Module MA2321: Analysis in Several Real Variables
Michaelmas Term 2016
Section 8: Differentiation of Functions of Several Real Variables

D. R. Wilkins

Contents

8 Differentiation of Functions of Several Real Variables 116

8.1 Review of Differentiability for Functions of One Real Variable 116
8.2 Derivatives of Functions of Several Variables 117
8.3 Differentiation of Square Matrices 119
8.4 Properties of Differentiable Functions of Several Real Variables 123
8.5 The Multidimensional Product Rule 124
8.6 The Multidimensional Chain Rule 126
8.7 Partial Derivatives and Continuous Differentiability 131
8.8 Functions with Continuous Partial Derivatives 132
8.9 Summary of Differentiability Results 137

8 Differentiation of Functions of Several Real Variables

8.1 Review of Differentiability for Functions of One Real Variable

Let $f: D \to \mathbb{R}$ be a real-valued function defined on a subset D of the set $\mathbb{R}$ of real numbers. Let s be an element in the interior of D. the function f is *differentiable* at s if and only if

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

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

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

Lemma 8.1 *Let $f: D \to \mathbb{R}$ be a real-valued function defined on some subset D of the set of real numbers. Let s be a real number in the interior of D. The function f is differentiable at s with derivative $f'(s)$ (where $f'(s)$ is some real number) if and only if*

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

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

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

if and only if

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

But

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

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

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

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

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

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

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

where

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

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

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

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

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

8.2 Derivatives of Functions of Several Variables

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

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

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

The derivative $(D\varphi)_\mathbf{p}$ of 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 8.2 *Let $T: \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation from $\mathbb{R}^m$ into $\mathbb{R}^n$. Then T is differentiable at each point $\mathbf{p}$ of $\mathbb{R}^m$, and $(DT)_\mathbf{p} = T$.*

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

Lemma 8.3 *Let $\varphi: X \to \mathbb{R}^n$ be a function, let $T: \mathbb{R}^m \to \mathbb{R}^n$ be a linear transformation, and let $\Omega: X \to \mathbb{R}^n$ be defined so that*

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

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

Proof It follows from the definition of differentiability that the function φ is differentiable, with derivative $T: \mathbb{R}^m \to \mathbb{R}^n$, if and only if $\lim_{\mathbf{x} \to \mathbf{p}} \Omega(\mathbf{x}) = \mathbf{0}$. But $\Omega(\mathbf{p}) = \mathbf{0}$. It follows that $\lim_{\mathbf{x} \to \mathbf{p}} \Omega(\mathbf{x}) = \mathbf{0}$ if and only if the function Ω is continuous at $\mathbf{p}$ (see Proposition 4.20). The result follows. ∎

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

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

for all real numbers x and y. Let p and q be fixed real numbers. Then

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

Now

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

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

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

8.3 Differentiation of Square Matrices

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

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

We now review the relationship between the Hilbert-Schmidt norm $\|A\|_{\mathrm{HS}}$ and the operator norm $\|A\|_{\mathrm{op}}$ of an $n\times n$ matrix A. The operator norm $\|A\|_{\mathrm{op}}$ is by definition the smallest non-negative real number with the property that $|A\mathbf{x}| \le \|A\|_{\mathrm{op}}|\mathbf{x}|$ for all $\mathbf{x} \in \mathbb{R}^m$. Its properties are set out in the statement of Lemma 7.5.

Proposition 8.4 *The Hilbert-Schmidt norm $\|A\|_{\mathrm{HS}}$ and the operator norm $\|A\|_{\mathrm{op}}$ of an $n \times n$ matrix satisfy the inequalities*

$$\|A\|_{\mathrm{op}} \le \|A\|_{\mathrm{HS}} \le \sqrt{n}\|A\|_{\mathrm{op}}.$$

Proof The inequality $|A\mathbf{x}| \le \|A\|_{\mathrm{HS}}\,|\mathbf{x}|$ is satisfied for all $\mathbf{x} \in \mathbb{R}^n$ (see Lemma 7.6). It follows that $\|A\|_{\mathrm{op}} \le \|A\|_{\mathrm{HS}}$.

Now let $A_{i,j}$ denote the coefficient of the matrix A in the ith row and jth column, and let $\mathbf{e}_i$ denote the vector whose ith component is equal to i and whose other components are equal to zero. Then

$$A\mathbf{e}_j = \sum_{i=1}^{n} A_{i,j}\mathbf{e}_i$$

for $i = 1, 2, \ldots, n$. It follows from the orthogonality of $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ that

$$\sum_{i=1}^{n} A_{i,j}^2 \leq \|A\|_{\mathrm{op}}^2.$$

Adding these inequalities for $j = 1, 2, \ldots, n$, we find that

$$\|A\|_{\mathrm{HS}}^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{i,j}^2 \leq n\|A\|_{\mathrm{op}}^2,$$

and therefore $\|A\|_{\mathrm{HS}} \leq \sqrt{n}\|A\|_{\mathrm{op}}$. The result follows. ∎

Lemma 8.5 *Let V be a subset of the set $M_n(\mathbb{R})$. Then V is open in $M_n(\mathbb{R})$ if and only if, given any $n \times n$ matrix A belonging to V, there exists some positive number δ such that*

$$\{B \in M_n : \|A - B\|_{\mathrm{op}} < \delta\} \subset V.$$

Proof Suppose that V is open in $M_n(\mathbb{R})$. Let some positive real number ε be given. Now $M_n(\mathbb{R})$ is a Euclidean space whose Euclidean norm is the Hilbert-Schmidt norm. It follows that there exists some positive real number δ_{HS} for which

$$\{B \in M_n(\mathbb{R}) : \|A - B\|_{\mathrm{HS}} < \delta_{\mathrm{HS}}\} \subset V.$$

Let δ_{op} be chosen so that $0 < \sqrt{n}\,\delta_{\mathrm{op}} \leq \delta_{\mathrm{HS}}$. It follows from Proposition 8.4 that an $n \times n$ matrix B satisfies $\|A - B\|_{\mathrm{op}} < \delta_{\mathrm{op}}$ then

$$\|A - B\|_{\mathrm{HS}} < \sqrt{n}\|A - B\|_{\mathrm{op}} < \sqrt{n}\,\delta_{\mathrm{op}} \leq \delta_{\mathrm{HS}}$$

and therefore $B \in V$. Thus

$$\{B \in M_n(\mathbb{R}) : \|A - B\|_{\mathrm{op}} < \delta_{op}\} \subset V.$$

Conversely let V be a subset of $M_n(\mathbb{R})$ with the property that, given any $n \times n$ matrix A belonging to $M_n(\mathbb{R})$, there exists some positive number δ such that

$$\{B \in M_n(\mathbb{R}) : \|A - B\|_{\mathrm{op}} < \delta\} \subset V.$$

Let $B \in M_n(\mathbb{R})$ satisfy $|A - B|_{\mathrm{HS}} < \delta$. Then

$$\|A - B\|_{\mathrm{op}} \leq \|A - B\|_{\mathrm{HS}} < \delta,$$

and therefore $B \in V$. It follows that the set V is open in $M_n(\mathbb{R})$. The result follows. ∎

Definition The group $\mathrm{GL}(n, \mathbb{R})$ is the group of invertible $n \times n$ matrices with real coefficients, the group operation being the operation of multiplication of $n \times n$ matrices.

Lemma 8.6 *The group* $\mathrm{GL}(n, \mathbb{R})$ *is an open set in the Euclidean space* $M_n(\mathbb{R})$.

Proof The group $\mathrm{GL}(n, \mathbb{R})$ consists of those $n \times n$ matrices whose determinant is non-zero. Now the determinant $\det A$ of an $n \times n$ matrix A is a sum of products of coefficients of A. It follows that the determinant function is continuous on $M_n(\mathbb{R})$. It then follows from a direct application of Proposition 4.17 that

$$\{A \in M_n(\mathbb{R}) : \det A \neq 0\}$$

is an open set in $M_n(\mathbb{R})$, as required. ∎

The determinant $\det A$ of a square $n \times n$ matrix A is a continuous function of the coefficients of the matrix. It follows from this that $\mathrm{GL}(n, \mathbb{R})$ is an open subset of $M_n(\mathbb{R})$. We denote the identity $n \times n$ matrix by I. It follows directly from the definition of the operator norm that $\|I\|_{\mathrm{op}} = 1$.

Proposition 8.7 *Let* $\varphi\colon \mathrm{GL}(n, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R})$ *be the function defined so that* $\varphi(A) = A^{-1}$ *for all invertible* $n \times n$ *matrices* A. *Then the function* $\varphi\colon \mathrm{GL}(n, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R})$ *is differentiable, and*

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

Proof Let A be an invertible $n \times n$ matrix. Then, given any $n \times n$ matrix H, the matrix $I + A^{-1}H$ is invertible if and only if $\det(I + A^{-1}H) \neq 0$. Moreover this determinant is a continuous function of the coefficients of the matrix H. It follows that there exists some positive number δ_0 such that $I + A^{-1}H$ is invertible whenever $\|H\|_{\mathrm{op}} < \delta_0$. Now the coefficients of the matrix $(I + A^{-1}H)^{-1}$ are continuous functions of the coefficients of H on the open set consisting of those $n \times n$ matrices H for which $\det(I + A^{-1}H) \neq 0$. It follows that the function mapping the matrix H to $\|(I + A^{-1}H)^{-1}\|_{\mathrm{op}}$ is continuous (see Lemma 7.3). This function takes the value 1 when H is the zero matrix. We can therefore choose a positive number δ_0 small enough to ensure that $I + A^{-1}H$ is invertible and $\|(I + A^{-1}H)^{-1}\|_{\mathrm{op}} < 2$ whenever $\|H\|_{\mathrm{op}} < \delta_0$.

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

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

and therefore

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

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

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

It follows that

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

Now

$$\begin{aligned} &|A^{-1}HA^{-1}H(I + A^{-1}H)^{-1}A^{-1}\mathbf{x}| \\ &\qquad \leq \|A^{-1}\|_{\mathrm{op}}\|H\|_{\mathrm{op}}\|A^{-1}\|_{\mathrm{op}}\|H\|_{\mathrm{op}}\|(I + A^{-1}H)^{-1}\|_{\mathrm{op}}\|A^{-1}\|_{\mathrm{op}}|\mathbf{x}| \end{aligned}$$

for all $\mathbf{x} \in \mathbb{R}^n$, and therefore

$$\begin{aligned} &\|A^{-1}HA^{-1}H(I + A^{-1}H)^{-1}A^{-1}\|_{\mathrm{op}} \\ &\qquad \leq \|A^{-1}\|_{\mathrm{op}}^3\|(I + A^{-1}H)^{-1}\|_{\mathrm{op}}\|H\|_{\mathrm{op}}^2. \end{aligned}$$

Moreover $\|(I + A^{-1}H)^{-1}\|_{\mathrm{op}} < 2$ whenever $\|H\|_{\mathrm{op}} < \delta_0$, and therefore

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

whenever $\|H\|_{\mathrm{op}} < \delta_0$. It follows that

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

and thus

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

Proposition 8.4 ensures that the corresponding inequality with Hilbert-Schmidt norms in place of operator norms is also satisfied. Therefore the function $\varphi\colon \mathrm{GL}(n,\mathbb{R}) \to \mathrm{GL}(n,\mathbb{R})$ is differentiable, where $\varphi(A) = A^{-1}$ for all invertible $n \times n$ matrices A with real coefficients, and moreover

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

as required. ∎

8.4 Properties of Differentiable Functions of Several Real Variables

Lemma 8.8 *Let $\varphi: X \to \mathbb{R}^n$ be a function which maps an open subset X of $\mathbb{R}^m$ into $\mathbb{R}^n$ which is differentiable at some point $\mathbf{p}$ of X. Then φ is continuous at $\mathbf{p}$.*

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

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

for all points $\mathbf{x}$ of X satisfying $\mathbf{x} \neq \mathbf{p}$. If $\varphi: X \to \mathbb{R}^n$ is differentiable at $\mathbf{p}$ then $\Omega: X \to \mathbb{R}^n$ is continuous at $\mathbf{p}$ (see Lemma 8.3). Moreover

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

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

Lemma 8.9 *Let $\varphi: X \to \mathbb{R}^n$ be a function which maps an open subset X of $\mathbb{R}^m$ into $\mathbb{R}^n$ which is differentiable at some point $\mathbf{p}$ of X. Let $(D\varphi)_{\mathbf{p}}: \mathbb{R}^m \to \mathbb{R}^n$ be the derivative of φ at $\mathbf{p}$. Let $\mathbf{u}$ be an element of $\mathbb{R}^m$. Then*

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

Thus the derivative $(D\varphi)_{\mathbf{p}}$ of 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 $(\mathbf{x} - \mathbf{p}) = t\mathbf{u}$, and $(\mathbf{x} - \mathbf{p}) = -t\mathbf{u}$, where t is a real variable, we can conclude that

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

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

It follows that

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

as required. ∎

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

Proposition 8.10 *Let X be an open set in $\mathbb{R}^m$, and let $f\colon X \to \mathbb{R}$ and $g\colon X \to \mathbb{R}$ be functions mapping X into $\mathbb{R}$. Let $\mathbf{p}$ be a point of X. Suppose that f and g are differentiable at $\mathbf{p}$. Then the functions $f+g$ and $f-g$ are differentiable at $\mathbf{p}$, and*

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

and

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

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

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

Therefore

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

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

8.5 The Multidimensional Product Rule

Proposition 8.11 (Product Rule) *Let X be an open set in $\mathbb{R}^m$, and let $f\colon X \to \mathbb{R}$ and $g\colon X \to \mathbb{R}$ be functions mapping X into $\mathbb{R}$. Let $\mathbf{p}$ be a point of X. Suppose that f and g are differentiable at $\mathbf{p}$. Then the function $f \cdot g$ is differentiable at $\mathbf{p}$, and*

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

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

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

that are defined throughout X so as to ensure that

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

and

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

for all $\mathbf{x} \in X$ (see Lemma 8.3).

Then

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

where

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

Now

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

where $\|(Df)_{\mathbf{p}}\|_{\mathrm{op}}$ denotes the operator norm of $(Df)_{\mathbf{p}}$ (see Lemma 7.5) Similarly

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

It follows that

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

and therefore

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

Next we note that

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

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

Similarly

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

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

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

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

Putting these results together, we see that

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

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

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

(see Lemma 8.3). This completes the proof. ∎

8.6 The Multidimensional Chain Rule

Proposition 8.12 (Chain Rule) *Let X be an open set in $\mathbb{R}^m$, and let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$. Let Y be an open set in $\mathbb{R}^n$ which contains $\varphi(X)$, and let $\psi\colon Y \to \mathbb{R}^k$ be a function mapping Y into $\mathbb{R}^k$. Let $\mathbf{p}$ be a point of X. Suppose that φ is differentiable at $\mathbf{p}$ and that ψ is differentiable at $\varphi(\mathbf{p})$. Then the composition $\psi\circ\varphi\colon \mathbb{R}^m \to \mathbb{R}^k$ (i.e., φ followed by ψ) is differentiable at $\mathbf{p}$. Moreover*

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

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

Proof Let $\mathbf{q} = \varphi(\mathbf{p})$. The functions $\varphi\colon X \to \mathbb{R}^n$ and $\psi\colon Y \to \mathbb{R}^k$ are differentiable at $\mathbf{p}$ and $\mathbf{q}$ respectively, and therefore there are well-defined functions

$\Omega_1 \colon X \to \mathbb{R}^n$ and $\Omega_2 \colon Y \to \mathbb{R}^k$ that are defined throughout X and Y respectively so as to ensure that

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

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

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

and

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

for all $\mathbf{y} \in Y$ (see Lemma 8.3).

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

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

where

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

Let

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

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

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

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

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

where $\|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$ denotes the operator norm of the linear operator $(D\varphi)_{\mathbf{p}}$ (see Lemma 7.5). It follows that

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

for all $\mathbf{x} \in X$ satisfying $\mathbf{x} \neq \mathbf{p}$. It follows from the continuity of the function Ω_1 at $\mathbf{p}$ that $M(\mathbf{x})$ remains bounded as $\mathbf{x}$ tends to $\mathbf{p}$ in X. Now

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

Also the function $\varphi: X \to \mathbb{R}^n$ is continuous at $\mathbf{p}$ and the function $\Omega_2: Y \to \mathbb{R}^k$ is continuous at $\varphi(\mathbf{p})$. It follows that the composition function $\Omega_2 \circ \varphi$ is continuous at $\mathbf{p}$ (see Lemma 4.4), and therefore

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

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

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

(see Proposition 4.22).

Linear operators on finite-dimensional vector spaces are continuous. Therefore

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

It follows that

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

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

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

(see Lemma 8.3). The result follows. ∎

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

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

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

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

is differentiable everywhere on $\mathbb{R}$, though its derivative $h': \mathbb{R} \to \mathbb{R}$ is not continuous at 0. Also the functions $(x, y) \mapsto x$ and $(x, y) \mapsto y$ are differentiable everywhere on $\mathbb{R}$ (by Lemma 8.2). Now $\varphi(x, y) = y^3 h(x)$. Using Proposition 8.10 and Proposition 8.12, we conclude that φ is differentiable everywhere on $\mathbb{R}^2$.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

cannot possibly exist. Because f is not continuous at $(0,0)$ we conclude from Lemma 8.8 that f cannot be differentiable at $(0,0)$. However it is easy to show that the partial derivatives

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

8.7 Partial Derivatives and Continuous Differentiability

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

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

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

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

for all $(w_1, w_2, \ldots, w_n) \in \mathbb{R}^n$. Then the partial derivative of the ith component f_i of the function φ with respect to the jth coordinate function x_j at a point $\mathbf{p}$ of X is determined by the formula

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

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

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

Lemma 8.13 *Let X be an open set in $\mathbb{R}^m$. and let $\varphi\colon X \to \mathbb{R}^n$ be a continuously differentiable function on X. Then the first order partial derivatives of the components of φ exist and are continuous throughout X.*

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

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

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

8.8 Functions with Continuous Partial Derivatives

Proposition 8.14 *Let X be an open set in $\mathbb{R}^m$, let $\varphi\colon X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$, and let $\mathbf{p}$ be a point of X, where $\mathbf{p} = (p_1, p_2, \ldots, p_m)$. Suppose that the partial derivatives of the components of φ with respect to the Cartesian coordinates exist and are continuous throughout X. Suppose also that the partial derivatives of the components of φ are all equal to zero at the point $\mathbf{p}$. Then, given any positive real number ε, there exists a positive real number δ such that $\mathbf{u} \in X$, $\mathbf{v} \in X$ and*

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

for all points $\mathbf{u}$ and $\mathbf{v}$ of $H(\mathbf{p}, \delta)$, where

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

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

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

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

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

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

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

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

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

for $j = 1, 2, \ldots, n$. Then $\mathbf{q}_m = \mathbf{u}$ and $\mathbf{q}_j \in H(\mathbf{p}, \delta)$ for $j = 1, 2, \ldots, m$. Now, for each integer j between 1 and m, the points $\mathbf{q}_j$ and $\mathbf{q}_{j-1}$ differ only in the jth coordinate. Applying the Mean Value Theorem of single-variable calculus (Theorem 2.2), we find that, given any pair of integers i and j, where $1 \leq i \leq n$ and $1 \leq j \leq m$, there exists some real number θ satisfying $0 < \theta < 1$ such that

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

It follows that

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

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

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

On applying Schwarz's Inequality (Lemma 4.1), we find that

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

It follows that

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

and therefore

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

It follows that

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

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

Remark The essential strategy underlying the proof of Proposition 8.14 can be presented, in the two-dimensional case, as follows. Consider a city laid out on a gridiron pattern, where all streets run either from north to south, or from east to west. To get from one street intersection to another, it is always possible to find a route that does not involve both northward and southward legs, and does not involve both eastward and westward legs. (Thus to get from one street intersection to another that lies to the northeast, one can choose a route that involves only travelling northwards or travelling eastwards along city streets.) Suppose that all streets have a maximum gradient equal to m. Then the height difference between any two intersections is bounded above by $\sqrt{2}md$, where d is the direct distance between those street intersections.

Corollary 8.15 *Let $\varphi: X \to \mathbb{R}^n$ be a continuously differentiable function defined over an open set X in $\mathbb{R}^m$, and let $\mathbf{p}$ be a point of X. Let M be a positive real number satisfying $M > \|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$, where $\|(D\varphi)_{\mathbf{p}}\|_{\mathrm{op}}$ denotes the operator norm of the derivative of φ at $\mathbf{p}$. Then there exists a positive real number δ such that*

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

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

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

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

for all $\mathbf{u} \in X$. Then $(D\psi)_{\mathbf{p}} = (D\varphi)_{\mathbf{p}} - (D\varphi)_{\mathbf{p}} = 0$. It follows from Proposition 8.14 that there exists a positive real number δ such that

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

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

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

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

Corollary 8.15 ensures that continuously differentiable functions of several real variables are *locally Lipschitz continuous.* This means that they satisfy a Lipschitz condition in some sufficiently small neighbourhood of any given point. This in turn ensures that standard theorems concerning the existence and uniqueness of ordinary differential equations can be applied to systems of ordinary differential equations specified in terms of continuously differentiable functions.

Theorem 8.16 *Let X be an open subset of $\mathbb{R}^m$ and let $\varphi: X \to \mathbb{R}^n$ be a function mapping X into $\mathbb{R}^n$. Suppose that the Jacobian matrix*

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

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

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

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

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

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

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

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

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

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

Let some positive real number ε be given. It follows from Proposition 8.14 that there exists some positive real number δ such that

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

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

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

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

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

at the point $\mathbf{p}$).

It follows from this that

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

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

Corollary 8.17 *Let X be an open set in $\mathbb{R}^m$. A function $\varphi: X \to \mathbb{R}^n$ is continuously differentiable if and only if the first order partial derivatives of the components of φ exist and are continuous throughout X.*

Proof The result follows directly on combining the results of Lemma 8.13 and Theorem 8.16. ∎

8.9 Summary of Differentiability Results

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

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

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

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

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

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

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

(v) Linear transformations are everywhere differentiable.

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

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

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