open in X. Now X itself is open in $\mathbb{R}^n$. It follows that $\mu(W)$ is indeed an open subset of $\mathbb{R}^n$.

Let $\mathbf{w} \in W$, and let $\mathbf{u} = \mu(\mathbf{w})$. Then $|\mathbf{u} - \mathbf{p}| < r$. Let some positive real number ε be given. The differentiability of the map φ at $\mathbf{u}$ ensures the existence of a positive real number δ such that $\eta + |\mathbf{u} - \mathbf{p}| \leq r$ and

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{u}) - (D\varphi)_{\mathbf{u}}(\mathbf{x} - \mathbf{u})| \leq \frac{\varepsilon}{c^2}|\mathbf{x} - \mathbf{u}|$$

for all $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{u}| \leq c\delta$. Let $\mathbf{y} \in W$ satisfy $|\mathbf{y} - \mathbf{w}| < \delta$, and let $\mathbf{x} = \mu(\mathbf{y})$. Then $\varphi(\mathbf{x}) = \mathbf{y}$ and $\varphi(\mathbf{u}) = \mathbf{w}$, and therefore

$$|\mathbf{x} - \mathbf{u}| \leq c|\varphi(\mathbf{x}) - \varphi(\mathbf{u})| = c|\mathbf{y} - \mathbf{w}| < c\delta.$$

It follows that

$$|\mathbf{y} - \mathbf{w} - (D\varphi)_{\mathbf{u}}(\mathbf{x} - \mathbf{u})| \leq \frac{\varepsilon}{c^2}|\mathbf{x} - \mathbf{u}| \leq \frac{\varepsilon}{c}|\mathbf{y} - \mathbf{w}|,$$

and therefore

$$\begin{array}{rcl} \left|(D\varphi)_{\mathbf{u}}^{-1}(\mathbf{y} - \mathbf{w}) - (\mathbf{x} - \mathbf{u})\right| & \leq & c\,|\mathbf{y} - \mathbf{w} - (D\varphi)_{\mathbf{u}}(\mathbf{x} - \mathbf{u})| \\ & \leq & \varepsilon|\mathbf{y} - \mathbf{w}|. \end{array}$$

But $\mathbf{x} - \mathbf{u} = \mu(\mathbf{y}) - \mu(\mathbf{w})$. We conclude therefore that, given any positive real number ε, there exists some positive real number δ such that

$$\left|\mu(\mathbf{y}) - \mu(\mathbf{w}) - (D\varphi)_{\mathbf{u}}^{-1}(\mathbf{y} - \mathbf{w})\right| \leq \varepsilon|\mathbf{y} - \mathbf{w}|$$

for all points $\mathbf{y}$ of W satisfying $|\mathbf{y}-\mathbf{w}| < \delta$. It follows that the map $\mu: W \to X$ is differentiable at $\mathbf{w}$, and moreover

$$(D\mu)_{\mathbf{w}} = (D\varphi)_{\mathbf{u}}^{-1} = (D\varphi)_{\mu(\mathbf{y})}^{-1}.$$

Now the map $\mu: W \to X$ is continuous, because it is differentiable. Also the coefficients of the Jacobian matrix representing the derivative of φ at points $\mathbf{x}$ of $\mu(W)$ are continuous functions of $\mathbf{x}$ on $\mu(W)$. It follows that the coefficients of the inverse of the Jacobian matrix of the map φ are also continuous functions of $\mathbf{x}$ on $\mu(W)$. Each coefficient of the Jacobian matrix of the map μ is thus the composition of the continuous map μ with a continuous real-valued function on $\mu(W)$, and must therefore itself be a continuous real-valued function on W. It follows that the map $\mu: W \to X$ is continuously differentiable on W. This completes the proof. ∎

10.3 The Implicit Function Theorem

Theorem 10.5 *Let X be an open set in $\mathbb{R}^n$, let $f_1, f_2, \ldots, f_m$ be a continuously differentiable real-valued functions on X, where $m < n$, let*

$$M = \{\mathbf{x} \in X : f_i(\mathbf{x}) = 0 \text{ for } i = 1, 2, \ldots, m\},$$

and let $\mathbf{p}$ *be a point of* M*. Suppose that* $f_1, f_2, \ldots, f_m$ *are zero at* $\mathbf{p}$ *and that the matrix*

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

is invertible at the point $\mathbf{p}$*. Then there exists an open neighbourhood* U *of* $\mathbf{p}$ *and continuously differentiable functions* $h_1, h_2, \ldots, h_m$ *of* $n - m$ *real variables, defined around* $(p_{m+1}, \ldots, p_n)$ *in* $\mathbb{R}^{n-m}$*, such that*

$$\begin{aligned} M \cap U \;=\; & \{(x_1, x_2, \ldots, x_n) \in U : \\ & \qquad x_i = h_i(x_{m+1}, \ldots, x_n) \text{ for } i = 1, 2, \ldots, m\}. \end{aligned}$$

Proof Let $\varphi\colon X \to \mathbb{R}^n$ be the continuously differentiable function defined such that

$$\varphi(\mathbf{x}) = \Big(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}), x_{m+1}, \ldots, x_n\Big)$$

for all $\mathbf{x} \in X$. (Thus the ith Cartesian component of the function φ is equal to f_i for $i \leq m$, but is equal to x_i for $m < i \leq n$.) Let J be the Jacobian matrix of φ at the point $\mathbf{p}$, and let $J_{i,j}$ denote the coefficient in the ith row and jth column of J. Then

$$J_{i,j} = \frac{\partial f_i}{\partial x_j}$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. Also $J_{i,i} = 1$ if $i > m$, and $J_{i,j} = 0$ if $i > m$ and $j \neq i$. The matrix J can therefore be represented in block form as

$$J = \left(\begin{array}{c|c} J_0 & A \\ \hline 0 & I_{n-m} \end{array}\right),$$

where J_0 is the leading $m \times m$ minor of the matrix J, A is an $m \times (n-m)$ minor of the matrix J and I_{n-m} is the identity $(n-m) \times (n-m)$ matrix. It follows from standard properties of determinants that $\det J = \det J_0$. Moreover the hypotheses of the theorem require that $\det J_0 \neq 0$. Therefore $\det J \neq 0$. The derivative $(D\varphi)_{\mathbf{p}}$ of φ at the point $\mathbf{p}$ is represented by the Jacobian matrix J. It follows that $(D\varphi)_{\mathbf{p}}\colon \mathbb{R}^n \to \mathbb{R}^n$ is an invertible linear transformation.

The Inverse Function Theorem (Theorem 10.4) now ensures the existence of a continuously differentiable map $\mu\colon W \to X$ with the properties that $\mu(W)$ is an open subset of X and $\varphi(\mu(\mathbf{y})) = \mathbf{y}$ for all $\mathbf{y} \in W$.

Let $\mathbf{y}$ be a point of W, and let $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. Then $\mathbf{y} = \varphi(\mu(\mathbf{y}))$, and therefore $y_i = f_i(\mu(\mathbf{y}))$ for $i = 1, 2, \ldots, m$, and y_i is equal to the ith component of $\mu(\mathbf{y})$ when $m < i \leq n$.

Now $\mathbf{p} \in \mu(W)$. Therefore there exists some point $\mathbf{q}$ of W satisfying $\mu(\mathbf{q}) = \mathbf{p}$. Now $\mathbf{p} \in M$, and therefore $f_i(\mathbf{p}) = 0$ for $i = 1, 2, \ldots, m$. But $q_i = f_i(\mu(\mathbf{q})) = f_i(\mathbf{p})$ when $1 \leq i \leq m$. It follows that $q_i = 0$ when $1 \leq i \leq m$. Also $q_i = p_i$ when $i > m$.

Let g_i denote the ith Cartesian component of the continuously differentiable map $\mu: W \to \mathbb{R}^n$ for $i = 1, 2, \ldots, n$. Then $g_i: W \to \mathbb{R}$ is a continuously differentiable real-valued function on W for $i = 1, 2, \ldots, n$. If $(y_1, y_2, \ldots, y_n) \in W$ then

$$(y_1, y_2, \ldots, y_n) = \varphi(\mu(y_1, y_2, \ldots, y_n)).$$

It then follows from the definition of the map φ that y_i is the ith Cartesian component of $\mu(y_1, y_2, \ldots, y_n)$ when $i > m$, and thus

$$y_i = g_i(y_1, y_2, \ldots, y_n) \quad \text{when} \quad i > m.$$

Now $\mu(W)$ is an open set, and $\mathbf{p} \in \mu(W)$. It follows that there exists some positive real number δ such that $H(\mathbf{p}, \delta) \subset \mu(W)$. where

$$\begin{aligned} &H(\mathbf{p}, \delta) \\ &\quad = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : p_i - \delta < x_i < p_i + \delta \text{ for } i = 1, 2, \ldots, n\}. \end{aligned}$$

Let

$$\begin{aligned} D = {} & \{(z_1, z_2, \ldots, z_{n-m}) \in \mathbb{R}^{n-m} : p_{m+j} - \delta < z_j < p_{m+j} + \delta \\ & \quad \text{for } j = 1, 2, \ldots, n-m\}, \end{aligned}$$

and let $h_i: D \to \mathbb{R}$ be defined so that

$$h_i(z_1, z_2, \ldots, z_{n-m}) = g_i(0, 0, \ldots, 0, z_1, z_2, \ldots, z_{n-m})$$

for $i = 1, 2, \ldots, m$.

Let $\mathbf{x} \in H(\mathbf{p}, \delta)$, where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Then $\mathbf{x} \in \mu(W)$. There therefore exists $\mathbf{w} \in W$ for which $\mu(\mathbf{w}) = \mathbf{x}$. But the properties of the map μ ensure that $\mathbf{w} = \varphi(\mu(\mathbf{w}))$. It follows that

$$\mathbf{x} = \mu(\mathbf{w}) = \mu(\varphi(\mu(\mathbf{w}))) = \mu(\varphi(\mathbf{x})).$$

Thus

$$\begin{aligned} (x_1, x_2, \ldots, x_n) &= \mu(\varphi(\mathbf{x})) \\ &= \mu\Big(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}), x_{m+1}, \ldots, x_n\Big). \end{aligned}$$

On equating Cartesian components we find that

$$x_i = g_i\Big(f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x}), x_{m+1}, \ldots, x_n\Big).$$

for $i = 1, 2, \ldots, n$.

In particular, if $\mathbf{x} \in H(\mathbf{p}, \delta) \cap M$ then

$$f_1(\mathbf{x}) = f_2(\mathbf{x}) = \cdots = f_m(\mathbf{x}) = 0,$$

and therefore

$$\begin{aligned} x_i &= g_i\Big(0, 0, \ldots, 0, x_{m+1}, \ldots, x_n\Big) \\ &= h_i\Big(x_{m+1}, \ldots, x_n\Big). \end{aligned}$$

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

$$\begin{aligned} M \cap H(\mathbf{p}, \delta) \subset \{&(x_1, x_2, \ldots, x_n) \in H(\mathbf{p}, \delta) : \\ &x_i = h_i(x_{m+1}, \ldots, x_n) \text{ for } i = 1, 2, \ldots, m\}. \end{aligned}$$

Now let $\mathbf{x}$ be a point of $H(\mathbf{x}, \delta)$ whose Cartesian components $x_1, x_2, \ldots, x_n$ satisfy the equations

$$x_i = h_i(x_{m+1}, \ldots, x_n)$$

for $i = 1, 2, \ldots, m$. Then

$$x_i = g_i(0, 0, \ldots, 0, x_{m+1}, \ldots, x_n)$$

for $i = 1, 2, \ldots, m$. Now it was shown earlier that

$$y_i = g_i(y_1, y_2, \ldots, y_n)$$

for all $(y_1, y_2, \ldots, y_n) \in W$ when $i > m$. It follows from this that

$$x_i = g_i(0, 0, \ldots, 0, x_{m+1}, \ldots, x_n)$$

when $m < i \leq n$. The functions $g_1, g_2, \ldots, g_n$ are the Cartesian components of the map $\mu\colon W \to X$. We conclude therefore that

$$(x_1, x_2, \ldots, x_n) = \mu(0, 0, \ldots, 0, x_{m+1}, \ldots, x_n),$$

Applying the function φ to both sides of this equation we see that

$$\begin{aligned} \varphi(x_1, x_2, \ldots, x_n) &= \varphi(\mu(0, 0, \ldots, 0, x_{m+1}, \ldots, x_n)) \\ &= (0, 0, \ldots, 0, x_{m+1}, \ldots, x_n). \end{aligned}$$

It then follows from the definition of the map φ that

$$f_i(x_1, x_2, \ldots, x_n) = 0,$$

for $i = 1, 2, \ldots, m$. We have thus shown that if $\mathbf{x}$ is a point of $H(\mathbf{x}, \delta)$ whose Cartesian components $x_1, x_2, \ldots, x_n$ satisfy the equations

$$x_i = h_i(x_{m+1}, \ldots, x_n)$$

for $i = 1, 2, \ldots, m$ then $\mathbf{x} \in M$. The converse of this result was proved earlier. The proof of the theorem is therefore completed on taking $U = H(\mathbf{p}, \delta)$. ∎

11 Second Order Partial Derivatives and the Hessian Matrix

11.1 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 11.1 *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 $\mathbb{R}^2$ and therefore there exists some positive real number L such that $(a+h, b+k) \in X$ for all $(h,k) \in \mathbb{R}^2$ satisfying $|h| < L$ and $|k| < L$. Let

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

for all real numbers h and k satisfying $|h| < L$ and $|k| < L$. We use the Mean Value Theorem (Theorem 8.2) 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 8.2) that there exists some real number v lying between b and $b+k$ for which $q(b+k) - q(b) = kq'(v)$. But $q'(v) = f_y(a+h, v) - f_y(a,v)$. It follows that

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

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

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

Now let some positive real number ε be given. The function f_{xy} is continuous. Therefore there exists some real number δ satisfying $0 < \delta < L$ such that $|f_{xy}(a+h, b+k) - f_{xy}(a,b)| \leq \varepsilon$ whenever $|h| < \delta$ and $|k| < \delta$. It follows that

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

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

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

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 11.2 *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.

11.2 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 11.3 *Let f be a continuous real-valued function defined throughout an open ball in $\mathbb{R}^n$ of radius R about some point $\mathbf{p}$. Suppose that the partial derivatives of f of orders one and two exist and are continuous throughout this open ball. Then there exists some real number θ satisfying $0 < \theta < 1$ for which*

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

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 9.20) that

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

and

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

where

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

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 θ satisfying $0 < \theta < 1$. (see Proposition 8.7). 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 11.3 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 θ satisfying $0 < \theta < 1$.

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

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

If the partial derivatives of f of second order exist and are continuous then $H_{i,j}(\mathbf{p}) = H_{j,i}(\mathbf{p})$ for all i and j, by Corollary 11.2. 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 11.4 *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 11.5 *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 θ satisfying $0 < \theta < 1$ (where θ depends on $\mathbf{h}$) such that

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

where

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

(see Lemma 11.3).

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 11.6 *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 θ satisfying $0 < \theta < 1$ (where θ depends on $\mathbf{h}$) such that

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

where

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

(see Lemma 11.3). Suppose that the Hessian matrix $(H_{i,j}(\mathbf{p}))$ is positive definite. It follows from Lemma 11.4 that there exists some $\varepsilon > 0$ such that if $|H_{i,j}(\mathbf{x}) - H_{i,j}(\mathbf{p})| < \varepsilon$ for all i and j then $(H_{i,j}(\mathbf{x}))$ is positive definite.

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

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

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

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

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

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

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