Course 223: Academic Year 1987–8

D. R. Wilkins

Contents

1 Summary of Some of the Basic Concepts of Analysis

1.1 The Real Number System

One of the first tasks in an introduction to analysis is to construct the system of *real numbers*. The best-known method of constructing the real number system is by means of *cuts*. It is due to Dedekind. We denote the set of all real numbers by $\mathbb{R}$.

Having constructed the real number system one must then investigate the basic properties of this system. These fall into three categories: namely (I) *algebraic properties*, (II) *ordering properties*, and (III) *the 'no gaps' property*.

(I) *Algebraic Properties.*

The set $\mathbb{R}$ of real numbers is a *field* under the operations of addition and multiplication. This means that the set $\mathbb{R}$ of all real numbers is an *Abelian group* under the operation of addition, the set $\mathbb{R} \setminus \{0\}$ of non-zero real numbers is also an *Abelian group* under the operation of multiplication, and the operations of addition and multiplication satisfy the *distributive laws*. Expressed in less formal terms, this simply means that there are operations of addition, subtraction, multiplication and division defined on the set $\mathbb{R}$ of real numbers, and these operations satisfy all the familiar rules (such as the commutative laws, the associative laws, the distributive laws etc.).

(II) *Ordering Properties.*

The set of real numbers is ordered. If x and y are real numbers and if x is strictly less than y then we write $x < y$. If $x < y$ then we can also denote this fact by writing $y > x$. We write $x \leq y$ if either $x < y$ or $x = y$. Similarly we write $x \geq y$ if either $x > y$ or $x = y$. This ordering relation has the following properties:

(i) if x and y are real numbers then one and only one of the statements

$$x < y, \qquad x = y, \qquad y < x$$

is true,

(ii) if x, y and z are real numbers, if $x < y$ and $y < z$ then $x < z$,

(iii) if x, y and z are real numbers and if $y < z$ then $x + y < x + z$,

(iv) if x, y and z are real numbers, if $x > 0$ and $y > 0$ then $xy > 0$.

(III) *'No Gaps' Property.*

In addition to the algebraic properties and the ordering properties discussed above, the real number system is characterized by the property that there are no 'gaps' or 'holes' in the set of real numbers. It is necessary to express this vague statement formally in language that is logically precise. The property of the real numbers that captures this notion of 'no gaps' is the *least upper bound principle* (or *least upper bound axiom*), which we now describe.

A subset S of the set $\mathbb{R}$ of real numbers is said to be *bounded above* if there exists a real number u such that $s \leq u$ for all $s \in S$. A real number u with this property is said to be an *upper bound* for the set S. A *least upper bound* (or *supremum*) for the set S is a real number l with the properties that l is an upper bound for the set S which is less than any other upper bound for the set S. Clearly any subset of the set of real numbers can have at most one least upper bound.

The *least upper bound principle* states that if S is a subset of the set $\mathbb{R}$ of real numbers which is bounded above, then the set S has a (unique) least upper bound (or *supremum*), denoted by $\sup S$.

One easily deduces from the least upper bound principle the corresponding result that if S is a subset of $\mathbb{R}$ which is bounded below then the set S has a (unique) *greatest lower bound* (or *infimum*), denoted by $\inf S$, which has the properties that $\inf S \leq s$ for all $s \in S$ and if v is a lower bound for the set S (i.e., $v \leq s$ for all $s \in S$) then $v \leq \inf S$. Indeed the greatest lower bound of the set S is given by the formula

$$\inf S = -\sup\{x \in \mathbb{R} : -x \in S\}.$$

We have now described the properties which characterize the set $\mathbb{R}$ of real numbers. It turns out that these properties (the algebraic properties, the ordering properties and the least upper bound principle) are sufficient to characterize the real number system uniquely. Thus it is not necessary to use the explicit definition of real numbers (in terms of cuts) in order to prove theorems about real numbers: we can deduce such theorems from the properties listed above. Of course the algebraic properties (i.e. the properties of addition, subtraction, multiplication and division) and the ordering properties should be very familiar! However, to prove many important theorems of analysis we need the 'no gaps' property that is expressed formally by the least upper bound principle.

When one studies number systems, one meets a famous theorem of classical Greek mathematics (given in Euclid) which states that there is no rational number x satisfying the equation $x^2 = 2$ (i.e. there do not exist integers p and q with the property that $p^2 = 2q^2$). One is also told that numbers like π and e are not rational numbers. Thus one learns that there is a system of *real numbers* which contains all the rational numbers together with numbers such as $\sqrt{2}$, $\sqrt{3}$, π and e. However if one wants to be able to prove theorems involving the real number system then it is not much use to know merely that "the set of real numbers consists of the set of rational numbers together with certain other numbers such as $\sqrt{2}$, $\sqrt{3}$, π and e". One needs a far more precise characterization of the real numbers in order to be able to prove significant theorems involving real numbers. This is where the least upper bound principle comes in. It is the key to a number of deep theorems concerning real numbers and continuous functions.

1.2 Euclidean Space

We define *n-dimensional Euclidean space* $\mathbb{R}^n$ to be the space which consists of all n-tuples $(x_1, x_2, \ldots, x_n)$ of real numbers. Thus $\mathbb{R}^1$ (denoted also by $\mathbb{R}$) is the real line, $\mathbb{R}^2$ is the (Euclidean) plane (i.e. the familiar plane in which are located all the triangles, parallel lines etc. of good old-fashioned schoolroom geometry) and $\mathbb{R}^3$ represents standard Euclidean space. The functions that one considers are often defined over some subset of $\mathbb{R}^n$ known as the *domain* of that function. Thus, for example, the function $x \mapsto 1/x$ is defined over the set

$$\{x \in \mathbb{R} : x \neq 0\}$$

consisting of all non-zero real numbers. If f is a real-valued function whose domain is a subset D of $\mathbb{R}^n$ for some n, then we write $f\colon D \to \mathbb{R}$ in order to denote the fact that the function f is defined on the set D (i.e., $f(x)$ is defined for all $x \in D$) and f maps the elements of D into the set $\mathbb{R}$ of real numbers.

We add and subtract elements of $\mathbb{R}^n$ and multiply them by scalars (i.e. real numbers) in the obvious fashion. Thus if we are given elements $\mathbf{x}$ and $\mathbf{y}$ of $\mathbb{R}^n$ represented by the n-tuples $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ then we define $\mathbf{x} + \mathbf{y}$ and $\mathbf{x} - \mathbf{y}$ to be the elements of $\mathbb{R}^n$ represented by the n-tuples $(x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$ and $(x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n)$ respectively. Similarly we denote by $\lambda\mathbf{x}$ the element of $\mathbb{R}^n$ represented by the n-tuple $(\lambda x_1, \lambda x_2, \ldots, \lambda x_n)$, for all real numbers λ.

There is an *inner product* (or *scalar product*) defined on n-dimensional Euclidean space: if $\mathbf{x}$ and $\mathbf{y}$ are elements of $\mathbb{R}^n$, represented by the n-tuples $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$ respectively then the inner product $\mathbf{x}.\mathbf{y}$

of $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$\mathbf{x}.\mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Note that $\mathbf{x}.\mathbf{y} = \mathbf{y}.\mathbf{x}$ and $(\lambda\mathbf{x}).\mathbf{y} = \lambda(\mathbf{x}.\mathbf{y}) = \mathbf{x}.(\lambda\mathbf{y})$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$, and that $(\mathbf{x}+\mathbf{y}).\mathbf{z} = \mathbf{x}.\mathbf{z}+\mathbf{y}.\mathbf{z}$ for all $\mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$. The inner product satisfies an important inequality known as *Schwarz' Inequality*.

Lemma 1.1 (Schwarz' Inequality) *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be elements of* $\mathbb{R}^n$. *Then* $|\mathbf{x}.\mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|$.

Proof Let $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, and let us consider the quantity $A(\lambda, \mu)$ defined by

$$A = \sum_{j=1}^{n} (\lambda x_j - \mu y_j)^2.$$

Note that $A(\lambda, \mu)$ is a sum of squares, so that $A(\lambda, \mu) \geq 0$. But if we expand out the expression defining $A(\lambda, \mu)$ we see that

$$\begin{aligned} A(\lambda, \mu) &= \sum_{j=1}^{n} \left(\lambda^2 x_j^2 - 2\lambda\mu x_j y_j + \mu^2 y_j^2\right) \\ &= \lambda^2 |\mathbf{x}|^2 - 2\lambda\mu \mathbf{x}.\mathbf{y} + \mu^2 |\mathbf{y}|^2, \end{aligned}$$

hence

$$\lambda^2 |\mathbf{x}|^2 - 2\lambda\mu \mathbf{x}.\mathbf{y} + \mu^2 |\mathbf{y}|^2 \geq 0$$

for all $\lambda, \mu \in \mathbb{R}$. In particular, if we take $\lambda = |\mathbf{y}|^2$ and $\mu = \mathbf{x}.\mathbf{y}$ we see that

$$|\mathbf{y}|^4 |\mathbf{x}|^2 - 2|\mathbf{y}|^2 (\mathbf{x}.\mathbf{y})^2 + (\mathbf{x}.\mathbf{y})^2 |\mathbf{y}|^2 \geq 0,$$

which simplifies to

$$|\mathbf{y}|^2 \left(|\mathbf{x}|^2 |\mathbf{y}|^2 - \mathbf{x}.\mathbf{y}^2\right) \geq 0.$$

It follows from this that $|\mathbf{x}.\mathbf{y}| \leq |\mathbf{x}||\mathbf{y}|$ when $\mathbf{y} \neq \mathbf{0}$. But the inequality is trivially satisfied when $\mathbf{y} = \mathbf{0}$. This proves Schwarz' inequality. ∎

The *length* (or *norm*) of an element $\mathbf{x}$ of $\mathbb{R}^n$ is denoted by $|\mathbf{x}|$, and is defined by $|\mathbf{x}| = \sqrt{\mathbf{x}.\mathbf{x}}$. Thus if $\mathbf{x}$ is represented by the n-tuple $(x_1, x_2, \ldots, x_n)$ then

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

If $\mathbf{x}$ and $\mathbf{y}$ are elements of $\mathbb{R}^n$ then the quantity $|\mathbf{x}-\mathbf{y}|$ measures the *distance* from the point $\mathbf{x}$ to the point $\mathbf{y}$. (Note that this notion of distance corresponds to the familiar notion of distance on the Euclidean plane $\mathbb{R}^2$ and on standard 3-dimensional Euclidean space $\mathbb{R}^3$, by Pythagoras' Theorem!). We can deduce from Schwarz' inequality the following inequality.

Lemma 1.2 *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *be elements of* $\mathbb{R}^n$. *Then*

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|.$$

Proof We expand out the quantity $|\mathbf{x} + \mathbf{y}|^2$, using the definition of this quantity as the inner product of $\mathbf{x} + \mathbf{y}$ with itself. Thus

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &= (\mathbf{x} + \mathbf{y}).(\mathbf{x} + \mathbf{y}) \\ &= \mathbf{x}.\mathbf{x} + \mathbf{y}.\mathbf{y} + 2\mathbf{x}.\mathbf{y} \\ &= |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2\mathbf{x}.\mathbf{y} \\ &\leq |\mathbf{x}|^2 + |\mathbf{y}|^2 + 2|\mathbf{x}||\mathbf{y}| \\ &= (|\mathbf{x}| + |\mathbf{y}|)^2 \end{aligned}$$

(where we have used Schwarz' inequality). Hence $|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$, as required. ∎

In particular, if $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{z}$ are points in $\mathbb{R}^n$ then

$$|\mathbf{x} - \mathbf{y}| \leq |\mathbf{x} - \mathbf{z}| + |\mathbf{z} - \mathbf{y}|$$

This important inequality is known as the *triangle inequality.* It describes the fact that, for any triangle in $\mathbb{R}^n$, the length of any side of that triangle is less than the sum of the lengths of the other two sides.

1.3 Open and Closed Sets

There is a particularly important class of subsets of n-dimensional Euclidean space $\mathbb{R}^n$. These sets are known as *open sets.* A subset D of $\mathbb{R}^n$ is said to be *open* if and only if, given any element $\mathbf{a}$ of D there exists some $\delta > 0$ (which might depend on $\mathbf{a}$) such that all elements of $\mathbb{R}^n$ which satisfy $|\mathbf{x} - \mathbf{a}| < \delta$ belong to D (i.e., D contains all points whose distance from $\mathbf{a}$ is less than δ). In more informal language, all that this definition is saying is that a subset D of $\mathbb{R}^n$ is open if and only if, for each point of D, the immediate neighbourhood of that point is contained in D.

Example Let c and d be real numbers satisfying $c < d$. The *open interval* (c, d) is defined to be the subset of $\mathbb{R}$ consisting of all real numbers t which satisfy the inequality $c < t < d$. We claim that the set (c, d) is an open subset of $\mathbb{R}$. For if t is a real number satisfying $c < t < d$ then we can choose $\delta > 0$ such that $t - \delta > c$ and $t + \delta < d$. Then all numbers x satisfying $|x - t| < \delta$ are contained in the set (c, d).

Let $\mathbf{a}$ be a point of n-dimensional Euclidean space $\mathbb{R}^n$. Given a strictly positive real number δ, let $B(\mathbf{a}, \delta)$ denote the set

$$\{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{a}| < \delta\}$$

consisting of all points $\mathbf{x}$ in $\mathbb{R}^n$ whose distance from the point $\mathbf{a}$ is less than δ. We claim that the set $B(\mathbf{a}, \delta)$ is an open set. In order to verify this, we must check that if $\mathbf{x}$ is any point belonging to $B(\mathbf{a}, \delta)$ then there exists some $\delta_1 > 0$ such that all points $\mathbf{y}$ of $\mathbb{R}^n$ which satisfy $|\mathbf{y} - \mathbf{x}| < \delta_1$ belong to $B(\mathbf{a}, \delta)$ (i.e., we must show that, given any point $\mathbf{x}$ of $B(\mathbf{a}, \delta)$, there exists some $\delta_1 > 0$ such that $B(\mathbf{x}, \delta_1) \subset B(\mathbf{a}, \delta)$). But this follows if we choose $\delta_1 = \delta - |\mathbf{x} - \mathbf{a}|$, for then $\delta_1 > 0$ (recall that $\mathbf{x}$ belongs to $B(\mathbf{a}, \delta)$ if and only if $|\mathbf{x} - \mathbf{a}| < \delta$) and if $|\mathbf{y} - \mathbf{x}| < \delta_1$ then

$$\begin{aligned} |\mathbf{y} - \mathbf{a}| &\leq |\mathbf{y} - \mathbf{x}| + |\mathbf{x} - \mathbf{a}| \\ &< \delta_1 + |\mathbf{x} - \mathbf{a}| = \delta, \end{aligned}$$

by the triangle inequality. The subset $B(\mathbf{a}, \delta)$ of $\mathbb{R}^n$ defined above is referred to as the *open ball with radius* δ about $\mathbf{a}$, for obvious geometric reasons. (In the case when $n = 2$, one also refers to this set as the *open disk with radius* δ about $\mathbf{a}$, again for obvious geometric reasons.)

The definition of an open set can be rephrased as follows: a subset D of $\mathbb{R}^n$ is *open* if and only if, for every point $\mathbf{a}$ of D, there exists some $\delta > 0$ (which will usually depend on the choice of the point $\mathbf{a}$) such that the open ball $B(\mathbf{a}, \delta)$ of radius δ about $\mathbf{a}$ is contained in D.

Now let X be some subset of n-dimensional Euclidean space $\mathbb{R}^n$, and let $\mathbf{a}$ be a point of X. Given $\delta > 0$ we define the subset $B_X(\mathbf{a}, \delta)$ of X by

$$B_X(\mathbf{a}, \delta) = \{\mathbf{x} \in X : |\mathbf{x} - \mathbf{a}| < \delta\}.$$

More informally, $B_X(\mathbf{a}, \delta)$ is the set consisting of all points $\mathbf{x}$ belonging to the set X whose distance from $\mathbf{a}$ is less than δ. We refer to $B_X(\mathbf{a}, \delta)$ as the *open ball in X of radius δ about* $\mathbf{a}$.

We say that a subset U of X is *open in* X if and only if, given any point $\mathbf{a}$ of U there exists some $\delta > 0$ such that the open ball $B_X(\mathbf{a}, \delta)$ in X of radius δ is contained in X. To express this another way: a subset U of X is *open in* X if and only if for all points $\mathbf{a}$ of U there exists some $\delta > 0$ (which may depend on $\mathbf{a}$) such that all points of X whose distance from $\mathbf{a}$ is less than δ are contained in U.

We now prove that the union of any collection of open subsets of n-dimensional Euclidean space is itself open. We also show that any *finite* intersection of open subsets of this Euclidean space is itself open.

Lemma 1.3 *Let* $(U_\alpha : \alpha \in A)$ *be a collection of open sets in* n*-dimensional Euclidean space* $\mathbb{R}^n$*. Then the union* $\bigcup_{\alpha \in A} U_\alpha$ *of all the open sets of this collection is itself an open set. Also let* $U_1, U_2, \ldots, U_k$ *be a finite collection of open sets. Then the intersection*

$$U_1 \cap U_2 \cap \cdots \cap U_k$$

of this finite collection of open sets is itself an open set.

Proof We show that the union of all the open sets in the collection $(U_\alpha : \alpha \in A)$ is itself an open set. Let $\mathbf{a}$ be a point which is an element of this union. Then $\mathbf{a}$ is an element of U_α for one of the open sets U_α of this collection. Therefore there exists some $\delta > 0$ such that if $\mathbf{x}$ is a point of $\mathbb{R}^n$ which satisfies $|\mathbf{x} - \mathbf{a}| < \delta$ (i.e. if the point $\mathbf{x}$ is within a distance δ of $\mathbf{a}$) then $\mathbf{x}$ is a member of U_α. But of course this means that all points of $\mathbb{R}^n$ that are within a distance δ of $\mathbf{a}$ belong to the union $\bigcup_{\alpha \in A} U_\alpha$ of all the open sets of the collection. Therefore this union is itself an open set.

Next we show that the intersection $U_1 \cap U_2 \cap \cdots \cap U_k$ of the open sets U_1, $U_2, \ldots, U_k$ is itself an open set. Let $\mathbf{a}$ be a point of this intersection. Then there exist strictly positive real numbers δ_1, $\delta_2, \ldots, \delta_k$ such that

$$\{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{a}| < \delta_i\} \subset U_i$$

for $i = 1, 2, \ldots, k$. Let δ be the minimum of $\delta_1, \delta_2, \ldots, \delta_k$. Then $\delta > 0$, and if $\mathbf{x}$ is a point of $\mathbb{R}^n$ which satisfies $|\mathbf{x} - \mathbf{a}| < \delta$ then $\mathbf{x}$ is an element of U_i for $i = 1, 2, \ldots, k$ and thus

$$\mathbf{x} \in U_1 \cap U_2 \cap \cdots \cap U_k.$$

Thus given any point $\mathbf{a}$ of $U_1 \cap U_2 \cap \cdots \cap U_k$ there exists a ball (whose radius is strictly positive) about the point $\mathbf{a}$ which is contained in the intersection $U_1 \cap U_2 \cap \cdots \cap U_k$. Thus this intersection of open sets is itself open. ∎

Let $\mathbf{x}$ be a point in $\mathbb{R}^n$. A subset N of $\mathbb{R}^n$ is said to be a *neighbourhood* of the point $\mathbf{x}$ if there exists an open set U such that $\mathbf{x} \in U$ and $U \subset N$. To express this another way, a subset N of $\mathbb{R}^n$ is a neighbourhood of the point $\mathbf{x}$ if and only if there exists some $\delta > 0$ such that the open ball $B(\mathbf{x}, \delta)$ of radius δ about $\mathbf{x}$ is contained in N. (For if there exists some $\delta > 0$ with this property then $B(\mathbf{x}, \delta)$ is an open set which contains $\mathbf{x}$ and which is a subset of N; conversely if U is an open set in $\mathbb{R}^n$ such that $\mathbf{x} \in U$ and $U \subset N$ then the fact that U is open implies that there exists some $\delta > 0$ such that

$$B(\mathbf{x}, \delta) \subset U \subset N.)$$

This gives us another characterization of open sets: a subset U of $\mathbb{R}^n$ is open if and only if U is a neighbourhood of every point that belongs to U.

A subset F of n-dimensional Euclidean space is said to be *closed* if the complement $\mathbb{R}^n \setminus F$ of the set F in $\mathbb{R}^n$ is open (where

$$\mathbb{R}^n \setminus F \equiv \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} \text{ is not an element of } F\}).$$

The *closure* $\overline{D}$ of a subset D of $\mathbb{R}^n$ is defined to be the smallest closed set in $\mathbb{R}^n$ which contains D. More formally, the closure $\overline{D}$ is characterized by the properties that $\overline{D}$ is a closed set containing D and if F is any other closed set containing D then $\overline{D} \subset F$.

We now show that a point $\mathbf{x}$ of $\mathbb{R}^n$ belongs to the closure $\overline{D}$ of the set D if and only if the intersection of the set D with the open ball $B(\mathbf{x}, \delta)$ of radius δ about $\mathbf{x}$ is non-empty for all strictly positive real numbers δ.

Lemma 1.4 *Let D be a subset of n-dimensional Euclidean space $\mathbb{R}^n$ and let $\overline{D}$ be the closure of D. Then a point $\mathbf{x}$ of $\mathbb{R}^n$ belongs to $\overline{D}$ if and only if for every $\delta > 0$ there exists a point $\mathbf{y}$ of D with the property that $|\mathbf{y} - \mathbf{x}| < \delta$.*

Proof Let E be the subset of $\mathbb{R}^n$ consisting of all points $\mathbf{x}$ of $\mathbb{R}^n$ which have the property that the intersection of the set D with the open ball $B(\mathbf{x}, \delta)$ of radius δ about $\mathbf{x}$ is non-empty for all strictly positive real numbers δ. We must show that $E = \overline{D}$.

First we show that E is closed. To prove this, we must show that the complement $\mathbb{R}^n \setminus E$ of E in $\mathbb{R}^n$ is open. But it follows from the definition of E that if a point $\mathbf{x}$ of $\mathbb{R}^n$ belongs to $\mathbb{R}^n \setminus E$ then there exists some $\delta_1 > 0$ with the property that $B(\mathbf{x}, \delta_1) \cap D = \emptyset$ (where $\emptyset$ is the empty set). Let δ be defined by $\delta = \frac{1}{2}\delta_1$. If $\mathbf{y}$ is an element of the open ball $B(\mathbf{x}, \delta)$ of radius δ about the point $\mathbf{x}$ then it follows from the triangle inequality that $B(\mathbf{y}, \delta) \subset B(\mathbf{x}, \delta_1)$ (since $\delta_1 = 2\delta$). But δ_1 was chosen such that $B(\mathbf{x}, \delta_1) \cap D = \emptyset$. It follows that if $\mathbf{y}$ belongs to $B(\mathbf{x}, \delta)$ then $B(\mathbf{y}, \delta) \cap D = \emptyset$, and so $\mathbf{y}$ belongs to the complement $\mathbb{R}^n \setminus E$ of E (by the definition of the set E). We have therefore shown that if $\mathbf{x}$ is an element of $\mathbb{R}^n \setminus E$ then there exists some strictly positive real number δ such that the open ball $B(\mathbf{x}, \delta)$ of radius δ is contained in $\mathbb{R}^n \setminus E$. Thus $\mathbb{R}^n \setminus E$ is open, and so E is closed.

Now let F be a closed set which contains D. We show that the complement $\mathbb{R}^n \setminus F$ of F in $\mathbb{R}^n$ is contained in $\mathbb{R}^n \setminus E$. But $\mathbb{R}^n \setminus F$ is open, hence if $\mathbf{x}$ is an element of $\mathbb{R}^n \setminus F$ then there exists some $\delta > 0$ such that

$$B(\mathbf{x}, \delta) \subset \mathbb{R}^n \setminus F.$$

But then $B(\mathbf{x}, \delta) \cap D = \emptyset$ (because $D \subset F$), and thus $\mathbf{x}$ belongs to $\mathbb{R}^n \setminus E$. This shows that the complement $\mathbb{R}^n \setminus F$ of F in $\mathbb{R}^n$ is contained in $\mathbb{R}^n \setminus E$, so that E is a subset of F.

We have therefore shown that E is a closed subset of $\mathbb{R}^n$ containing D which is contained in every other closed set F that contains D. Therefore the set E is the closure of D in $\mathbb{R}^n$, which is what we set out to prove. ∎

Example The closure of the subset $(0, 1)$ of $\mathbb{R}$ is the set $[0, 1]$ (where $(0, 1)$ is the set of all real numbers t satisfying $0 < t < 1$ and $[0, 1]$ is the set of all real numbers satisfying $0 \leq t \leq 1$).

Example Let D be the open unit ball in $\mathbb{R}^3$, consisting of those points (x_1, x_2, x_3) in $\mathbb{R}^3$ which satisfy

$$x_1^2 + x_2^2 + x_3^2 < 1.$$

Then the closure $\overline{D}$ consists of all points (x_1, x_2, x_3) of $\mathbb{R}^3$ which satisfy

$$x_1^2 + x_2^2 + x_3^2 \leq 1.$$

1.4 Continuity

A real-valued function $f\colon D \to \mathbb{R}$ defined on a subset D of $\mathbb{R}^n$ is said to be *continuous* at an element $\mathbf{a}$ of D if and only if, for all $\varepsilon > 0$ there exists some $\delta > 0$ (which may depend on $\mathbf{a}$) such that

$$|f(\mathbf{x}) - f(\mathbf{a})| < \varepsilon$$

whenever $\mathbf{x}$ belongs to D and

$$|\mathbf{x} - \mathbf{a}| < \delta.$$

If f is a real-valued function defined on a subset D of some Euclidean space, and if f is continuous at every point of D then we say that f is *continuous on D*.

There is no excuse for not being able to quote this definition of continuity. Thus if you are not at present able to reproduce this definition (e.g., on the back of an envelope) without looking it up in your lecture notes, then you should learn it NOW! You should also have some idea as to what the definition of continuity means in practical terms. For example, let us consider a real-valued function $f\colon \mathbb{R}^3 \to \mathbb{R}$ defined on 3-dimensional Euclidean space $\mathbb{R}^3$. Let $\mathbf{a}$ be a point of $\mathbb{R}^3$. We suppose that some 'margin of error' is specified, measured by some strictly positive real number ε. To say that the function f is *continuous* at $\mathbf{a}$ is to say that, however small the margin of error that we allow ourselves (i.e., however close ε is to zero), we can always find some ball about the point $\mathbf{a}$ so that the value of f at any point of this ball agrees with the value of f at the point $\mathbf{a}$ to within the specified 'margin of error' ε (i.e., we can find a ball $B(\mathbf{a}, \delta)$ about $\mathbf{a}$ such that

$$f(\mathbf{a}) - \varepsilon < f(\mathbf{x}) < f(\mathbf{a}) + \varepsilon$$

at all points $\mathbf{x}$ within this ball).

We can rephrase the definition of continuity in terms of 'neighbourhoods'. Recall that if N is a subset of $\mathbb{R}^n$ containing the point $\mathbf{a}$ then N is said to be a *neighbourhood* of the point $\mathbf{a}$ if and only if there exists some $\delta > 0$ such that the ball $B(\mathbf{a}, \delta)$ of radius δ about $\mathbf{a}$ is contained in N. Thus a set N is a neighbourhood of the point $\mathbf{a}$ if it contains all points of $\mathbb{R}^n$ that are 'sufficiently close' to the point $\mathbf{a}$. One can readily verify that a function f defined on some subset D of some Euclidean space is continuous at a point $\mathbf{a}$ of D if and only if, for all $\varepsilon > 0$, there exists a neighbourhood of $\mathbf{a}$ in D such that the value of f at any point of this neighbourhood agrees with the value of f at the point $\mathbf{a}$ to within an error of ε.

An important result states that compositions of continuous functions are themselves continuous. Let $f\colon D \to \mathbb{R}$ be a real-valued function defined on a subset D of $\mathbb{R}^n$ and let $g\colon E \to \mathbb{R}$ be a real-valued function defined on a subset E of $\mathbb{R}$. If $f(D) \subset E$ (i.e., if the image of the set D under the map f is contained in the domain E of g, so that $g(f(\mathbf{x}))$ is well-defined for all $\mathbf{x} \in D$) then we can form the *composition* $g \circ f$ of f and g. This is the function defined by $(g \circ f)(\mathbf{x}) = g(f(\mathbf{x}))$ for all $x \in D$.

Lemma 1.5 *Let $f\colon D \to \mathbb{R}$ be a real-valued function defined on a subset D of $\mathbb{R}^n$ and let $g\colon E \to \mathbb{R}$ be a real-valued function defined on a subset E of $\mathbb{R}$, where $f(D) \subset E$. Suppose that f is continuous at some point $\mathbf{a}$ of D and that g is continuous at $f(\mathbf{a})$. Then the composition $g \circ f$ of f and g is continuous at $\mathbf{a}$.*

Proof Let $\varepsilon > 0$ be given. We must show that there exists some $\delta > 0$ such that

$$|g(f(\mathbf{x})) - g(f(\mathbf{a}))| < \varepsilon$$

for all points $\mathbf{x}$ of D which satisfy $|\mathbf{x} - \mathbf{a}| < \delta$. Now the definition of continuity, applied to the function g (with δ replaced by η), shows that, given $\varepsilon > 0$, there exists some $\eta > 0$ such that $|g(t) - g(f(\mathbf{a}))| < \varepsilon$ for all $t \in E$ which satisfy $|t - f(\mathbf{a})| < \eta$. Also the definition of continuity, applied to the function f shows that there exists $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{a})| < \eta$ for all $\mathbf{x} \in D$ which satisfy $|\mathbf{x} - \mathbf{a}| < \delta$. Thus if $\mathbf{x}$ is an element of D which satisfies $|\mathbf{x} - \mathbf{a}| < \delta$ then

$$|g(f(\mathbf{x})) - g(f(\mathbf{a}))| < \varepsilon.$$

This is what we set out to prove. ∎

1.5 Limits

Let $f\colon D \to \mathbb{R}$ be a real-valued function defined on a subset D of $\mathbb{R}^n$. Let $\mathbf{a}$ be a point of the closure $\overline{D}$ of D. Let l be a real number. We say that l is the *limit* of $f(\mathbf{x})$ is $\mathbf{x}$ *tends to* $\mathbf{a}$ if, for all $\varepsilon > 0$ there exists some $\delta > 0$ (which may depend on $\mathbf{a}$) such that

$$|f(\mathbf{x}) - l| < \varepsilon$$

whenever $\mathbf{x}$ belongs to D and

$$0 < |\mathbf{x} - \mathbf{a}| < \delta.$$

(Note that in the above definition of the limit of a function $f(\mathbf{x})$ as $\mathbf{x}$ tends to a point $\mathbf{a}$ we require the point $\mathbf{a}$ to belong to the closure $\overline{D}$ of the domain D of the function f. This is done merely in order to ensure that it is indeed sensible to talk about the limit of the function. For example, if f is a function defined on the set $\{t \in \mathbb{R} : t > 0\}$ of positive real numbers then it makes sense to talk about the limit of $f(t)$ as t tends to 1 or as t tends to 0, but it makes little sense to talk about the limit of f as t tends to -2 (say).)

The concept of a limit is very closely related to the concept of continuity. Indeed if $D\colon D \to \mathbb{R}$ is a real-valued function defined over a subset D of $\mathbb{R}^n$ then f is continuous at a point $\mathbf{a}$ of D if and only if $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})$ exists and is equal to $f(\mathbf{a})$. This follows immediately on comparing the formal definition of a limit given above with the formal definition of continuity.

1.6 'Calculus of Limits'

We shall prove various useful results concerning the behaviour of limits. The proofs involve so-called 'epsilon-delta' arguments. Any student who has difficulty in constructing proofs of this sort would be well-advised to study the the proofs of these results in some detail, since they exhibit many of the techniques used in constructing 'epsilon-delta' proofs.

Lemma 1.6 *Let f and g be real-valued functions defined on some subset D of $\mathbb{R}^n$. Let $\mathbf{a}$ be a point of the closure $\overline{D}$ of D. Suppose that*

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = l$$

and

$$\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = m.$$

Then

$$\lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x}) + g(\mathbf{x})) = l + m$$

and
$$\lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x}) - g(\mathbf{x})) = l - m.$$

Proof In order to prove that
$$\lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x}) + g(\mathbf{x})) = l + m,$$
we must show that, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that
$$|(f(\mathbf{x}) + g(\mathbf{x})) - (l + m)| < \varepsilon$$
for all points $\mathbf{x}$ of D which satisfy $0 < |\mathbf{x} - \mathbf{a}| < \delta$. Similarly, in order to prove that
$$\lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x}) - g(\mathbf{x})) = l - m,$$
we must show that, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that
$$|(f(\mathbf{x}) - g(\mathbf{x})) - (l - m)| < \varepsilon$$
for all points $\mathbf{x}$ of D which satisfy $0 < |\mathbf{x} - \mathbf{a}| < \delta$.

Now
$$|(f(\mathbf{x}) + g(\mathbf{x})) - (l + m)| \leq |f(\mathbf{x}) - l| + |g(\mathbf{x}) - m|$$
and
$$|(f(\mathbf{x}) - g(\mathbf{x})) - (l - m)| \leq |f(\mathbf{x}) - l| + |g(\mathbf{x}) - m|$$
(by Lemma 1.2). On applying the definition of limits to the function f, we conclude that, given any $\varepsilon_1 > 0$, there exists some $\delta_1 > 0$ such that $|f(\mathbf{x}) - l| < \varepsilon_1$ for all points $\mathbf{x}$ of D which satisfy $0 < |\mathbf{x} - \mathbf{a}| < \delta_1$. Similarly, applying the definition of limits to the function g, we conclude that, given any $\varepsilon_2 > 0$, there exists some $\delta_2 > 0$ such that $|g(\mathbf{x}) - m| < \varepsilon_2$ for all points $\mathbf{x}$ of D which satisfy $0 < |\mathbf{x} - \mathbf{a}| < \delta_2$. In particular, suppose that we choose $\varepsilon_1 = \frac{1}{2}\varepsilon$ and $\varepsilon_2 = \frac{1}{2}\varepsilon$. We conclude that there exist $\delta_1 > 0$ and $\delta_2 > 0$ such that $|f(\mathbf{x}) - l| < \frac{1}{2}\varepsilon$ at every point $\mathbf{x}$ of D satisfying $0 < |\mathbf{x} - \mathbf{a}| < \delta_1$, $|g(\mathbf{x}) - m| < \frac{1}{2}\varepsilon$ at every point $\mathbf{x}$ of D satisfying $0 < |\mathbf{x} - \mathbf{a}| < \delta_2$. Let δ be the minimum of δ_1 and δ_2. Then $\delta > 0$. Moreover, if $\mathbf{x}$ is a point of D which satisfies $0 < |\mathbf{x} - \mathbf{a}| < \delta$ then
$$\begin{aligned} &|(f(\mathbf{x}) + g(\mathbf{x})) - (l + m)| \\ &\qquad \leq |f(\mathbf{x}) - l| + |g(\mathbf{x}) - m| \\ &\qquad < \tfrac{1}{2}\varepsilon + \tfrac{1}{2}\varepsilon = \varepsilon \end{aligned}$$
and similarly
$$|(f(\mathbf{x}) - g(\mathbf{x})) - (l - m)| < \varepsilon.$$
But this is what we set out to prove. ∎

Next we show that the limit of a product of two functions is the product of the limits of these functions, provided that these limits exist. The method of proof is the same as that used in proving Lemma 1.6, but the details of the proof are somewhat more complicated.

Lemma 1.7 *Let f and g be real-valued functions defined on some subset D of $\mathbb{R}^n$. Let $\mathbf{a}$ be a point of the closure $\overline{D}$ of D. Suppose that*

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = l$$

and

$$\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = m.$$

Then

$$\lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x})g(\mathbf{x})) = lm.$$

Proof To show that $\lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x})g(\mathbf{x})) = lm.$ we must show that given any $\varepsilon > 0$ there exists some $\delta > 0$ such that

$$|f(\mathbf{x})g(\mathbf{x}) - lm| < \varepsilon$$

whenever $0 < |\mathbf{x} - \mathbf{a}| < \delta$. Now

$$f(\mathbf{x})g(\mathbf{x}) - lm = f(\mathbf{x})\,(g(\mathbf{x}) - m) + (f(\mathbf{x}) - l)\,m,$$

and hence

$$|f(\mathbf{x})g(\mathbf{x}) - lm| \leq |f(\mathbf{x})|\,|g(\mathbf{x}) - m| + |f(\mathbf{x}) - l|\,|m|.$$

It therefore suffices to show that

$$|f(\mathbf{x})|\,|g(\mathbf{x}) - m| < \tfrac{1}{2}\varepsilon$$

and

$$|f(\mathbf{x}) - l|\,|m| < \tfrac{1}{2}\varepsilon$$

for all points $\mathbf{x}$ of D that are sufficiently close to the point $\mathbf{a}$.

Now $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = l$. Thus, given any $\varepsilon_1 > 0$, there exists some $\delta_1 > 0$ such that $|f(\mathbf{x}) - l| < \varepsilon_1$ for all points $\mathbf{x}$ of D which satisfy $0 < |\mathbf{x} - \mathbf{a}| < \delta_1$. We first apply this result when

$$\varepsilon_1 = \frac{\varepsilon}{2(|m| + 1)}.$$

We conclude that there exists $\delta_1 > 0$ such that if $\mathbf{x}$ is a point of D which satisfies $0 < |\mathbf{x} - \mathbf{a}| < \delta_1$ then

$$|f(\mathbf{x}) - l| < \frac{\varepsilon}{2(|m| + 1)},$$

and hence

$$|f(\mathbf{x}) - l|\,|m| < \tfrac{1}{2}\varepsilon.$$

If we apply the definition of the limit of f as $\mathbf{x}$ tends to $\mathbf{a}$ (given above) in the particlar case where $\varepsilon_1 = 1$ we see that there exists $\delta_3 > 0$ such that if $0 < |\mathbf{x} - \mathbf{a}| < \delta_3$ then $|f(\mathbf{x}) - l| < 1$. Thus $|f(\mathbf{x})| < |l| + 1$ for all points $\mathbf{x}$ of D satisfying $0 < |\mathbf{x} - \mathbf{a}| < \delta_3$. Also the formal definition of the limit, applied to the function g, states that, given any $\varepsilon_2 > 0$, there exists some $\delta_2 > 0$ such that $|g(\mathbf{x}) - m| < \varepsilon_2$ for all points $\mathbf{x}$ of D which satisfy $0 < |\mathbf{x} - \mathbf{a}| < \delta_2$. We apply this result when

$$\varepsilon_2 = \frac{\varepsilon}{2(|l| + 1)}$$

in order to conclude that there exists $\delta_2 > 0$ such that if $\mathbf{x}$ is a point of D which satisfies $0 < |\mathbf{x} - \mathbf{a}| < \delta_2$ then

$$|g(\mathbf{x}) - m| < \frac{\varepsilon}{2(|l| + 1)}.$$

Thus if $x \in D$ satisfies the inequalities $0 < |\mathbf{x} - \mathbf{a}| < \delta_2$ and $0 < |\mathbf{x} - \mathbf{a}| < \delta_3$ then

$$|f(\mathbf{x})|\,|g(\mathbf{x}) - m| < (|l| + 1)\frac{\varepsilon}{2(|l| + 1)} = \tfrac{1}{2}\varepsilon.$$

Let us choose δ to be the minimum of δ_1, δ_2 and δ_3. Then $\delta > 0$. If $\mathbf{x}$ is any point of D which satisfies $0 < |\mathbf{x} - \mathbf{a}| < \delta$ then

$$|f(\mathbf{x}) - l|\,|m| < \tfrac{1}{2}\varepsilon.$$

and

$$|f(\mathbf{x})|\,|g(\mathbf{x}) - m| < \tfrac{1}{2}\varepsilon,$$

(since $0 < |\mathbf{x} - \mathbf{a}|$ is less than δ_1, δ_2 and δ_3). Thus if $0 < |\mathbf{x} - \mathbf{a}| < \delta$ then

$$|f(\mathbf{x})g(\mathbf{x}) - lm| < \varepsilon.$$

This shows that

$$\lim_{\mathbf{x} \to \mathbf{a}} (f(\mathbf{x})g(\mathbf{x})) = lm. \quad \blacksquare$$

It is perhaps worthwhile to describe in a more informal fashion the reasoning underlying the above proof. First of all we should have a clear idea of the objective: given any strictly positive real number ε, no matter how small, we have to show that the value of $f(\mathbf{x})g(\mathbf{x})$ is within ε of the number lm for all points $\mathbf{x}$ of D that are 'sufficiently close' to the point $\mathbf{a}$. The basic trick involved is to observe that

$$f(\mathbf{x})g(\mathbf{x}) - lm = f(\mathbf{x})\,(g(\mathbf{x}) - m) + (f(\mathbf{x}) - l)\,m.$$

Thus we can ensure that $f(\mathbf{x})g(\mathbf{x})$ is close to lm by arranging matters so that $f(\mathbf{x})$ is close to l and $g(\mathbf{x})$ is close to m. More specifically, if we arrange matters such that

$$|f(\mathbf{x})|\,|g(\mathbf{x}) - m| < \tfrac{1}{2}\varepsilon$$

and

$$|f(\mathbf{x}) - l|\,|m| < \tfrac{1}{2}\varepsilon$$

then

$$|f(\mathbf{x})g(\mathbf{x}) - lm| < \varepsilon.$$

This illustrates an obvious but useful technique: if we have a sum of k terms and we wish to ensure that the modulus of the sum is less than some strictly positive number ε then it suffices to arrange matters so that the modulus of each individual term is less than ε/k. We now have to work on the terms $|f(\mathbf{x})|\,|g(\mathbf{x}) - m|$ and $|f(\mathbf{x}) - l|\,|m|$. The second of these expressions is the easier to work with: to show that $|f(\mathbf{x}) - l|\,|m|$ is less than $\frac{1}{2}\varepsilon$, we merely have to show that $|f(\mathbf{x}) - l|$ is less than $\varepsilon/(2(|m| + 1))$ for all points $\mathbf{x}$ of D 'sufficiently close' to $\mathbf{a}$. To do this we simply apply the definition of the limit (applied to the function f).

It remains to show that $|f(\mathbf{x})|\,|g(\mathbf{x}) - m|$ is less than $\frac{1}{2}\varepsilon$ at all points $\mathbf{x}$ of D that are 'sufficiently close' to $\mathbf{a}$. We know that $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = l$. Therefore the $f(\mathbf{x})$ is close to l at all points $\mathbf{x}$ of D that are 'sufficiently close' to $\mathbf{a}$. In particular, if $\mathbf{x}$ is 'sufficiently close' to $\mathbf{a}$, then $|f(\mathbf{x})| < |l| + 1$. Thus it suffices to show that $|g(\mathbf{x}) - m|$ is less than $\varepsilon/(2(|l| + 1))$ for all points $\mathbf{x}$ of D that are 'sufficiently close' to $\mathbf{a}$. To do this we simply apply the definition of the limit (applied to the function g).

Thus in order to show that

$$|f(\mathbf{x})g(\mathbf{x}) - lm| < \varepsilon$$

at all points $\mathbf{x}$ of D that are sufficiently close to the point $\mathbf{a}$ it suffices to choose δ to be a strictly positive real number that is chosen sufficiently small that if $\mathbf{x}$ is any point of D whose distance $|\mathbf{x} - \mathbf{a}|$ from $\mathbf{a}$ is less than δ then all three inequalities

$$|f(\mathbf{x}) - l| < \frac{\varepsilon}{2(|m| + 1)},$$

$$|g(\mathbf{x}) - m| < \frac{\varepsilon}{2(|l| + 1)},$$

$$|f(\mathbf{x})| < |l| + 1$$

are valid simultaneously, then the inequality

$$|f(\mathbf{x})g(\mathbf{x}) - lm| < \varepsilon$$

holds for all values $\mathbf{x}$ of D for which $0 < |\mathbf{x} - \mathbf{a}| < \delta$. This is what we need to prove in order to show that

$$\lim_{\mathbf{x} \to \mathbf{a}} (f(\mathbf{x})g(\mathbf{x})) = lm.$$

Next we examine the effect on limit of a function on composing that function with some other continuous function.

Lemma 1.8 *Let $f \colon D \to \mathbb{R}$ be a real-valued function defined on a subset D of $\mathbb{R}^n$, and let $\mathbf{a}$ be a point of the closure $\overline{D}$ of the set D. Suppose that*

$$\lim_{\mathbf{x} \to \mathbf{a}} f(\mathbf{x}) = l$$

for some real number l. Let $g \colon E \to \mathbb{R}$ be a continuous real-valued function defined on a subset E of $\mathbb{R}$, where E contains an open neighbourhood of l. Then

$$\lim_{\mathbf{x} \to \mathbf{a}} g(f(\mathbf{x})) = g(l).$$

Proof Let $\varepsilon > 0$ be given. We must show that there exists some $\delta > 0$ such that

$$|g(f(\mathbf{x})) - g(l)| < \varepsilon$$

for all points $\mathbf{x}$ of D which satisfy $0 < |\mathbf{x} - \mathbf{a}| < \delta$. Now the definition of continuity, applied to the function g (with δ replaced by η), shows that, given $\varepsilon > 0$, there exists some $\eta > 0$ such that $|g(t) - g(l)| < \varepsilon$ for all $t \in E$ which satisfy $|t - l| < \eta$. Also the definition of limits, applied to the function f shows that there exists $\delta > 0$ such that $|f(\mathbf{x}) - l| < \eta$ for all $\mathbf{x} \in D$ which satisfy $0 < |\mathbf{x} - \mathbf{a}| < \delta$. Thus if $\mathbf{x}$ is an element of D which satisfies $0 < |\mathbf{x} - \mathbf{a}| < \delta$ then

$$|g(f(\mathbf{x})) - g(l)| < \varepsilon.$$

This is what we set out to prove. ∎

Lemma 1.9 *Let f and g be real-valued functions defined on some open subset D of $\mathbb{R}^n$. Let $\mathbf{a}$ be a point of the closure $\overline{D}$ of D. Suppose that*

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = l$$

and

$$\lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) = m,$$

where $m \neq 0$. Then

$$\lim_{\mathbf{x}\to\mathbf{a}} \left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) = \frac{l}{m}.$$

Proof Let $r: \mathbb{R} \setminus \{0\} \to \mathbb{R}$ be the reciprocal function on the set $\mathbb{R} \setminus \{0\}$ of non-zero real numbers defined by $r(t) = 1/t$. We show that this function r is continuous on $\mathbb{R} \setminus \{0\}$. Once we have shown this, then we can apply Lemmas 1.7 and 1.8 in order to deduce the desired result. Indeed

$$\lim_{\mathbf{x}\to\mathbf{a}} \left(\frac{1}{g(\mathbf{x})}\right) = \lim_{\mathbf{x}\to\mathbf{a}} r(g(\mathbf{x})) = r(m) = \frac{1}{m}$$

by Lemma 1.8, and hence

$$\lim_{\mathbf{x}\to\mathbf{a}} \left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) = \frac{l}{m},$$

by Lemma 1.7.

We must therefore show that the reciprocal function r is continuous on $\mathbb{R} \setminus \{0\}$. Let s be a non-zero real number. Suppose that we are given some $\varepsilon > 0$. Let δ be the mininum of $\frac{1}{2}|s|$ and $\frac{1}{2}\varepsilon|s|^2)$. If t is a real number satisfying $|t - s| < \delta$ then $|t| \geq |s| - |t - s|$ and so $|t| > \frac{1}{2}|s|$. Thus if t is any non-zero real number satisfying $|t - s| < \delta$ then

$$\left|\frac{1}{t} - \frac{1}{s}\right| = \left|\frac{s - t}{ts}\right| < \frac{2}{|s|^2}|t - s| < \varepsilon$$

(since $|t - s| < \delta \leq \frac{1}{2}\varepsilon|s|^2)$. We have therefore shown that, given any $\varepsilon > 0$ there exists some $\delta > 0$ such that if t is any non-zero real number satisfying $|t - s| < \delta$ then $|r(t) - r(s)| < \varepsilon$. Thus we have verified that the reciprocal function r is continuous at s, where s is an arbitrary non-zero real number. This completes the proof, since we have already shown that the desired result follows from the continuity of the reciprocal function. ∎

1.7 Properties of Continuous Functions

Let $f\colon D \to \mathbb{R}$ be a real-valued function defined on a subset D of $\mathbb{R}^n$. We noted above that the function f is continuous at a point $\mathbf{a}$ of D if and only if the limit $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})$ exists and is equal to $f(\mathbf{a})$.

If f and g are real-valued functions defined on a subset D of $\mathbb{R}^n$ then we denote by $f+g$,$f-g$ and $f.g$ the *sum*, *difference* and *product* respectively of the functions f and g. Naturally these functions are defined by

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

If g is not identically zero then the *quotient* f/g of f and g is defined on the set $\{\mathbf{x} \in D : g(\mathbf{x}) \neq 0\}$ (where $(f/g)(\mathbf{x}) = f(\mathbf{x})/g(\mathbf{x})$).

Theorem 1.10 *Let f and g be real-valued functions defined on a subset D of n-dimensional Euclidean space $\mathbb{R}^n$. Let $\mathbf{a}$ be a point of D. Suppose that the functions f and g are continuous at the point $\mathbf{a}$. Then the functions $f+g$ and $f-g$ and $f.g$ are also continuous at $\mathbf{a}$. If also $g(\mathbf{a})$ is non-zero then the function f/g is continuous at $\mathbf{a}$.*

Proof We know that

$$\begin{aligned} \lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) &= f(\mathbf{a}), \\ \lim_{\mathbf{x}\to\mathbf{a}} g(\mathbf{x}) &= g(\mathbf{a}), \end{aligned}$$

because f and g are continuous at $\mathbf{a}$. We conclude from Lemmas 1.6 and 1.7 that

$$\begin{aligned} \lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x}) + g(\mathbf{x})) &= f(\mathbf{a}) + g(\mathbf{a}), \\ \lim_{\mathbf{x}\to\mathbf{a}} (f(\mathbf{x}) - g(\mathbf{x})) &= f(\mathbf{a}) - g(\mathbf{a}), \\ \lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})g(\mathbf{x}) &= f(\mathbf{a})g(\mathbf{a}). \end{aligned}$$

Hence $f+g$, $f-g$ and $f.g$ are continuous at $\mathbf{a}$. Similarly it follows from Lemma 1.9 that f/g is continuous at $\mathbf{a}$. ∎

We have already proved that a composition of continuous functions is itself continuous (see Lemma 1.5). Using this result and Theorem 1.10 we obtain a plentiful supply of continuous functions: every polynomial function is continuous, and a rational function (i.e., a quotient of polynomial functions) is continuous at all points at which it is well-defined.

1.8 'Vector-Valued Functions'

Let D be a subset of n-dimensional Euclidean space $\mathbb{R}^n$. We shall be considering 'vector-valued' functions $\varphi: D \to \mathbb{R}^m$ that map D into $\mathbb{R}^m$ for some positive integer m. A typical example of such a function is provided by a vector field on 3-dimensional Euclidean space (such as an electric field or a magnetic field). Such a vector field, defined over some subset D of 3-dimensional Euclidean space can be regarded as a function from D to $\mathbb{R}^3$ (i.e., it maps a point of D to a triple of real numbers which specify the 3 components of the vector field at that point).

The definitions of continuity and of limits for such functions are analogous to those for real-valued functions that are given above. Thus a function $\varphi: D \to \mathbb{R}$ which maps a subset D of $\mathbb{R}^n$ into $\mathbb{R}^m$ is said to be *continuous* at an element $\mathbf{a}$ of D if and only if, for all $\varepsilon > 0$ there exists some $\delta > 0$ (which may depend on $\mathbf{a}$) such that

$$|\varphi(\mathbf{x}) - \varphi(\mathbf{a})| < \varepsilon$$

whenever $\mathbf{x}$ belongs to D and

$$|\mathbf{x} - \mathbf{a}| < \delta.$$

(Here of course $|\varphi(\mathbf{x}) - \varphi(\mathbf{a})|$ measures the distance from $\varphi(\mathbf{x})$ to $\varphi(\mathbf{a})|$ in the Euclidean space $\mathbb{R}^m$.

If $\mathbf{a}$ is an element of the closure of D and if $\mathbf{v}$ is an element of $\mathbb{R}^m$ then we say that $\mathbf{v}$ is the *limit* of $\varphi(\mathbf{x})$ as $\mathbf{x}$ tends to $\mathbf{a}$ if, for every $\varepsilon > 0$ there exists some $\delta > 0$ such that $|\varphi(\mathbf{x}) - \mathbf{v}| < \varepsilon$ whenever $\mathbf{x}$ belongs to D and $0 < |\mathbf{x} - \mathbf{a}| < \delta$.

If $\varphi: D \to \mathbb{R}^m$ is a function which maps a subset D of $\mathbb{R}^n$ into $\mathbb{R}^m$ then the *components* of the function φ are the real-valued functions $\varphi_1, \varphi_2, \ldots, \varphi_m$ defined such that

$$\varphi(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots, \varphi_n(\mathbf{x}))\,.$$

The following lemma enables us to deduce results concerning 'vector-valued functions' directly from corresponding results for real-valued functions.

Lemma 1.11 *Let $\varphi: D \to \mathbb{R}^m$ be a function which maps a subset D of $\mathbb{R}^n$ into $\mathbb{R}^m$. Let $\mathbf{a}$ be a point of the closure $\overline{D}$ of D and let $\mathbf{v}$ be an element of $\mathbb{R}^m$. Then*

$$\lim_{\mathbf{x} \to \mathbf{a}} \varphi(\mathbf{x}) = \mathbf{v}$$

if and only if

$$\lim_{\mathbf{x}\to\mathbf{a}} \varphi_i(\mathbf{x}) = v_i$$

for $i = 1, 2, \ldots, m$, where $\varphi_i(\mathbf{x})$ denotes the ith component of $\varphi(\mathbf{x})$ and v_i denotes the ith component of $\mathbf{v}$.

Proof Suppose that $\lim_{\mathbf{x}\to\mathbf{a}} \varphi(\mathbf{x}) = \mathbf{v}$. We show that $\lim_{\mathbf{x}\to\mathbf{a}} \varphi_i(\mathbf{x}) = v_i$ for all i. Let $\varepsilon > 0$ be given. Then there exists some $\delta > 0$ such that $|\varphi(\mathbf{x}) - \mathbf{v}| < \varepsilon$ whenever $\mathbf{x}$ belongs to D and $0 < |\mathbf{x} - \mathbf{a}| < \delta$. But

$$|\varphi_i(\mathbf{x}) - v_i| \leq |\varphi(\mathbf{x}) - \mathbf{v}|$$

for all i, hence $|\varphi_i(\mathbf{x}) - v_i| < \varepsilon$ whenever $\mathbf{x}$ belongs to D and $0 < |\mathbf{x} - \mathbf{a}| < \delta$. This shows that $\lim_{\mathbf{x}\to\mathbf{a}} \varphi_i(\mathbf{x}) = v_i$ for all i.

Conversely we must show that if $\lim_{\mathbf{x}\to\mathbf{a}} \varphi_i(\mathbf{x}) = v_i$ for $i = 1, 2, \ldots, m$ then $\lim_{\mathbf{x}\to\mathbf{a}} \varphi(\mathbf{x}) = \mathbf{v}$. Let $\varepsilon > 0$ be given. Applying the definition of the limit of a function (with ε replaced by $\varepsilon/\sqrt{m}$ we see that there exist strictly positive real numbers δ_1, δ_2,. . ., δ_m such that

$$|\varphi_i(\mathbf{x}) - v_i| < \frac{\varepsilon}{\sqrt{m}}$$

whenever $\mathbf{x}$ belongs to D and $0 < |\mathbf{x} - \mathbf{a}| < \delta_i$. Let δ be the minimum of δ_1, δ_2,. . ., δ_m. Then $\delta > 0$, and if $\mathbf{x}$ is a point of D which satisfies $0 < |\mathbf{x} - \mathbf{a}| < \delta$ then

$$|\varphi_i(\mathbf{x}) - v_i| < \frac{\varepsilon}{\sqrt{m}}$$

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

$$|\varphi(\mathbf{x}) - \mathbf{v}|^2 = \sum_{i=1}^{m} |\varphi_i(\mathbf{x}) - v_i|^2,$$

hence $|\varphi(\mathbf{x}) - \mathbf{v}|^2 < \varepsilon^2$. We have therefore shown that, given any $\varepsilon > 0$ there exists some $\delta > 0$ such that $|\varphi(\mathbf{x}) - \mathbf{v}| < \varepsilon$ whenever $\mathbf{x}$ belongs to D and $0 < |\mathbf{x} - \mathbf{a}| < \delta$. We deduce that $\lim_{\mathbf{x}\to\mathbf{a}} \varphi(\mathbf{x}) = \mathbf{v}$, as required. ▮

Using this result together with Lemma 1.6 we deduce immediately that if $\varphi\colon D \in \mathbb{R}^m$ and $\psi\colon D \in \mathbb{R}^m$ are functions which map D into $\mathbb{R}^m$ and if $\lim_{\mathbf{x}\to\mathbf{a}} \varphi(\mathbf{x}) = \mathbf{v}$ and $\lim_{\mathbf{x}\to\mathbf{a}} \psi(\mathbf{x}) = \mathbf{w}$ then

$$\lim_{\mathbf{x}\to\mathbf{a}} (\varphi(\mathbf{x}) + \psi(\mathbf{x})) = \mathbf{v} + \mathbf{w}$$

and

$$\lim_{\mathbf{x}\to\mathbf{a}} (\varphi(\mathbf{x}) - \psi(\mathbf{x})) = \mathbf{v} - \mathbf{w}.$$

Similarly we can use Lemma 1.7 together with Lemma 1.11 to deduce that

$$\lim_{\mathbf{x}\to\mathbf{a}} \varphi(\mathbf{x}).\psi(\mathbf{x}) = \mathbf{v}.\mathbf{w}.$$

Also if $f: D \to \mathbb{R}$ is a real-valued function on D and if $\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x}) = \lambda$ then

$$\lim_{\mathbf{x}\to\mathbf{a}} f(\mathbf{x})\varphi(\mathbf{x}) = \lambda\mathbf{v}.$$

Thus if the functions φ, ψ and f are continuous at $\mathbf{a}$ then so are $\varphi+\psi$, $\varphi-\psi$, $\varphi.\psi$ and $f\varphi$ (where $(\varphi.\psi)(\mathbf{x}) = \varphi(\mathbf{x}).\psi(\mathbf{x})$).

We now show that a composition of continuous 'vector-valued' functions is continuous. The proof is closely modelled on that of Lemma 1.5.

Lemma 1.12 *Let $\varphi: D \to \mathbb{R}^m$ be a function defined on a subset D of $\mathbb{R}^n$ and let $\psi: E \to \mathbb{R}^p$ be a real-valued function defined on a subset E of $\mathbb{R}^m$, where $\varphi(D) \subset E$. Suppose that φ is continuous at some point $\mathbf{a}$ of D and that ψ is continuous at $\varphi(\mathbf{a})$. Then the composition $\psi \circ \varphi$ of φ and ψ is continuous at $\mathbf{a}$.*

Proof Let $\varepsilon > 0$ be given. We must show that there exists some $\delta > 0$ such that

$$|\psi(\varphi(\mathbf{x})) - \psi(\varphi(\mathbf{a}))| < \varepsilon$$

for all points $\mathbf{x}$ of D which satisfy $|\mathbf{x}-\mathbf{a}| < \delta$. Now the definition of continuity, applied to the function ψ (with δ replaced by η), shows that, given $\varepsilon > 0$, there exists some $\eta > 0$ such that $|\psi(\mathbf{u}) - \psi(\varphi(\mathbf{a}))| < \varepsilon$ for all $\mathbf{u} \in E$ which satisfy $|\mathbf{u} - \varphi(\mathbf{a})| < \eta$. Also the definition of continuity, applied to the function φ shows that there exists $\delta > 0$ such that $|\varphi(\mathbf{x}) - \varphi(\mathbf{a})| < \eta$ for all $\mathbf{x} \in D$ which satisfy $|\mathbf{x} - \mathbf{a}| < \delta$. Thus if $\mathbf{x}$ is an element of D which satisfies $|\mathbf{x} - \mathbf{a}| < \delta$ then

$$|\psi(\varphi(\mathbf{x})) - \psi(\varphi(\mathbf{a}))| < \varepsilon.$$

This is what we set out to prove. ▮

1.9 The Relationship between Continuous Functions and Open Sets

One can give a very elegant characterization of the notion of continuity in terms of open sets. For simplicity, we consider functions defined over the whole of $\mathbb{R}^n$.

Let $\varphi: \mathbb{R}^n \to \mathbb{R}^m$ be a function that maps $\mathbb{R}^n$ into $\mathbb{R}^m$. Given any subset V of $\mathbb{R}^m$, we denote by $\varphi^{-1}(V)$ the subset

$$\varphi^{-1}(V) = \{\mathbf{x} \in \mathbb{R}^n : \varphi(\mathbf{x}) \in V\}$$

of $\mathbb{R}^n$ consisting of all points of $\mathbb{R}^n$ that are mapped into V by φ.

Theorem 1.13 *Let $\varphi: \mathbb{R}^n \to \mathbb{R}^n$ be a real-valued function over the whole of n-dimensional Euclidean space $\mathbb{R}^n$. Then the function φ is continuous on $\mathbb{R}^n$ if and only if $\varphi^{-1}(V)$ is an open subset of $\mathbb{R}^n$ for every open subset V of $\mathbb{R}^m$*

Proof Suppose that $\varphi: \mathbb{R}^n \to \mathbb{R}^m$ is continuous. Let V be an open set in $\mathbb{R}^m$. We must show that $\varphi^{-1}(V)$ is open. Let $\mathbf{a}$ be a point of $\varphi^{-1}(V)$. Then $\varphi(\mathbf{a}) \in V$ and hence there exists some $\varepsilon > 0$ such that the open ball $B(\varphi(\mathbf{a}), \varepsilon)$ of radius ε about $\varphi(\mathbf{a})$ is contained in the open set V. Applying the definition of continuity to the function φ we see that there exists some $\delta > 0$ such that $|\varphi(\mathbf{x}) - \varphi(\mathbf{a})| < \varepsilon$ whenever $|\mathbf{x} - \mathbf{a}| < \delta$. This means that the image $\varphi(B(\mathbf{a}, \delta))$ of the open ball $B(\mathbf{a}, \delta)$ under the map φ is contained in $B(\varphi(\mathbf{a}), \varepsilon)$ and is thus contained in V. Therefore

$$B(\mathbf{a}, \delta) \subset \varphi^{-1}(V).$$

We have therefore shown that given any point $\mathbf{a}$ of $\varphi^{-1}(V)$ there exists an open ball about $\mathbf{a}$ which is contained in the set $\varphi^{-1}(V)$. Thus we have shown that $\varphi^{-1}(V)$ is open.

Conversely, suppose that $\varphi: \mathbb{R}^n \to \mathbb{R}^m$ is a map with the property that $\varphi^{-1}(V)$ is an open subset of $\mathbb{R}^n$ for every open subset V of $\mathbb{R}^m$. We must show that φ is continuous. Thus we must show that, given any point $\mathbf{a}$ of $\mathbb{R}^n$ and given any $\varepsilon > 0$ there exists some $\delta > 0$ (where δ might depend on the choice of the point $\mathbf{a}$) such that $|\varphi(\mathbf{x}) - \varphi(\mathbf{a})| < \varepsilon$ whenever $|\mathbf{x} - \mathbf{a}| < \delta$. But if we are given $\varepsilon > 0$ then the open ball $B(f(\mathbf{a}), \varepsilon)$ of radius ε about $\mathbf{a}$ is an open set in $\mathbb{R}^m$, and therefore $\varphi^{-1}(B(f(\mathbf{a}), \varepsilon))$ is an open set in $\mathbb{R}^n$. But $\mathbf{a}$ is a point of this set, hence there exists some $\delta > 0$ such that

$$B(\mathbf{a}, \delta) \subset \varphi^{-1}(B(f(\mathbf{a}), \varepsilon))$$

(where $B(\mathbf{a}, \delta)$ is the open ball of radius δ about $\mathbf{a}$). But this means that if $\mathbf{x}$ is a point of $\mathbb{R}^n$ which satisfies $|\mathbf{x} - \mathbf{a}| < \delta$ then $|\varphi(\mathbf{x}) - \varphi(\mathbf{a})| < \varepsilon$. Thus φ is continuous at $\mathbf{a}$. ∎

Applying this theorem in the case of a function mapping $\mathbb{R}^n$ into $\mathbb{R}$ we deduce the following result.

Corollary 1.14 *Let $f: \mathbb{R}^n \to \mathbb{R}$ be a continuous real-valued function defined over the whole of $\mathbb{R}^n$. Then, for every real number c, the sets*

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) > c\}$$

and

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < c\}$$

are open.

Proof Note that

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) > c\} = f^{-1}\{t \in \mathbb{R} : t > c\}$$

and

$$\{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) < c\} = f^{-1}\{t \in \mathbb{R} : t < c\}.$$

Also the sets $\{t \in \mathbb{R} : t > c\}$ and $\{t \in \mathbb{R} : t > c\}$ are open subsets of $\mathbb{R}$. The result therefore follows immediately from the preceding theorem. ∎

1.10 Sequences

Let $(\mathbf{x}_i : i \in \mathbb{N})$ be a sequence of points in $\mathbb{R}^n$. The sequence is said to *converge* to a point $\mathbf{a}$ of $\mathbb{R}^n$ if, for any $\varepsilon > 0$ there exists some natural number N such that $|\mathbf{x}_i - \mathbf{a}| < \varepsilon$ for all i satisfying $i \geq N$. If the sequence $(\mathbf{x}_i : i \in \mathbb{N})$ converges to the point $\mathbf{a}$ then we say that $\mathbf{a}$ is the *limit* of the sequence, and we write

$$\mathbf{a} = \lim_{i \to +\infty} \mathbf{x}_i.$$

One can prove that if the sequences $(\mathbf{x}_i : i \in \mathbb{N})$ $(\mathbf{y}_i : i \in \mathbb{N})$ in $\mathbb{R}^n$ converge to points $\mathbf{a}$ and $\mathbf{b}$ of $\mathbb{R}^n$, then

$$\begin{aligned} \lim\,(\mathbf{x}_i + \mathbf{y}_i) &= \mathbf{a} + \mathbf{b}, \\ \lim\,(\mathbf{x}_i . \mathbf{y}_i) &= \mathbf{a} . \mathbf{b}. \end{aligned}$$

If also $(\lambda_i : i \in \mathbb{N})$ is a sequence of real numbers which converges to a real number β then

$$\lim \lambda_i \mathbf{x}_i = \beta \mathbf{a}.$$

We now prove two important properties of continuous functions.

Lemma 1.15 *Let D be a subset of n-dimensional Euclidean space $\mathbb{R}^n$, and let $(\mathbf{x}_i : i \in \mathbb{N})$ be a sequence of points in D which converges to some point $\mathbf{a}$ of D. Let $\phi: D \to \mathbb{R}^m$ be a continuous function mapping D into $\mathbb{R}^m$ for some m. Then*

$$\lim_{i \to +\infty} \phi(\mathbf{x}_i) = \phi(\mathbf{a}).$$

Proof Suppose that we are given some $\varepsilon > 0$. Then there exists some $\delta > 0$ such that $|\phi(\mathbf{x}) - \phi(\mathbf{a})| < \varepsilon$ for all points $\mathbf{x}$ of D which satisfy $|\mathbf{x} - \mathbf{a}| < \delta$. But from the definition of convergence (with δ playing the role of ϵ) we see that there exists some positive integer N such that $|\mathbf{x}_i - \mathbf{a}| < \delta$ for all i satisfying $i \geq n$. Thus $|\phi(\mathbf{x}_i) - \phi(\mathbf{a})| < \varepsilon$ for all i satisfying $i \geq N$. Thus $\phi(\mathbf{x}_i)$ converges to $\phi(\mathbf{a})$ as i tends to $+\infty$, as required. ∎

Lemma 1.16 *Let D be a subset of n-dimensional Euclidean space $\mathbb{R}^n$, and let $(\mathbf{x}_i : i \in \mathbb{N})$ be a sequence of points in D which converges to some point $\mathbf{a}$ of D. Let $f: D \to \mathbb{R}$ be a continuous real-valued function on D. If $f(\mathbf{x}_i) \geq c$ for all positive integers i, where c is some real number, then $f(\mathbf{a}) \geq c$. Similarly, if $f(\mathbf{x}_i) \leq d$ for all positive integers i, where d is some real number, then $f(\mathbf{a}) \leq d$.*

Proof Suppose that $f(\mathbf{x}_i) \geq c$ for all i. We must show that $f(\mathbf{a}) \geq c$. We shall show that a contradiction would arise, were it the case that $f(\mathbf{a}) < c$. Indeeed suppose that the inequality $f(\mathbf{a}) < c$ were to hold. Then there would exist some $\varepsilon > 0$ such that $f(\mathbf{a}) + \varepsilon < c$. There would then exist some $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{a})| < \varepsilon$ for all points $\mathbf{x}$ of D which satisfy $|\mathbf{x} - \mathbf{a}| < \delta$, since f is continuous. In particular, $f(\mathbf{x}) < c$ for all points $\mathbf{x}$ of D which satisfy $|\mathbf{x} - \mathbf{a}| < \delta$. But then there would exist some positive integer N such that $|\mathbf{x}_i - \mathbf{a}| < \delta$ for all i satisfying $i \geq N$, since the sequence $(\mathbf{x}_i : i \in \mathbb{N})$ converges to $\mathbf{a}$. Thus we would have $f(\mathbf{x}_i) < c$ for all i satisfying $i \geq N$, contradicting the condition that $f(\mathbf{x}_i) \geq c$ for all i. Thus we have shown that $f(\mathbf{a}) \geq c$. An analogous proof shows that if $f(\mathbf{x}_i) \leq d$ for all i then $f(\mathbf{a}) \leq d$. ∎

We now prove an important property of closed sets in Euclidean space.

Lemma 1.17 *Let F be a closed subset of n-dimensional Euclidean space $\mathbb{R}^n$. Let $(\mathbf{x}_i : i \in \mathbb{N})$ be a sequence of points of F which converges to some point $\mathbf{a}$ of $\mathbb{R}^n$. Then the point $\mathbf{a}$ belongs to F.*

Proof Suppose that it were the case that $\mathbf{a}$ does not belong to F. We show that this leads to a contradiction. Now the definition of a closed set tells us that the complement of F in $\mathbb{R}^n$ is open. Thus if it were the case that $\mathbf{a}$ does not belong to F then there would exist some $\delta > 0$ such that if $\mathbf{x}$ is any point of $\mathbb{R}^n$ which satisfies $|\mathbf{x} - \mathbf{a}| < \delta$ then $\mathbf{x}$ does not belong to F. But then there would exist some positive integer N such that $|\mathbf{x}_i - \mathbf{a}| < \delta$ for all i satisfying $i \geq N$, since the sequence $(\mathbf{x}_i)$ converges to $\mathbf{a}$. Thus we would have that $\mathbf{x}_i$ does not belong to F if $i \geq N$. This the required contradiction. We conclude therefore that $\mathbf{a}$ belongs to F. ∎

Note that one can use Lemma 1.15 and Lemma 1.17 to give another proof of Lemma 1.16. For let $(\mathbf{x}_i : i \in \mathbb{N})$ be a sequence of points contained in some subset D of $\mathbb{R}^n$ which converges to a point $\mathbf{a}$ of $\mathbb{R}^n$, and let $f: D \to \mathbb{R}$ be a continuous real-valued function on D. Then $f(\mathbf{x}_i)$ converges to $f(\mathbf{a})$ as i tends to $+\infty$, by Lemma 1.15. Suppose that $f(\mathbf{x}_i) \geq c$ for all i. Now the subset $[c, +\infty)$ consisting of all real numbers t satisfying $t \geq c$ is a closed subset of $\mathbb{R}$. But $f(\mathbf{x}_i) \in [c, +\infty)$ for all i, hence $f(\mathbf{a}) \in [c, +\infty)$, by Lemma 1.17. Thus $f(\mathbf{a}) \geq c$. Similarly, if $f(\mathbf{x}_i) \leq d$ for all i then $f(\mathbf{a}) \leq d$.

2 The Bolzano-Weierstrass Theorem

The theorems proved in this section make use of the *least upper bound principle.* This principle states that any subset S of the set $\mathbb{R}$ of real numbers which is bounded above has a *least upper bound* (or *supremum*), denoted by $\sup S$. Similarly any subset of $\mathbb{R}$ which is bounded below has a *greatest lower bound* or infimum.

One of the consequences of the least upper bound principle is the fact that bounded increasing sequences of real numbers always converge, as do bounded decreasing sequences of real numbers.

Lemma 2.1 *Let $(s_i : i \in \mathbb{N})$ be an increasing sequence of real numbers. Suppose that there exists some constant C such that $s_i \leq C$ for all $i \in \mathbb{N}$. Then the sequence (s_i) converges, and the limit s of this sequence satisfies $s \leq C$.*

Similarly let $(t_i : i \in \mathbb{N})$ be a decreasing sequence of real numbers. Suppose that there exists some constant C' such that $t_i \geq C'$ for all $i \in \mathbb{N}$. Then the sequence (t_i) converges, and the limit t of this sequence satisfies $t \geq C'$.

Proof Let S be the set $S = \{s_i : i \in \mathbb{N}\}$ consisting of all of the elements of the sequence (s_i). Then the constant C is an upper bound for the set S. By the least upper bound principle we conclude that the set S has a least upper bound $\sup S$. Define $s = \sup S$. We claim that s is the limit of the sequence (s_i).

Let $\varepsilon > 0$ be given. Now there must exist some element s_N of the sequence (s_i) with the property that $s_N > s - \varepsilon$, since otherwise $s - \varepsilon$ would be an upper bound for the set S, contradicting the definition of s as the *least* upper bound of this set. But the sequence (s_i) is increasing, so that if $i \geq N$ then $s_i \geq s_N > s - \varepsilon$. Also $s_i \leq s$ for all i (because s is an upper bound for the set S. Thus if $i \geq N$ then $s - \varepsilon < s_i < s$, and hence $|s_i - s| < \varepsilon$.

We have therefore shown that, given any $\varepsilon > 0$ there exists some positive integer N such that $|s_i - s| < \varepsilon$ whenever $i \geq N$. Thus we have shown that s is the limit of the sequence s, and moreover $s \leq C$ (because C is an upper bound for the set S and s is by definition the least upper bound of this set).

An analogous proof shows that the bounded decreasing sequence $(t_i : i \in \mathbb{N})$ converges. Indeed the limit of this sequence is $\inf T$, where $T = \{t_i : i \in \mathbb{N}\}$. ∎

We now define the concept of a *subsequence* of a sequence $(\mathbf{s}_i : i \in \mathbb{N})$ of elements of $\mathbb{R}^n$. A *subsequence* of the sequence $(\mathbf{s}_i)$ is defined to be a sequence of the form $(\mathbf{s}_{i(j)} : j \in \mathbb{N})$, where $(i(j) : j \in \mathbb{N})$ is an increasing sequence of positive integers. Thus a subsequence is a sequence obtained by making a selection of elements of the original sequence. For example, if $(\mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3, \ldots)$ is a sequence of elements of $\mathbb{R}^n$ then $(\mathbf{s}_2, \mathbf{s}_4, \mathbf{s}_6, \ldots)$ is a subsequence of the original sequence.

An important theorem of analysis is the Bolzano-Weierstrass Theorem, which states that every bounded sequence of real numbers possesses a convergent subsequence. We now prove this result. As we shall see, the proof given here makes essential use of the Least Upper Bound Principle.

Theorem 2.2 (Bolzano-Weierstrass) *Every sequence of real numbers that is bounded above and below possesses a convergent subsequence.*

Proof Let $(s_i : i \in \mathbb{N})$ be a sequence of real numbers that is bounded above and below. For all positive integers m let us define

$$S_m = \{s_i \in \mathbb{R} : i \geq m\}.$$

Then each S_m is a subset of $\mathbb{R}$ that is bounded above and below. Moreover $S_r \subset S_m$ whenever $r > m$. Let us define $a_m = \sup S_m$ for all positive integers m. (Note that a_m exists, by the Least Upper Bound Principle, because each S_m is bounded above.) Now there exists some constant C such that $s_i \geq C$ for all i, since the sequence (s_i) is bounded below. But $a_m \geq s_m$, hence $a_m \geq C$ for all positive integers m. It follows from Lemma 2.1 that the sequence $(a_m : m \in \mathbb{N})$ converges to some real number c. We shall show that there exists a subsequence of the sequence $(s_i : i \in \mathbb{N})$ which converges to c.

We construct a subsequence $(s_{i(j)} : j \in \mathbb{N})$ of the sequence (s_i) with the property that $|s_{i(j)} - c| < 1/j$ for all positive integers j. Suppose that we have found $i(1), i(2), \ldots i(k-1)$ such that $|s_{i(j)} - c| < 1/j$ for $j = 1, 2, \ldots, k-1$. We can find some positive integer m with the properties that $m > i(k-1)$ and $a_m < c + 1/k$ (since (a_m) is a decreasing sequence of real numbers which converges to c). Thus if $i \geq m$ then $s_i < c + 1/k$. But $c - 1/k$ is not an upper bound for the set S_m (because $c - 1/k < a_m$ and a_m is the *least* upper bound of S_m). Therefore there must exist some $i(k) \geq m$ such that $s_{i(k)} > c - 1/k$. Then $|s_{i(k)} - c| < 1/k$. In this way we can construct, by induction on j, a subsequence $(s_{i(j)} : j \in \mathbb{N})$ of the original sequence with the property that $|s_{i(j)} - c| < 1/j$ for all j. Then $c = \lim_{j \to +\infty} s_{i(j)}$, as required. ∎

Remark Given a sequence $(s_i : i \in \mathbb{N})$ of real numbers that is bounded above and below, we denote by $\limsup_{i \to +\infty} s_i$ and $\liminf_{i \to +\infty} s_i$ the quantities defined by

$$\begin{aligned} \limsup_{i \to +\infty} &= \lim_{m \to +\infty} \sup\{s_i : i \geq m\}, \\ \liminf_{i \to +\infty} &= \lim_{m \to +\infty} \inf\{s_i : i \geq m\}. \end{aligned}$$

The proof of the Bolzano-Weierstrass theorem given above shows that the sequence (s_i) has a subsequence converging to $\limsup_{i\to+\infty} s_i$. A similar proof shows that it has a subsequence converging to $\liminf_{i\to+\infty} s_i$.

We can generalize the Bolzano-Weierstrass Theorem to n-dimensional Euclidean space $\mathbb{R}^n$. Thus let $(\mathbf{x}_i : i \in \mathbb{N})$ be a sequence in $\mathbb{R}^n$. We say that this sequence is *bounded* if there exists a positive constant C such that $|\mathbf{x}_i| \leq C$ for all $i \in \mathbb{N}$. We shall show that bounded sequences in $\mathbb{R}^n$ possess convergent subsequences. First we prove a lemma which shows that if a sequence in $\mathbb{R}^n$ converges 'componentwise' then it converges in $\mathbb{R}^n$.

Lemma 2.3 *Let $(\mathbf{x}_i : i \in \mathbb{N})$ be a sequence of points in n-dimensional Euclidean space $\mathbb{R}^n$. Suppose that the jth components of the elements of this sequence form a sequence of real numbers converging to some $c_j \in \mathbb{R}$. Then the sequence $(\mathbf{x}_i)$ converges to the point $\mathbf{c}$ of $\mathbb{R}^n$, where $\mathbf{c} = (c_1, c_2, \ldots, c_n)$.*

Proof Let us denote by $x_i^{(j)}$ the jth component of $\mathbf{x}_i$ for $j = 1, 2, \ldots, n$. Let $\varepsilon > 0$ be given. Then there exist positive integers $N_1, N_2, \ldots, N_n$ such that

$$|x_i^{(j)} - c_j| < \frac{\varepsilon}{\sqrt{n}}$$

for all i satisfying $i \geq N_j$. Let N be the maximum of $N_1, N_2, \ldots, N_n$. Then

$$|\mathbf{x}_i - \mathbf{c}|^2 = \sum_{j=1}^{n} |x_i^{(j)} - c_j|^2,$$

and hence $|\mathbf{x}_i - \mathbf{c}| < \varepsilon$ for all integers i satisfying $i \geq N$. This shows that the sequence $(\mathbf{x}_i : i \in \mathbb{N})$ converges to $\mathbf{c}$, as required. ∎

The following theorem generalizes the Bolzano-Weierstrass theorem to sequences in $\mathbb{R}^n$.

Theorem 2.4 *Every bounded sequence of points in n-dimensional Euclidean space $\mathbb{R}^n$ has a convergent subsequence.*

Proof Let $(\mathbf{x}_i)$ be a bounded sequence in $\mathbb{R}^n$. By the Bolzano-Weierstrass Theorem, applied to the sequence of real numbers represented by the 1st components of the elements of the sequence $(\mathbf{x}_i)$, there exists a subsequence of the sequence $(\mathbf{x}_i)$ such that the 1st components of the elements of this subsequence constitute a convergent sequence of real numbers. We apply the Bolzano-Weierstrass Theorem to the 2nd components of the elements of this subsequence to conclude that there exists a subsequence of the subsequence just constructed for which the 2nd components of the elements of this new subsequence constitute a convergent sequence of real numbers. Of course the 1st components of the elements of this new subsequence also constitute a convergent sequence of real numbers (since subsequences of convergent sequences are convergent).

We can continue in this fashion to construct the required subsequence of the original sequence. Thus suppose that, for some integer k between 2 and n, we have found a subsequence of $(\mathbf{x}_i)$ such that the jth components of the elements of this subsequence constitute a convergent sequence of real numbers for $j = 1, 2, \ldots, k-1$. We can apply the Bolzano-Weierstrass Theorem to the kth components of this subsequence to extract a further subsequence from the subsequence already found with the property that the kth components of the elements of this new subsequence constitute a convergent sequence of real numbers. We have thus constructed a subsequence of the original sequence with the property that the jth components of the elements of this subsequence constitute a convergent sequence of real numbers for $j = 1, 2, \ldots, k$. By repeatedly applying this procedure we obtain a subsequence of the original sequence such that the jth components of this subsequence constitute a convergent sequence for $j = 1, 2, \ldots, n$. This subsequence is convergent, by Lemma 2.3. ∎

We now introduce the notion of a *cluster point* of a subset S of n-dimensional Euclidean space $\mathbb{R}^n$. We say that a point $\mathbf{a}$ of $\mathbb{R}^n$ is a *cluster point* (or *limit point*) of the set S if, for every $\varepsilon > 0$, there exist an infinite number of elements $\mathbf{s}$ of the set S with the property that $|\mathbf{s} - \mathbf{a}| < \varepsilon$. Thus a point $\mathbf{a}$ of $\mathbb{R}^n$ is a cluster point of the set S if and only if every neighbourhood of the point $\mathbf{a}$ contains infinitely many of the elements of S.

Example Let S be the subset of $\mathbb{R}$ consisting of all numbers of the form $1/n$, where n is a non-zero integer. Then 0 is a cluster point for this set, since given any $\varepsilon > 0$ there exist infinitely many non-zero integers n with the property that $|1/n| < \varepsilon$. Indeed 0 is the only cluster point of the set S.

Example Let a be any real number. Then a is a cluster point of the set $\mathbb{Q}$ of *rational numbers* (i.e., numbers of the form p/q where p and q are integers and $q \neq 0$). Indeed for any $\varepsilon > 0$ there exist infinitely many rational numbers r satisfying $a - \varepsilon < r < a + \varepsilon$.

Corollary 2.5 *Let S be an infinite subset of $\mathbb{R}^n$. Suppose that S is bounded. Then S possesses a cluster point.*

Proof There exists a sequence $(\mathbf{s}_i : i \in \mathbb{N})$ of *distinct* points of S, since S is infinite. This sequence has a subsequence which converges to some point $\mathbf{c}$ of $\mathbb{R}^n$, by Theorem 2.4. Then $\mathbf{c}$ is a cluster point of the set S.

3 Continuous Functions on Closed Bounded Subsets of Euclidean Space

We shall use the n-dimensional generalization of the Bolzano-Weierstrass Theorem (Theorem 2.4) to show that continuous functions are bounded above and below on closed bounded subsets of Euclidean space. (A subset K of $\mathbb{R}^n$ is said to be *bounded* if there exists some $R > 0$ such that $|\mathbf{x}| \leq R$ for all $\mathbf{x} \in K$.)

Theorem 3.1 *Let K be a closed bounded subset of $\mathbb{R}^n$ and let $f: K \to \mathbb{R}$ be a continuous function on K. Then there exist constants A and B such that $A \leq f(\mathbf{x}) \leq B$ for all points $\mathbf{x}$ of K.*

Proof Suppose that the function f were not bounded above. Then there would exist a sequence $(\mathbf{x}_i : i \in \mathbb{N})$ of points of K such that $f(\mathbf{x}_i) \geq i$ for all positive integers i. This sequence would have a convergent subsequence $(\mathbf{x}_{i(j)} : j \in \mathbb{N})$, by Theorem 2.4. Let $\mathbf{c}$ be the limit of this convergent subsequence. The set K is closed, hence $\mathbf{c}$ is a point of K (by Lemma 1.17). Now f is continuous at $\mathbf{c}$. Applying the definition of continuity, we see that there would exist some $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{c})| < 1$ for all points $\mathbf{x}$ of K satisfying $|\mathbf{x} - \mathbf{c}| < \delta$. There would also exist some positive integer N such that $|\mathbf{x}_{i(j)} - \mathbf{c}| < \delta$ for all j satisfying $j \geq N$ (since the subsequence converges to $\mathbf{c}$). However this leads to a contradiction, for we would then have $f(\mathbf{x}_{i(j)}) < f(\mathbf{c}) + 1$ for all sufficiently large j, contradicting the fact that $f(\mathbf{x}_{i(j)}) \geq i(j)$ for all j (where $i(j) \to +\infty$ as $j \to +\infty$). This contradiction shows that f is bounded above on K. Similarly the continuous function f is bounded below on K.

The next theorem shows that continuous functions attain their bounds on closed bounded subsets of Euclidean space.

Theorem 3.2 *Let K be a closed bounded subset of $\mathbb{R}^n$ and let $f: K \to \mathbb{R}$ be a continuous function on K. Then there exist points $\mathbf{a}$ and $\mathbf{b}$ of K such that*

$$f(\mathbf{a}) \leq f(\mathbf{x}) \leq f(\mathbf{b})$$

for all $\mathbf{x} \in K$.

Proof The function f is bounded above on K, by Theorem 3.1. Define

$$c = \sup\{f(\mathbf{x}) : \mathbf{x} \in K\}.$$

We must show that there exists some point $\mathbf{b}$ of K such that $f(\mathbf{b}) = c$. Now for each positive integer i there exists a point $\mathbf{x}_i$ of K for which $c - 1/i < f(\mathbf{x}_i) \leq c$. Now the sequence $(\mathbf{x}_i)$ has a convergent subsequence $(\mathbf{x}_{i(j)} : j \in \mathbb{N})$, by Theorem 2.4. Let $\mathbf{b}$ be the limit of this subsequence. The point $\mathbf{b}$ is in K (by Lemma 1.17), since K is closed. Also $f(\mathbf{b}) = \lim_{j \to +\infty} f(\mathbf{x}_{i(j)})$, by Lemma 1.15. But $\lim_{i \to +\infty} f(\mathbf{x}_i) = c$. Hence $f(\mathbf{b}) = c$. It follows from the definition of c that $f(\mathbf{x}) \leq f(\mathbf{b})$ for all $\mathbf{x} \in K$. A similar argument shows that there exists some point $\mathbf{a}$ of K such that $f(\mathbf{x}) \geq f(\mathbf{a})$ for all $\mathbf{x} \in K$. ∎

Let D be a subset of $\mathbb{R}^n$ and let $f \colon D \to R$ be a real-valued function on D. The function f is said to be *uniformly continuous* on D if and only if, for all $\varepsilon > 0$ there exists some $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{y})| < \varepsilon$ for all points $\mathbf{x}$ and $\mathbf{y}$ in D which satisfy $|\mathbf{x} - \mathbf{y}| < \delta$, *where δ does not depend on either $\mathbf{x}$ or $\mathbf{y}$*.

Example Let D be the set $\{t \in \mathbb{R} : t > 0\}$ of positive real numbers, and let $r \colon D \to \mathbb{R}$ be the reciprocal function, defined by $r(t) = 1/t$. Then r is continuous on D (as is shown in the proof of Lemma 1.9). However r is not uniformly continuous on D. For suppose that $\varepsilon > 0$ is given. If r were uniformly continuous, then there would exist some $\delta > 0$ such that $|r(s) - r(t)| < \varepsilon$ for all positive real numbers s and t which satisfy $|s - t| < \delta$. But for any such $\delta > 0$ there exists some positive integer n with the property that $n > \varepsilon$ and $1/n < \delta$. Thus if we set $s = 1/n$ and $t = 1/2n$ then $|s - t| < \delta$, but $|r(s) - r(t)| = |n - 2n| = n$, so that $|r(s) - r(t)| \geq \varepsilon$. This shows that r is not uniformly continuous on D, even though r is continuous on D.

The next theorem shows that if a subset K of $\mathbb{R}^n$ is both bounded and closed then every continuous function on K is also uniformly continuous on K. The proof uses Theorem 2.4, which is the generalization of the Bolzano-Weierstrass Theorem to sequences in n-dimensional Euclidean space. We shall use this theorem later on in the course, when we discuss the properties of the Riemann integral. It plays a crucial role in showing that all continuous functions are Riemann-integrable.

Theorem 3.3 *Let K be a closed bounded subset of $\mathbb{R}^n$ and let $f \colon K \to \mathbb{R}$ be a continuous real-valued function on K. Then f is uniformly continuous on K.*

Proof We prove the theorem by showing that if f were not uniformly continuous on K then a contradiction would arise. Suppose therefore that f were not uniformly continuous on K. Then there would exist some $\varepsilon_0 > 0$ with the property that, for every strictly positive real number δ there would exist points $\mathbf{x}$ and $\mathbf{y}$ of K for which $|\mathbf{x} - \mathbf{y}| < \delta$ but $|f(\mathbf{x}) - f(\mathbf{y})| \geq \varepsilon_0$. In particular suppose we apply this with $\delta = 1/i$ for all positive integers i. We conclude that if f were not uniformly continuous then, for all positive integers i, there would exist points $\mathbf{x}_i$ and $\mathbf{y}_i$ of K with the property that

$$|\mathbf{x}_i - \mathbf{y}_i| < \frac{1}{i}$$

and

$$|f(\mathbf{x}) - f(\mathbf{y})| \geq \varepsilon_0.$$

Now the sequence $(\mathbf{x}_i : i \in \mathbb{N})$ is bounded, because K is a bounded subset of $\mathbb{R}^n$. We conclude from Theorem 2.4 that the sequence $(\mathbf{x}_i : i \in \mathbb{N})$ has a convergent subsequence $(\mathbf{x}_{i(j)} : j \in \mathbb{N})$. Let $\mathbf{a}$ be the limit of this subsequence. Now K is closed, therefore $\mathbf{a}$ belongs to K, by Lemma 1.17.

Now f is continuous on K, by hypothesis. In particular, f is continuous at $\mathbf{a}$. Thus, for every $\varepsilon > 0$ there exists $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{a})| < \varepsilon$ for all points $\mathbf{x} \in K$ which satisfy $|\mathbf{x} - \mathbf{a}| < \delta$. In particular, let us choose $\varepsilon = \frac{1}{2}\varepsilon_0$. We conclude that there exists some $\delta > 0$ such that

$$|f(\mathbf{x}) - f(\mathbf{a})| < \tfrac{1}{2}\varepsilon_0$$

for all points $\mathbf{x} \in K$ which satisfy $|\mathbf{x} - \mathbf{a}| < \delta$. But $\mathbf{a}$ is the limit of the sequence $(x_{i(j)} : j \in \mathbb{N})$ as j tends to $+\infty$. Thus there exists some positive integer J such that $|\mathbf{x}_{i(j)} - \mathbf{a}| < \frac{1}{2}\delta$ for all j satisfying $j \geq J$. Let us choose j sufficiently large so that $j \geq J$ and $i(j) > 2/\delta$. Then $|\mathbf{x}_{i(j)} - \mathbf{a}| < \frac{1}{2}\delta$ and

$$\begin{aligned} |\mathbf{y}_{i(j)} - \mathbf{a}| &\leq |\mathbf{y}_{i(j)} - \mathbf{x}_{i(j)}| + |\mathbf{x}_{i(j)} - \mathbf{a}| \\ &< \frac{1}{i(j)} + \tfrac{1}{2}\delta < \delta. \end{aligned}$$

Therefore

$$\begin{aligned} |f(\mathbf{x}_{i(j)}) - f(\mathbf{a})| &< \tfrac{1}{2}\varepsilon_0, \\ |f(\mathbf{y}_{i(j)}) - f(\mathbf{a})| &< \tfrac{1}{2}\varepsilon_0. \end{aligned}$$

We conclude from the triangle inequality that

$$|f(\mathbf{x}_{i(j)}) - f(\mathbf{y}_{i(j)})| < \varepsilon_0.$$

But this is a contradiction, for $\mathbf{x}_{i(j)}$ and $\mathbf{y}_{i(j)}$ were chosen so that the inequality

$$|f(\mathbf{x}_{i(j)}) - f(\mathbf{y}_{i(j)})| \geq \varepsilon_0$$

is satisfied. From this contradiction we conclude that f must be uniformly continuous on K. ∎

4 The Riemann Integral

Let a and b be real numbers satisfying $a < b$ and let $f\colon [a, b] \to \mathbb{R}$ be a real-valued function defined on $[a, b]$ which is bounded above and below, so that there exist real numbers m and M with the property that $m \leq f(t) \leq M$ for all $t \in [a, b]$.

We define a *partition* P of $[a, b]$ to be a set $\{t_0, t_1, \ldots, t_n\}$ of real numbers satisfying

$$a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b.$$

Given such a partition P of $[a, b]$, we define the quantities $L(P, f)$ and $U(P, f)$ by

$$L(P, f) = \sum_{i=1}^{n} m_i (t_i - t_{i-1}),$$

$$U(P, f) = \sum_{i=1}^{n} M_i (t_i - t_{i-1}),$$

where

$$m_i = \inf\{f(t) : t_{i-1} \leq t \leq t_i\},$$

$$M_i = \sup\{f(t) : t_{i-1} \leq t \leq t_i\}$$

(so that m_i is the greatest lower bound on the values of f on the interval $[t_{i-1}, t_i]$ and M_i is the least upper bound on the values of f on this interval). Clearly $L(P, f) \leq U(P, f)$.

Suppose that $m \leq f(t) \leq M$ for all $t \in [a, b]$. Then $m_i \geq m$ for all integers i between 1 and n, hence

$$L(P, f) \geq m \sum_{i=1}^{n} (t_i - t_{i-1}) = m(b - a),$$

Similarly $U(P, f) \leq M(b - a)$. Thus

$$m(b - a) \leq L(P, f) \leq U(P, f) \leq M(b - a)$$

for all partitions P of $[a, b]$.

Definition Let a and b be real numbers satisfying $a < b$ and let $f\colon [a,b] \to \mathbb{R}$ be a real-valued function on $[a,b]$ which is bounded above and below. Define the *upper Riemann integral*

$$\mathcal{U}\int_a^b f(t)\,dt$$

and the *lower Riemann integral*

$$\mathcal{L}\int_a^b f(t)\,dt$$

of the function f on $[a,b]$ by

$$\begin{aligned} \mathcal{U}\int_a^b f(t)\,dt &= \inf\{U(P,f) : P \text{ is a partition of } [a,b]\}, \\ \mathcal{L}\int_a^b f(t)\,dt &= \sup\{L(P,f) : P \text{ is a partition of } [a,b]\}. \end{aligned}$$

If the upper Riemann integral of f on $[a,b]$ is equal to the lower Riemann integral of f on $[a,b]$ then f is said to be *Riemann-integrable* on $[a,b]$, and the *Riemann integral*

$$\int_a^b f(t)\,dt$$

of f on $[a,b]$ is defined to be the common value of the upper and lower Riemann integrals of f.

Observe that, for every bounded function f on $[a,b]$, the upper and lower Riemann integrals of the function f are well-defined. However such a function need not be Riemann-integrable on $[a,b]$.

Let $f\colon [a,b] \to \mathbb{R}$ be a bounded real-valued function on $[a,b]$. We shall prove that

$$\mathcal{L}\int_a^b f(t)\,dt \le \mathcal{U}\int_a^b f(t)\,dt.$$

In order to prove this fact, we introduce the notion of a *refinement* of a partition, and we show that if a partition Q of $[a,b]$ is a refinement of a partition P, then $L(Q,f) \ge L(P,f)$ and $U(Q,f) \le U(P,f)$.

Let P and Q be partitions of $[a,b]$, given by $P = \{t_0, t_1, \ldots, t_n\}$ and $Q = \{s_0, s_1, \ldots, s_m\}$. We say that the partition Q is a *refinement* of P if the set P is contained in the set Q, so that for every t_i in P there exists some s_j in Q such that $t_i = s_j$.

Lemma 4.1 *Let a and b be real numbers satisfying $a < b$. Let $f\colon [a,b] \to \mathbb{R}$ be a real-valued function on $[a,b]$ which is bounded above and below on $[a,b]$. Let P be a partition of $[a,b]$ and let R be a refinement of P. Then*

$$L(R,f) \geq L(P,f),$$

$$U(R,f) \leq U(P,f).$$

Proof First we establish notation to be used in the proof. Let the partitions P and R be given by

$$P = \{t_0, t_1, \ldots, t_{n-1}, t_n\},$$

$$R = \{s_0, s_1, \ldots, s_{m-1}, s_m\},$$

where

$$a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b,$$

$$a = s_0 < s_1 < \cdots < s_{m-1} < s_m = b.$$

Now for each $t_i \in P$ there exists some $s_j \in R$ such that $t_i = s_j$. Given i between 1 and n, let us define $j(i)$ so that $t_i = s_{j(i)}$. Note that $j(0) = 0$ and $j(n) = m$. Let us define

$$\begin{aligned} m_i[P,f] &= \inf\{f(t) : t_{i-1} \leq t \leq t_i\}, \\ M_i[P,f] &= \sup\{f(t) : t_{i-1} \leq t \leq t_i\}, \\ m_j[R,f] &= \inf\{f(t) : s_{j-1} \leq t \leq s_j\}, \\ M_j[R,f] &= \sup\{f(t) : s_{j-1} \leq t \leq s_j\}, \end{aligned}$$

so that

$$\begin{aligned} L(P,f) &= \sum_{i=1}^{n} m_i[P,f](t_i - t_{i-1}), \\ U(P,f) &= \sum_{i=1}^{n} M_i[P,f](t_i - t_{i-1}), \\ L(R,f) &= \sum_{j=1}^{m} m_j[R,f](s_j - s_{j-1}), \\ U(R,f) &= \sum_{j=1}^{m} M_j[R,f](s_j - s_{j-1}), \end{aligned}$$

Note that

$$L(R,f) = \sum_{i=1}^{n} \left(\sum_{j=j(i-1)+1}^{j(i)} m_j[R,f](s_j - s_{j-1}) \right).$$

Now if $j(i-1)+1 \leq j \leq j(i)$ then

$$t_{i-1} \leq s_{j-1} < s_j \leq t_i,$$

(because $t_{i-1} = s_{j(i-1)}$ and $t_i = s_{j(i)}$) and so $m_j[R,f] \geq m_i[P,f]$ (because $m_i[R,f]$ is the infimum of f on the closed interval $[t_{i-1}, t_i]$ and $m_j[R,f]$ is the infimum of f on the closed interval $[s_{j-1}, s_j]$, and this interval is contained in the closed interval $[t_{i-1}, t_i]$). Therefore

$$\begin{aligned} \sum_{j=j(i-1)+1}^{j(i)} m_j[R,f](s_j - s_{j-1}) \quad &\geq \quad m_i[P,f] \sum_{j=j(i-1)+1}^{j(i)} (s_j - s_{j-1}) \\ &= m_i[P,f](s_{j(i)} - s_{j(i-1)}) \\ &= m_i[P,f](t_i - t_{i-1}). \end{aligned}$$

Therefore

$$L(R,f) \geq \sum_{i=1}^{n} m_i[P,f](t_i - t_{i-1}) = L(P,f).$$

An analogous argument shows that $U(R,f) \leq U(P,f)$. For

$$U(R,f) = \sum_{i=1}^{n} \left(\sum_{j=j(i-1)+1}^{j(i)} M_j[R,f](s_j - s_{j-1}) \right).$$

But if $j(i-1)+1 \leq j \leq j(i)$ then

$$t_{i-1} \leq s_{j-1} < s_j \leq t_i,$$

and so $M_j[R,f] \leq M_i[P,f]$, and hence

$$\begin{aligned} \sum_{j=j(i-1)+1}^{j(i)} M_j[R,f](s_j - s_{j-1}) \quad &\leq \quad M_i[P,f] \sum_{j=j(i-1)+1}^{j(i)} (s_j - s_{j-1}) \\ &= M_i[P,f](s_{j(i)} - s_{j(i-1)}) \\ &= M_i[P,f](t_i - t_{i-1}). \end{aligned}$$

Therefore

$$U(R,f) \leq \sum_{i=1}^{n} M_i[P,f](t_i - t_{i-1}) = U(P,f). \quad \blacksquare$$

Given any two partitions P and Q of $[a, b]$ there exists a partition R of $[a, b]$ which is a refinement of both P and Q. For example, we can take R to be the partition of $[a, b]$ consisting of all of the elements of the union $P \cup Q$ of P and Q, ordered in increasing order. Such a partition is said to be a *common refinement* of the partitions P and Q.

Let $f: [a, b] \to \mathbb{R}$ be a real-valued function that is bounded above and below on $[a, b]$. Let P and Q be partitions of $[a, b]$. Let R be a common refinement of the partitions P and Q. On applying Lemma 4.1 we see that

$$L(P, f) \leq L(R, f) \leq U(R, f) \leq U(Q, f),$$

We have therefore shown that $L(P, f) \leq U(Q, f)$ for all partitions P and Q of $[a, b]$. We use this fact in the proof of the following lemma.

Lemma 4.2 *Let $f: [a, b] \to \mathbb{R}$ be a real-valued function that is bounded above and below on $[a, b]$. Then*

$$\mathcal{L}\int_a^b f(t)\,dt \leq \mathcal{U}\int_a^b f(t)\,dt.$$

Proof Let P and Q be partitions of $[a, b]$. Then

$$L(P, f) \leq U(Q, f),$$

by Corollary 4.2. Taking the supremum of the left hand side of this inequality as P ranges over all partitions of $[a, b]$ we see that

$$\mathcal{L}\int_a^b f(t)\,dt \leq U(Q, f)$$

for all partitions Q of $[a, b]$. Taking the infimum of the right hand side of this inequality as Q ranges over all partitions of $[a, b]$ we see that

$$\mathcal{L}\int_a^b f(t)\,dt \leq \mathcal{U}\int_a^b f(t)\,dt,$$

as required. ∎

Remark Let us for the moment consider the consider when f is non-negative and bounded. Let us consider the area of the region R in the Euclidean plane defined by

$$R = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, \quad 0 \leq y \leq f(x)\},$$

where a and b are real numbers satisfying $a < b$. Thus R is the region 'under the graph of the function'. We suppose that f is 'regular enough' to ensure that this area is indeed well-defined. We now explain why

$$\text{area of } R = \int_a^b f(x)\,dx.$$

Let P be a partition of $[a, b]$ given by $P = \{t_0, t_1, \ldots, t_n\}$, where

$$a = t_0 < t_1 < \cdots < t_n = b$$

Consider the regions R_1 and R_2 defined by

$$\begin{aligned} R_1 &= \bigcup_{i=1}^{n} \{(x, y) \in \mathbb{R}^2 : t_{i-1} \le x \le t_i, 0 \le y \le m_i\}, \\ R_2 &= \bigcup_{i=1}^{n} \{(x, y) \in \mathbb{R}^2 : t_{i-1} \le x \le t_i, 0 \le y \le M_i\}, \end{aligned}$$

where

$$\begin{aligned} m_i &= \inf\{f(x) : t_{i-1} \le x \le t_i\}, \\ M_i &= \sup\{f(x) : t_{i-1} \le x \le t_i\}. \end{aligned}$$

Note that $R_1 \subset R \subset R_2$. Also R_1 and R_2 are unions of rectangles. Let $A(R)$, $A(R_1)$ and $A(R_2)$ denote the areas of R, R_1 and R_2 respectively. Then

$$\begin{aligned} A(R_1) &= \sum_{i=1}^{n} m_i(t_i - t_{i-1}) = L(P, f), \\ A(R_2) &= \sum_{i=1}^{n} M_i(t_i - t_{i-1}) = U(P, f), \end{aligned}$$

Now if the area $A(R)$ of R is well-defined then clearly $A(R_1) \le A(R) \le A(R_2)$, since $R_1 \subset R \subset R_2$. Thus

$$L(P, f) \le \int_a^b f(x)\,dx \le U(P, f).$$

If we take the supremum of the left hand side of this inequality over all partitions P of $[a, b]$, and if we take the infimum of the right hand side of this inequality over all partitions P of $[a, b]$ we deduce that

$$\mathcal{L}\int_a^b f(t)\,dt \le A(R) \le \mathcal{U}\int_a^b f(t)\,dt.$$

We conclude therefore that if f is bounded and non-negative on $[a, b]$, if the area $A(R)$ of the region R defined above is well-defined and if f is Riemann-integrable on $[a, b]$ then

$$A(R) = \int_a^b f(t)\, dt.$$

Thus the integral of a function measures the area 'under the graph of the function' if the function is bounded, non-negative and Riemann-integrable. Similarly if the function is bounded, non-positive and Riemann-integrable on $[a, b]$ then

$$A(R') = -\int_a^b f(x)\, dx,$$

where $A(R')$ is the area of the region R' 'over the graph of the function' defined by

$$R' = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, \quad f(x) \leq y \leq 0\}.$$

Theorem 4.3 *Let $f\colon [a, b] \to \mathbb{R}$ be a bounded real-valued function on $[a, b]$. Then f is Riemann-integrable on $[a, b]$ if and only if, for every $\varepsilon > 0$ there exists a partition P of $[a, b]$ for which*

$$U(P, f) - L(P, f) < \varepsilon.$$

Proof Suppose that f is Riemann-integrable on $[a, b]$. Let $\varepsilon > 0$ be any positive real number. Then there exists a partition Q of $[a, b]$ such that

$$\int_a^b f(t)\, dt - L(Q, f) < \tfrac{1}{2}\varepsilon,$$

since the Riemann integral of f on $[a, b]$ is equal to the lower Riemann integral on $[a, b]$ and hence

$$\int_a^b f(t)\, dt = \sup \{L(Q, f) : Q \text{ is a partition of } [a, b]\}.$$

Similarly there exists a partition R of $[a, b]$ such that

$$U(R, f) - \int_a^b f(t)\, dt < \tfrac{1}{2}\varepsilon,$$

since

$$\int_a^b f(t)\, dt = \inf \{U(R, f) : R \text{ is a partition of } [a, b]\}.$$

Then $U(R,f) - L(Q,f) < \varepsilon$. Let the partition P of $[a,b]$ be a common refinement of the partitions Q and R. Using Lemma 4.1 we see that

$$L(Q,f) \leq L(P,f) \leq U(P,f) \leq U(R,f),$$

and hence $U(P,f) - L(P,f) < \varepsilon$. This shows that if f is Riemann-integrable then, given any $\varepsilon > 0$, there exists a partition P of $[a,b]$ such that $U(P,f) - L(P,f) < \varepsilon$.

Conversely, let $f\colon [a,b] \to \mathbb{R}$ be a bounded real-valued function on $[a,b]$ with the property that, given any $\varepsilon > 0$, there exists a partition P of $[a,b]$ such that $U(P,f) - L(P,f) < \varepsilon$. We must show that f is Riemann-integrable on $[a,b]$. Now

$$L(P,f) \leq \mathcal{L}\int_a^b f(t)\,dt \leq \mathcal{U}\int_a^b f(t)\,dt \leq U(P,f)$$

for all partitions P of $[a,b]$. Therefore we conclude that

$$\mathcal{U}\int_a^b f(t)\,dt - \mathcal{L}\int_a^b f(t)\,dt < \varepsilon$$

for all $\varepsilon > 0$. But this implies that

$$\mathcal{U}\int_a^b f(t)\,dt - \mathcal{L}\int_a^b f(t)\,dt \leq 0.$$

But we have already shown that

$$\mathcal{U}\int_a^b f(t)\,dt \geq \mathcal{L}\int_a^b f(t)\,dt.$$

Hence

$$\mathcal{U}\int_a^b f(t)\,dt = \mathcal{L}\int_a^b f(t)\,dt,$$

and thus f is Riemann-integrable on $[a,b]$, as required. ∎

Corollary 4.4 *Let $f\colon [a,b] \to \mathbb{R}$ be a bounded real-valued function on $[a,b]$. Suppose that f is Riemann-integrable on $[a,b]$. Then for every $\varepsilon > 0$ there exists a partition P of $[a,b]$, where $P = \{t_0, t_1, \ldots, t_n\}$, such that*

$$\left| \int_a^b f\,dt - \sum_{i=1}^{n} f(x_i)(t_i - t_{i-1}) \right| < \varepsilon$$

for all collections $\{x_1, x_2, \ldots, x_n\}$ of points which satisfy the inequalities $t_{i-1} \leq x_i \leq t_i$.

Proof Given any $\varepsilon > 0$ there exists a partition P of $[a, b]$ such that $U(P, f) - L(P, f) < \varepsilon$ (by Theorem 4.3). But

$$L(P, f) = \sum_{i=1}^{n} m_i (t_i - t_{i-1}),$$

$$U(P, f) = \sum_{i=1}^{n} M_i (t_i - t_{i-1}),$$

where

$$m_i = \inf\{f(t) : t_{i-1} \le t \le t_i\},$$
$$M_i = \sup\{f(t) : t_{i-1} \le t \le t_i\}.$$

Thus if $t_{i-1} \le x_i \le t_i$ then $m_i \le f(x_i) \le M_i$, and hence

$$L(P, f) \le \sum_{i=1}^{n} f(x_i)(t_i - t_{i-1}) \le U(P, f).$$

But

$$L(P, f) \le \int_a^b f(t)\,dt \le U(P, f)$$

and $U(P, f) - L(P, f) < \varepsilon$. The required result follows directly from these inequalities. ∎

Example Let f be the function defined by $f(t) = ct + d$, where c and d are constants. For simplicity, we restrict our attention to the case when $c \ge 0$. We shall show that f is Riemann-integrable on $[0, 1]$ and evaluate

$$\int_0^1 f(t)\,dt$$

from first principles. In order to show that f is Riemann-integrable, it suffices to show that, given any $\varepsilon > 0$, there exists some partition P of $[0, 1]$ such that $U(P, f) - L(P, f) < \varepsilon$. In order to accomplish this we shall consider the partition P_n of $[0, 1]$, where n is a positive integer and P_n is a partition of $[0, 1]$ into n subintervals of equal length. We shall show that

$$\lim_{n \to +\infty} (U(P, f) - L(P, f)) = 0,$$

so that, given any $\varepsilon > 0$, $U(P_n, f) - L(P_n, f) < \varepsilon$ for all sufficiently large n.

Now $P_n = \{t_0, t_1, \ldots, t_n\}$, where $t_i = i/n$. Now f takes values between $(i-1)c/n + d$ and $ic/n + d$ on the interval $[t_{i-1}, t_i]$. We are considering the case when $c \geq 0$. Thus if

$$\begin{aligned} m_i &= \inf\{f(t) : t_{i-1} \leq t \leq t_i\}, \\ M_i &= \sup\{f(t) : t_{i-1} \leq t \leq t_i\}, \end{aligned}$$

then

$$m_i = \frac{(i-1)c}{n} + d, \qquad M_i = \frac{ic}{n} + d,$$

hence

$$\begin{aligned} L(P_n, f) &= \sum_{i=1}^{n} m_i(t_i - t_{i-1}) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{ci}{n} + d - \frac{c}{n} \right) \\ &= \frac{c(n+1)}{2n} + d - \frac{c}{n} \\ &= \frac{c}{2} + d - \frac{c}{2n}, \\ U(P_n, f) &= \sum_{i=1}^{n} M_i(t_i - t_{i-1}) \\ &= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{ci}{n} + d \right) \\ &= \frac{c(n+1)}{2n} + d \\ &= \frac{c}{2} + d + \frac{c}{2n}. \end{aligned}$$

Thus $U(P_n, f) - L(P_n, f) = c/n$. We conclude that, given any $\varepsilon > 0$, $U(P_n, f) - L(P_n, f) < \varepsilon$ for all sufficiently large n. Therefore f is Riemann-integrable on $[0, 1]$. Moreover

$$L(P_n, f) \leq \int_0^1 f(t)\,dt \leq U(P_n, f)$$

for all positive integers n. Thus

$$\frac{c}{2} + d - \frac{c}{2n} \leq \int_0^1 f(t)\,dt \leq \frac{c}{2} + d + \frac{c}{2n}$$

for all positive integers n. Therefore we must have

$$\int_0^1 f(t)\,dt = \frac{c}{2} + d.$$

One can use a similar argument to prove this result also for the case when $c \leq 0$.

The above example illustrates a useful technique when one is attempting to show that a function is Riemann-integrable and evaluate the integral from first principles. In order to show from first principles that a function f is Riemann-integrable on some interval [a,b], where f is bounded on $[a, b]$, it suffices to show that, given any $\varepsilon > 0$, there exists a partition P of $[a, b]$ such that $U(P, f) - L(P, f) < \varepsilon$, by Theorem 4.3. In order to find such a partition it is often sufficient to consider partitions P_n for all positive integers n, where P_n is a partition of the interval $[a, b]$ into n intervals of length $(b-a)/n$. (Thus $P_n = \{t_0, t_1, \ldots, t_n\}$, where $t_i = ((n-i)a + ib)/n$.) If one can show that

$$\lim_{n \to +\infty} (U(P_n, f) - L(P_n, f)) = 0$$

then one can deduce from this that the function f is Riemann-integrable on $[a, b]$. Moreover one knows that

$$L(P_n, f) \leq \int_a^b f(t)\,dt \leq U(P_n, f)$$

for all positive integers n, hence

$$\lim_{n \to +\infty} L(P_n, f) \leq \int_a^b f(t)\,dt \leq \lim_{n \to +\infty} U(P_n, f).$$

This enables us to calculate the Riemann integral of f on $[a, b]$ from first principles.

We now prove that sums and scalar multiples of Riemann-integrable functions are Riemann-integrable.

Theorem 4.5 *Let a and b be real numbers satisfying $a < b$ and let $f\colon [a, b] \to \mathbb{R}$ and $g\colon [a, b] \to \mathbb{R}$ be bounded Riemann-integrable functions on $[a, b]$. Let c be a real number. Then the functions $f + g$ and cf are Riemann-integrable on $[a, b]$, and*

$$\begin{aligned} \int_a^b (f(t) + g(t))\,dt &= \int_a^b f(t)\,dt + \int_a^b g(t)\,dt, \\ \int_a^b (cf(t))\,dt &= c\int_a^b f(t)\,dt. \end{aligned}$$

Proof Let R be a partition $\{t_0, t_1, \ldots, t_n\}$ of $[a, b]$, where

$$a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b.$$

Given any bounded real-valued function h on $[a, b]$ we define

$$\begin{aligned} m_i[R, h] &= \inf\{h(t) : t_{i-1} \le t \le t_i\}, \\ M_i[R, h] &= \sup\{h(t) : t_{i-1} \le t \le t_i\}, \end{aligned}$$

so that

$$\begin{aligned} L(R, h) &= \sum_{i=1}^{n} m_i[R, h](t_i - t_{i-1}), \\ U(R, h) &= \sum_{i=1}^{n} M_i[R, h](t_i - t_{i-1}). \end{aligned}$$

for any bounded real-valued function h on $[a, b]$.

If $t_{i-1} \le t \le t_i$ then $f(t) \ge m_i[R, f]$ and $g(t) \ge m_i[R, g]$, hence

$$f(t) + g(t) \ge m_i[R, f] + m_i[R, g].$$

for all $t \in [t_{i-1}, t_i]$. It follows from the definition of $m_i[R, f + g]$ that

$$m_i[R, f + g] \ge m_i[R, f] + m_i[R, g].$$

A similar argument shows that

$$M_i[R, f + g] \le M_i[R, f] + M_i[R, g].$$

We conclude from the definitions of $L(R, f + g)$ and $U(R, f + g)$ that

$$\begin{aligned} L(R, f + g) &\ge L(R, f) + L(R, g), \\ U(R, f + g) &\le U(R, f) + U(R, g), \end{aligned}$$

Let $\varepsilon > 0$ be given. Now it follows from the fact that f and g are Riemann-integrable that there exist partitions P and Q of $[a, b]$ such that

$$\begin{aligned} L(P, f) &> \int_a^b f(t)\,dt - \tfrac{1}{2}\varepsilon, \\ L(Q, g) &> \int_a^b g(t)\,dt - \tfrac{1}{2}\varepsilon. \end{aligned}$$

Now there exists a partition R of $[a, b]$ which is a common refinement of P and Q. Then $L(R, f) \geq L(P, f)$ and $L(R, g) \geq L(Q, g)$, by Lemma 4.1, hence

$$\begin{aligned} L(R, f+g) &\geq L(R, f) + L(R, g) \\ &\geq L(P, f) + L(Q, g) \\ &> \int_a^b f(t)\,dt + \int_a^b g(t)\,dt - \varepsilon. \end{aligned}$$

But

$$\mathcal{L}\int_a^b (f(t) + g(t))\,dt \geq L(R, f+g),$$

hence

$$\mathcal{L}\int_a^b (f(t) + g(t))\,dt > \int_a^b f(t)\,dt + \int_a^b g(t)\,dt - \varepsilon$$

for all $\varepsilon > 0$, hence

$$\mathcal{L}\int_a^b (f(t) + g(t))\,dt \geq \int_a^b f(t)\,dt + \int_a^b g(t)\,dt.$$

Similarly, given any $\varepsilon > 0$ we can find partitions P and Q of $[a, b]$ such that

$$\begin{aligned} U(P, f) &< \int_a^b f(t)\,dt + \tfrac{1}{2}\varepsilon, \\ U(Q, g) &< \int_a^b g(t)\,dt + \tfrac{1}{2}\varepsilon. \end{aligned}$$

Let R be a partition of $[a, b]$ which is a common refinement of P and Q. Using Lemma 4.1 we see that

$$\begin{aligned} \mathcal{U}\int_a^b (f(t) + g(t))\,dt &\leq U(R, f+g) \\ &\leq U(R, f) + U(R, g) \\ &\leq U(P, f) + U(Q, g) \\ &< \int_a^b f(t)\,dt + \int_a^b g(t)\,dt + \varepsilon. \end{aligned}$$

Since this inequality holds for all $\varepsilon > 0$, we conclude that

$$\mathcal{U}\int_a^b (f(t) + g(t))\,dt \leq \int_a^b f(t)\,dt + \int_a^b g(t)\,dt.$$

But

$$\mathcal{L}\int_a^b (f(t)+g(t))\,dt \leq \mathcal{U}\int_a^b (f(t)+g(t))\,dt,$$

hence

$$\mathcal{L}\int_a^b (f(t)+g(t))\,dt = \mathcal{U}\int_a^b (f(t)+g(t))\,dt = \int_a^b f(t)\,dt + \int_a^b g(t)\,dt.$$

We conclude that $f+g$ is Riemann-integrable and

$$\int_a^b (f(t)+g(t))\,dt = \int_a^b f(t)\,dt + \int_a^b g(t)\,dt.$$

Finally we show that cf is Riemann-integrable on $[a,b]$ for all real numbers c, and

$$\int_a^b cf(t)\,dt = c\int_a^b f(t)\,dt.$$

This follows immediately from the fact that if P is any partition of $[a,b]$ then $L(P,cf) = cL(P,f)$ and $U(P,cf) = cU(P,f)$ (by definition of $L(P,cf)$ and $U(P,cf)$), so that

$$\begin{aligned} \mathcal{L}\int_a^b cf(t)\,dt &= c\mathcal{L}\int_a^b f(t)\,dt, \\ \mathcal{U}\int_a^b cf(t)\,dt &= c\mathcal{U}\int_a^b f(t)\,dt. \end{aligned}$$ ∎

Theorem 4.6 *Let a, b and c be real numbers satisfying $a < b < c$. Let $f\colon [a,c] \to \mathbb{R}$ be a bounded real-valued function on $[a,c]$. Suppose that f is Riemann-integrable on the intervals $[a,b]$ and $[b,c]$. Then f is Riemann-integrable on $[a,c]$, and*

$$\int_a^c f(t)\,dt = \int_a^b f(t)\,dt + \int_b^c f(t)\,dt.$$

Proof We now establish the notation to be used in the proof. Given real numbers u and v satisfying $u < v$, where f is defined and bounded on $[u,v]$, and given a *partition* P of $[u,v]$ with $P = \{t_0, t_1, \ldots, t_n\}$, where

$$u = t_0 < t_1 < \cdots < t_{n-1} < t_n = v,$$

we define

$$L(P,f,u,v) = \sum_{i=1}^{n} m_i(t_i - t_{i-1}),$$

$$U(P, f, u, v) = \sum_{i=1}^{n} M_i(t_i - t_{i-1}),$$

where

$$m_i = \inf\{f(t) : t_{i-1} \leq t \leq t_i\},$$
$$M_i = \sup\{f(t) : t_{i-1} \leq t \leq t_i\}$$

(i.e., the quantities $L(P, f, u, v)$ and $U(P, f, u, v)$ are the quantities that we denoted previously by $L(P, f)$ and $U(P, f)$ respectively, but they are denoted in this proof by $L(P, f, u, v)$ and $U(P, f, u, v)$ in order to emphasize the dependence of these quantities on the endpoints of the chosen interval $[u, v]$.

Let $\varepsilon > 0$ be given. We show that there exists a partition P of $[a, c]$ such that $U(P, f, a, c) - L(P, f, a, c) < \varepsilon$. Given this fact, it then follows from Theorem 4.3 that f is Riemann-integrable on $[a, c]$.

Now f is Riemann-integrable on the intervals $[a, b]$ and $[b, c]$, hence there exist partitions Q_1 and Q_2 of $[a, b]$ and partitions R_1 and R_2 of $[b, c]$ with the property that

$$\begin{aligned} L(Q_1, f, a, b) &> \int_a^b f(t)\,dt - \frac{\varepsilon}{4}, \\ U(Q_2, f, a, b) &< \int_a^b f(t)\,dt + \frac{\varepsilon}{4}, \\ L(R_1, f, b, c) &> \int_b^c f(t)\,dt - \frac{\varepsilon}{4}, \\ U(R_2, f, b, c) &< \int_b^c f(t)\,dt + \frac{\varepsilon}{4}. \end{aligned}$$

Let Q be a common refinement of the partitions Q_1 and Q_2 and let R be a common refinement of the partitions R_1 and R_2. Then

$$\int_a^b f(t)\,dt - \frac{\varepsilon}{4} < L(Q, f, a, b) < U(Q, f, a, b) < \int_a^b f(t)\,dt + \frac{\varepsilon}{4},$$
$$\int_b^c f(t)\,dt - \frac{\varepsilon}{4} < L(R, f, b, c) < U(R, f, b, c) < \int_b^c f(t)\,dt + \frac{\varepsilon}{4},$$

by Lemma 4.1. Let P be the partition of $[a, c]$ consisting of all points of the partitions Q and R. Thus if $Q = \{t_0, t_1, \ldots, t_n\}$ and $R = \{s_0, s_1, \ldots, s_m\}$, where

$$a = t_0 < t_1 < \cdots < t_n = b, \qquad b = s_0 < s_1 < \cdots < s_m = c,$$

then $P = \{u_0, u_1, \ldots, u_{n+m}\}$, where

$$u_i = \begin{cases} t_i & \text{if } 0 \leq i \leq n; \\ s_{i-n} & \text{if } n \leq i \leq n+m. \end{cases}$$

Then

$$\begin{aligned} L(P, f, a, c) &= L(Q, f, a, b) + L(Q, f, b, c), \\ U(P, f, a, c) &= U(Q, f, a, b) + U(Q, f, b, c), \end{aligned}$$

hence

$$\begin{aligned} L(P, f, a, c) &> \int_a^b f(t)\,dt + \int_b^c f(t)\,dt - \frac{\varepsilon}{2}, \\ U(P, f, a, c) &< \int_a^b f(t)\,dt + \int_b^c f(t)\,dt + \frac{\varepsilon}{2}, \end{aligned}$$

Thus $U(P, f, a, c) - L(P, f, a, c) < \varepsilon$. It follows from Theorem 4.3 that f is Riemann-integrable on $[a, c]$. Moreover

$$L(P, f, a, c) \leq \int_a^c f(t)\,dt \leq U(P, f, a, c),$$

hence

$$\left| \int_a^c f(t)\,dt - \int_a^b f(t)\,dt - \int_b^c f(t)\,dt \right| < \frac{\varepsilon}{2}.$$

But this inequality must hold for *every* $\varepsilon > 0$. Hence

$$\int_a^c f(t)\,dt = \int_a^b f(t)\,dt + \int_b^c f(t)\,dt. \quad \blacksquare$$

We now show that continuous functions are Riemann-integrable.

Theorem 4.7 *Let a and b be real numbers satisfying $a < b$ and let $f\colon [a, b] \to \mathbb{R}$ be a continuous function on the interval $[a, b]$. Then f is Riemann-integrable on $[a, b]$.*

Proof The closed interval $[a, b]$ is a closed bounded set, and continuous functions are bounded above and below on closed bounded sets, by Theorem 3.2. Thus f is bounded above and below on $[a, b]$. We show that, given there exists a partition P of $[a, b]$ such that $U(P, f) - L(P, f) < \varepsilon$. It will then follow from Theorem 4.3 that f is Riemann-integrable on $[a, b]$. (As usual, the quantities $L(P, f)$ and $U(P, f)$ are defined by

$$L(P, f) = \sum_{i=1}^{n} m_i(t_i - t_{i-1}),$$

$$U(P, f) = \sum_{i=1}^{n} M_i(t_i - t_{i-1}),$$

where

$$m_i = \inf\{f(t) : t_{i-1} \leq t \leq t_i\},$$
$$M_i = \sup\{f(t) : t_{i-1} \leq t \leq t_i\}.)$$

Now f is uniformly continuous on the closed bounded interval $[a, b]$, by Theorem 3.3. Thus, given $\varepsilon > 0$, there exists some $\delta > 0$ such that

$$|f(x) - f(y)| < \frac{\varepsilon}{2(b-a)}$$

for all $x, y \in [a, b]$ which satisfy $|x - y| < \delta$. Now chose a partition P of $[a, b]$, where $P = \{t_0, t_1, \ldots, t_n\}$, such that $|t_i - t_{i-1}| < \delta$ for $i = 1, 2, \ldots, n$ (e.g., let P be the partition of $[a, b]$ into n intervals of length $(a-b)/n$, where $n > (a - b)/\delta$). Thus if x and t belong to the interval $[t_{i-1}, t_i]$ then

$$|f(x) - f(t)| < \frac{\varepsilon}{2(b-a)}$$

so that

$$f(t) < f(x) + \frac{\varepsilon}{2(b-a)}.$$

Thus if $t_{i-1} \leq x \leq t_i$ then

$$f(x) + \frac{\varepsilon}{2(b-a)}$$

is an upper bound for the set $\{f(t) : t_{i-1} \leq t \leq t_i\}$, and hence

$$M_i \leq f(x) + \frac{\varepsilon}{2(b-a)},$$

where

$$M_i = \sup\{f(t) : t_{i-1} \leq t \leq t_i\}.)$$

But then

$$M_i - \frac{\varepsilon}{2(b-a)}$$

is a lower bound for the set $\{f(t) : t_{i-1} \leq t \leq t_i\}$, and hence

$$M_i - \frac{\varepsilon}{2(b-a)} \leq m_i,$$

where

$$m_i = \inf\{f(t) : t_{i-1} \leq t \leq t_i\}.$$

Thus

$$M_i - m_i \leq \frac{\varepsilon}{2(b-a)} \qquad (i = 1, 2, \ldots, n).$$

Hence

$$\begin{aligned} U(P,f) - L(P,f) &= \sum_{i=1}^{n} ((M_i - m_i)(t_i - t_{i-1})) \\ &\leq \frac{\varepsilon}{2(b-a)} \sum_{i=1}^{n} (t_i - t_{i-1}) = \frac{\varepsilon}{2} < \varepsilon. \end{aligned}$$

It follows from Theorem 4.3 that f is Riemann-integrable on $[a, b]$, as required. ▮

5 The Fundamental Theorem of Calculus

We have defined the Riemann integral for the class of bounded Riemann-integrable functions on some closed bounded interval, and we have shown that all continuous functions are Riemann-integrable. However the task of calculating the Riemann integral of such a function directly from the definition of this integral is somewhat difficult in practice for all but the simplest functions. We need some simpler way of finding integrals of continuous functions. To do this one uses the Fundamental Theorem of Calculus, which shows that the derivative of an indefinite integral of a function is the function itself.

Theorem 5.1 (The Fundamental Theorem of Calculus) *Let $f: [a, b] \to \mathbb{R}$ be a continuous real-valued function defined on the interval $[a, b]$, where a and b are real numbers satisfying $a < b$. Define a function $F: [a, b] \to \mathbb{R}$ by*

$$F(x) = \int_a^x f(t)\, dt.$$

Then

$$\frac{dF}{dx} = f(x)$$

for all x satisfying $a < x < b$. Similarly

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

Proof Let x be an real number satisfying $x \leq x \leq b$. Choose $\varepsilon > 0$. Because $f\colon [a,b] \to \mathbb{R}$ is continuous there exists some $\delta > 0$ such that $|f(t) - f(x)| < \frac{1}{2}\varepsilon$ for all $t \in [a,b]$ satisfying $|t - x| < \delta$.

Consider the case when $a \leq x < b$. Suppose that $0 < h < \delta$ and $[x, x+h] \subset [a,b]$. Then

$$f(x) - \tfrac{1}{2}\varepsilon \leq f(t) \leq f(x) + \tfrac{1}{2}\varepsilon$$

for all $t \in [x, x+h]$. If we integrate the terms occuring in these inequalities over the interval $[x, x+h]$ (which is of length h) we deduce that

$$(f(x) - \tfrac{1}{2}\varepsilon)h \leq \int_x^{x+h} f(t)\,dt \leq (f(x) + \tfrac{1}{2}\varepsilon)h.$$

But

$$\int_a^{x+h} f(t)\,dt = \int_a^x f(t)\,dt + \int_x^{x+h} f(t)\,dt,$$

by Theorem 4.6, hence

$$\int_x^{x+h} f(t)\,dt = F(x+h) - F(x).$$

Thus if $0 < h < \delta$ and $[x, x+h] \subset [a,b]$ then

$$f(x) - \tfrac{1}{2}\varepsilon \leq \frac{F(x+h) - F(x)}{h} \leq f(x) + \tfrac{1}{2}\varepsilon.$$

A similar argument shows that if $a < x \leq b$, if $0 < k < \delta$ and $[x-k, x] \subset [a,b]$ then

$$(f(x) - \tfrac{1}{2}\varepsilon)k \leq \int_{x-k}^{x} f(t)\,dt \leq (f(x) + \tfrac{1}{2}\varepsilon)k,$$

so that

$$f(x) - \tfrac{1}{2}\varepsilon \leq \frac{F(x) - F(x-k)}{k} \leq f(x) + \tfrac{1}{2}\varepsilon.$$

Suppose that $a < x < b$. Given $\varepsilon > 0$, we choose $\delta > 0$ such that $(x - \delta, x + \delta) \subset [a,b]$ and

$$f(x) - \tfrac{1}{2}\varepsilon < f(t) < f(x) + \tfrac{1}{2}\varepsilon$$

for all $t \in [a,b]$ satisfying $|t - x| < \delta$. (This is possible by the continuity of f.) If we apply the above results (considering the cases $0 < h < \delta$ and $-\delta < h < 0$ separately) we see that

$$\left| \frac{F(x+h) - F(x)}{h} - f(x) \right| \leq \tfrac{1}{2}\varepsilon < \varepsilon.$$

for all h satisfying $0 < |h| < \delta$. It follows from the formal definition of limits of functions that

$$\frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = f(x).$$

Similarly

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

We can use the Fundamental Theorem of Calculus to prove many very familiar theorems used when calculating integrals of continuous functions. To prove these results, we shall also make use of the following lemma.

Lemma 5.2 *Let $f\colon [a,b] \to \mathbb{R}$ and $g\colon [a,b] \to \mathbb{R}$ be continuous real-valued functions on a closed interval $[a,b]$ (where $a < b$) that are differentiable on the open interval (a,b). Suppose that $f(a) = g(a)$ and that $f'(x) = g'(x)$ for all $x \in (a,b)$ (where f' and g' are the derivatives of f and g respectively). Then $f(x) = g(x)$ for all $x \in [a,b$*

Proof Consider the function $h\colon [a,b] \to \mathbb{R}$ defined by $h = f - g$. Given x satisfying $a < x \leq b$ we apply the Mean Value Theorem to the function h on the interval $[a,x]$ to conclude that $h(x) - h(a) = h'(t)(x - a)$ for some t satisfying $a < t < x$. But $h'(t) = 0$ for all $t \in (a,b)$. Hence $h(x) - h(a) = 0$. But $h(a) = 0$. Hence $h = 0$, and thus $f = g$ on $[a,b]$. $\blacksquare$

We now describe a number of familiar corollaries of the Fundamental Theorem of Calculus.

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 $[a,b]$ with derivative $f'\colon [a,b] \to \mathbb{R}$ if and only if the derivative

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

exists for all x satisfying $a < x < b$, the one-sided derivatives

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

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

exist at the endpoints of $[a, b]$, and the function mapping t to $f'(t)$ is continuous throughout $[a, b]$ (and in particular is continuous at the endpoints of this closed interval).

Corollary 5.3 *Let $f\colon [a, b] \to \mathbb{R}$ be a continuously differentiable function on the closed interval $[a, b]$. Then*

$$\int_a^b f'(t)\, dt = f(b) - f(a)$$

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

$$g(x) = f(a) + \int_a^x f'(t)\, dt.$$

Now $g(a) = f(a)$. Also g is continuously differentiable on $[a, b]$ and $g'(x) = f'(x)$ for all $x \in [a, b]$, by the Fundamental Theorem of Calculus. Hence $g = f$ on $[a, b]$, by Lemma 5.2. In particular

$$f(b) = f(a) + \int_a^b f'(t)\, dt,$$

as required. ∎

Corollary 5.4 (Integration by Parts) *Let $f\colon [a, b] \to \mathbb{R}$ and $g\colon [a, b] \to \mathbb{R}$ be continuously differentiable functions on the closed interval $[a, b]$. Then*

$$\int_a^b f(t)g'(t)\, dt + \int_a^b f'(t)g(t)\, dt = f(b)g(b) - f(a)g(a).$$

Proof Apply Corollary 5.3 to the product function $t \mapsto f(t)g(t)$ and apply the Product Rule, which states that if h is the function defined by $h(t) = f(t)g(t)$ then $h'(t) = f(t)g'(t) + f'(t)g(t)$. ∎

Corollary 5.5 (Integration by Substitution) *Let $u\colon [a, b] \to \mathbb{R}$ be a continuously differentiable monotonically increasing real-valued function on the closed interval $[a, b]$. Define $c = u(a)$ and $d = u(b)$. Let $f\colon [c, d] \to \mathbb{R}$ be a continuous function on the closed interval $[c, d]$. Then*

$$\int_c^d f(t)\, dt = \int_a^b f(u(x))u'(x)\, dx.$$

Proof Define $F\colon [a,b] \to \mathbb{R}$ and $G\colon [c,d] \to \mathbb{R}$ by

$$\begin{aligned} F(x) &= \int_a^x f(u(y))u'(y)dy, \\ G(s) &= \int_c^s f(t)dt. \end{aligned}$$

and let us define $H\colon [a,b] \to \mathbb{R}$ by $H = G \circ u$ (so that $H(x) = G(u(x))$ for all $x \in [a,b]$). It follows from the Chain Rule and the Fundamental Theorem of Calculus that

$$\begin{aligned} H'(x) &= G'(u(x))u'(x) \\ &= f(u(x))u'(x), \\ F'(x) &= f(u(x))u'(x) = H'(x), \end{aligned}$$

for all $x \in [a,b]$. Also $F(a) = 0 = G(c) = H(a)$. We conclude from Lemma 5.2 that $H = F$ on $[a,b]$. In particular, $H(b) = F(b)$, so that

$$\int_c^d f(t)\,dt = \int_a^b f(u(x))u'(x)\,dx. \quad \blacksquare$$

6 Uniform Convergence, Limits and Integrals

Definition Let $(f_j : j \in \mathbb{N})$ be a sequence of real-valued functions defined on some set X and let f be a real-valued function on X. The sequence (f_j) is said to converge *uniformly* on X to the function f as $j \to +\infty$ if and only if, for every $\varepsilon > 0$, there exists some positive integer N (chosen *independently* of the point $\mathbf{x} \in X$), such that $|f_j(\mathbf{x}) - f(\mathbf{x})| < \varepsilon$ for all $\mathbf{x} \in X$ and for all j satisfying $j \geq N$.

Theorem 6.1 *Let $(f_j : j \in \mathbb{N})$ be a sequence of continuous real-valued functions defined over some subset X of $\mathbb{R}^n$. Suppose that the sequence (f_j) converges uniformly on X to some real-valued function f on X as $j \to +\infty$. Then f is continuous.*

Proof Let $\mathbf{a}$ be a point of X. We show that f is continuous at $\mathbf{a}$. Let $\varepsilon > 0$ be given. Then there exists some positive integer N (independent of the choice of $\mathbf{x} \in X$ such that $|f_j(\mathbf{x}) - f(\mathbf{x})| < \frac{1}{3}\varepsilon$ for all $\mathbf{x} \in X$ and for all $j \geq N$, since the sequence (f_j) of real-valued functions on X converges uniformly to f as $j \to +\infty$. Choose any integer j satisfying $j \geq N$. Now the function f_j is continuous at $\mathbf{a}$, hence there exists some $\delta > 0$ such that

$|f(\mathbf{x}) - f(\mathbf{a})| < \frac{1}{3}\varepsilon$ for all $\mathbf{x} \in X$ satisfying $|\mathbf{x} - \mathbf{a}| < \delta$. Thus if $j \geq N$ and $|\mathbf{x} - \mathbf{a}| < \delta$ then

$$\begin{aligned} |f(\mathbf{x}) - f_j(\mathbf{x})| &< \tfrac{1}{3}\varepsilon, \\ |f_j(\mathbf{x}) - f_j(\mathbf{a})| &< \tfrac{1}{3}\varepsilon, \\ |f(\mathbf{a}) - f_j(\mathbf{a})| &< \tfrac{1}{3}\varepsilon. \end{aligned}$$

It follows from the Triangle Inequality that

$$\begin{aligned} |f(\mathbf{x}) - f(\mathbf{a})| &\leq |f(\mathbf{x}) - f_j(\mathbf{x})| + |f_j(\mathbf{x}) - f_j(\mathbf{a})| + |f_j(\mathbf{a}) - f(\mathbf{a})| \\ &< \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon + \tfrac{1}{3}\varepsilon = \varepsilon. \end{aligned}$$

We have therefore shown that, given any $\varepsilon > 0$, there exists some $\delta > 0$ such that $|f(\mathbf{x}) - f(\mathbf{a})| < \varepsilon$ whenever $\mathbf{x} \in X$ satisfies $|\mathbf{x} - \mathbf{a}| < \delta$. Thus f is continuous at $\mathbf{a}$. This shows that f is continuous on X, as required. ∎

Theorem 6.2 *Let a and b be real numbers satisfying $a < b$ and let $(f_j : j \in \mathbb{N})$ be a sequence of continuous real-valued functions on $[a, b]$ which converges uniformly to some continuous real-valued function f on $[a, b]$. Then*

$$\lim_{j \to +\infty} \int_a^b f_j(t)\, dt = \int_a^b f(t)\, dt.$$

Proof Let $\varepsilon > 0$ be given. We must show that there exists some positive integer N such that

$$\left| \int_a^b f_j(t)\, dt - \int_a^b f(t)\, dt \right| < \varepsilon$$

for all integers j satisfying $j \geq N$.

Now $f_j(t) - f(j) \leq |f_j(t) - f(t)|$ and $f(t) - f_j(j) \leq |f_j(t) - f(t)|$ for all $t \in [a, b]$, hence

$$\int_a^b f_j(t)\, dt - \int_a^b f(t)\, dt = \int_a^b (f_j(t) - f(t))\, dt \leq \int_a^b |f_j(t) - f(t)|\, dt$$

and

$$\int_a^b f(t)\, dt - \int_a^b f_j(t)\, dt = \int_a^b (f_j(t) - f(t))\, dt \leq \int_a^b |f_j(t) - f(t)|\, dt.$$

Hence

$$\left| \int_a^b f_j(t)\, dt - \int_a^b f(t)\, dt \right| \leq \int_a^b |f_j(t) - f(t)|\, dt.$$

for all positive integers j.

Now the sequence $(f_j : j \in \mathbb{N})$ converges uniformly to f on $[a, b]$. Thus, given any $\varepsilon_0 > 0$, there exists some positive integer N such that $|f_j(t) - f(t)| < \varepsilon_0$ for all $t \in [a, b]$ and for all integers j satisfying $j \geq N$. Let us choose ε_0 such that

$$0 < \varepsilon_0 < \frac{\varepsilon}{b - a}.$$

We conclude that there exists a positive integer N such that

$$|f_j(t) - f(t)| < \varepsilon_0 < \frac{\varepsilon}{b - a},$$

for all $t \in [a, b]$ and for all integers j satisfying $j \geq N$. Thus if $j \geq N$ then

$$\begin{aligned} \left| \int_a^b f_j(t)\, dt - \int_a^b f(t)\, dt \right| &\leq \int_a^b |f_j(t) - f(t)|\, dt \\ &\leq \varepsilon_0(b - a) < \varepsilon, \end{aligned}$$

as required. ∎

Example Let $(f_j : j \in \mathbb{N})$ be the sequence of continuous functions on the interval $[0, 1]$ defined by

$$f_j(t) = j(t^j - t^{2j}).$$

Note that $f_j(0) = f_j(1) = 0$ for all $j \in \mathbb{N}$. Suppose that $0 < t < 1$.

We show that

$$\lim_{j \to +\infty} f_j(t) = 0$$

for all t satisfying $0 \leq t \leq 1$. Given t satisfying $0 < t < 1$ let us choose u such that $t < u < 1$. There exists some positive integer k such that

$$\frac{k + 1}{k} < \frac{u}{t}.$$

Define $s_j = jt^j$. Then

$$\frac{s^{j+1}}{s^j} = \frac{(j + 1)t}{j} \leq \frac{(k + 1)t}{k} < u$$

for all integers j satisfying $j \geq k$ (where we have used the fact that

$$\frac{j + 1}{j} \leq \frac{k + 1}{k}$$

for all integers j satisfying $j \geq k$). It follows by induction on j that if $j > k$ then $s_j/s_k < u^{j-k}$, so that

$$0 < jt^j < kt^k u^{j-k}.$$

Now $|jt^{2j}| \leq |jt^j|$, hence $|f_j(t)| < 2kt^k u^{j-k}$ for all integers j satisfying $j > k$. But

$$\lim_{j \to +\infty} 2kt^k u^{j-k} = 0,$$

since $0 < u < 1$. We conclude that

$$\lim_{j \to +\infty} f_j(t) = 0$$

for all t satisfying $0 < t < 1$. But this result also holds when $t = 0$ and $t = 1$, since $f_j(0) = f_j(1) = 0$ for all j. Thus

$$\lim_{j \to +\infty} f_j(t) = 0 \qquad (0 \leq t \leq 1).$$

We claim however that the sequence $(f_j : j \in \mathbb{N})$ of real-valued functions on $[a, b]$ *does not converge uniformly* on $[0, 1]$ to the zero function. Let us define

$$\|f_j\| = \sup\{|f_j(t)| : 0 \leq t \leq 1\}.$$

It is easy to show from the definition of uniform convergence that if the sequence $(f_j : j \in \mathbb{N})$ were to converge uniformly to the zero function on $[0, 1]$, then we would have $\|f_j\| \to 0$ as $j \to +\infty$. (Indeed, if $\varepsilon > 0$ is given and if $f_j \to 0$ uniformly on $[0, 1]$ then there would exist a positive integer N such that $|f_j(t)| < \frac{1}{2}\varepsilon$ for all $t \in [0, 1]$ and for all j satisfying $j \geq N$, so that $\|f_j\| \leq \frac{1}{2}\varepsilon < \varepsilon$ for all $j \geq N$, showing that $\|f_j\| \to 0$ as $j \to +\infty$.)

Let us evaluate $\|f_j\|$. Note that $f(t) \geq 0$ for all $t \in [0, 1]$. Now

$$f_j'(t) = j^2 t^{j-1}(1 - 2t^j).$$

We see that f_j is non-negative on $[0, 1]$, it is increasing on $[0, x_j]$ and it is decreasing on $[x_j, 1]$, where $x_j^j = \frac{1}{2}$. Now $f_j(x^j) = j/4$. We conclude that $\|f_j\| = j/4$. Thus the sequence $(f_j : j \in \mathbb{N})$ does not converge uniformly to the zero function on $[0, 1]$.

Now

$$\int_0^1 f_j(t)\,dt = \frac{j}{j+1} - \frac{j}{2j+1},$$

hence

$$\lim_{j \to +\infty} \int_0^1 f_j(t)\,dt = \frac{1}{2}.$$

This shows that

$$\lim_{j\to+\infty} \int_0^1 f_j(t)\,dt \neq 0,$$

even though $f_j(t) \to 0$ as $j \to +\infty$ for all $t \in [0,1]$. *Let this example be a warning to you, demonstrating that it is not always possible to interchange limits and integrals.* Perhaps the most remarkable fact about the above example is that the functions f_j used are not weird 'pathological' functions, but are merely simple polynomial functions.

Example Consider the continuous functions $g_j\colon [0,1] \to \mathbb{R}$ on $[0,1]$ (for all positive integers j) defined by

$$g_j(t) = \begin{cases} 0 & \text{if } 0 \leq t \leq \frac{1}{4j}, \\ 4j(4jt-1) & \text{if } \frac{1}{4j} \leq t \leq \frac{1}{2j}, \\ 4j(3-4jt) & \text{if } \frac{1}{2j} \leq t \leq \frac{3}{4j}, \\ 0 & \text{if } \frac{3}{4j} \leq t \leq 1. \end{cases}$$

Now a simple calculation shows that

$$\int_0^1 g_j(t)\,dt = 1.$$

(Indeed this integral is the area of a triangle in the plane whose base is of length $1/(2j)$ and whose height is $4j$.) Thus

$$\int_0^1 g_j(t)\,dt \to +\infty$$

as $j \to +\infty$. But

$$\lim_{j\to+\infty} g_j(t) = 0$$

for all $t \in [0,1]$. Indeed $g_j(0) = 0$ for all j, and if $0 < t \leq 1$ then $g_j(t) = 0$ for all $j \geq 3/(4t)$. However it is readily seen that the sequence $(g_j : j \in \mathbb{N})$ does not converge uniformly to the zero function on $[0,1]$.

6.1 Integrals over Unbounded Intervals

Definition Let $f\colon [a,+\infty)$ be a bounded real-valued function on the (unbounded) interval $[a,+\infty)$ which is Riemann-integrable over each closed bounded subinterval of $[a,+\infty)$. We define

$$\int_a^{+\infty} f(t)\,dt = \lim_{b\to+\infty} \int_a^b f(t)\,dt,$$

provided that this limit exists. Similarly if $f\colon(-\infty,b]$ is a bounded real-valued function on the (unbounded) interval $(-\infty,b]$ which is Riemann-integrable over each closed bounded subinterval of $(-\infty,b]$ then we define

$$\int_{-\infty}^{b} f(t)\,dt = \lim_{a\to-\infty} \int_a^b f(t)\,dt,$$

provided that this limit exists. If $f\colon\mathbb{R}\to\mathbb{R}$ is a real-valued function defined over the whole of $\mathbb{R}$ which is Riemann-integrable over each closed bounded subinterval of $\mathbb{R}$ then we define

$$\int_{-\infty}^{+\infty} f(t)\,dt = \lim_{a\to-\infty,b\to+\infty} \int_a^b f(t)\,dt,$$

provided that this limit exists.

Example Consider the functions $h_j\colon\mathbb{R}\to\mathbb{R}$ defined by

$$h_j(t) = \frac{j}{t^2+j^2}$$

for all $t\in\mathbb{R}$ and for all $j\in\mathbb{N}$. Note that

$$\int_{-\infty}^{+\infty} h_j(t)\,dt = \int_{-\infty}^{+\infty} \frac{du}{u^2+1} = \pi.$$

Note also that $h_j(t)\to 0$ as $j\to+\infty$ for all $t\in\mathbb{R}$. Indeed $h_j(t)\leq 1/j$ for all $t\in\mathbb{R}$, hence the sequence $(h_j : j\in\mathbb{N})$ converges uniformly to the zero function on $\mathbb{R}$.

Remark We know from Theorem 6.2 that if $(f_j : j\in\mathbb{N})$ is a sequence of continuous real-valued functions on a *closed bounded* interval $[a,b]$ which converges *uniformly* on $[a,b]$ to a continuous function f then

$$\lim_{j\to+\infty} \int_a^b f_j(t)\,dt = \int_a^b f(t)\,dt.$$

The above example shows that the corresponding result does not hold for integrals over *unbounded* intervals. (It is instructive to examine the proof of Theorem 6.2 in order to see why the proof cannot be generalized to cover the case where the integrals are taken over an unbounded interval.)

6.2 Integrals of Unbounded Functions

Let $f\colon [a,b]$ be a real-valued function on an interval $[a,b]$ and let c satisfy $a < c < b$. Suppose that f is unbounded around c but has the property that f is bounded and Riemann-integrable on the intervals $[a,u]$ and $[v,b]$ for all u, v satisfying $a < u < c < v < b$. Then we define

$$\int_a^b f(t)\,dt = \lim_{u\to c^-} \int_a^u f(t)\,dt + \lim_{v\to c^+} \int_v^b f(t)\,dt,$$

provided that this limit exists. Similarly if f is unbounded around a but has the property that f is bounded and Riemann-integrable on $[v,b]$ for all v satisfying $a < v < b$, then we define

$$\int_a^b f(t)\,dt = \lim_{v\to a^+} \int_v^b f(t)\,dt,$$

provided that this limit exists, and if f is f is unbounded around b but has the property that f is bounded and Riemann-integrable on $[a,u]$ for all u satisfying $a < u < b$, then we define

$$\int_a^b f(t)\,dt = \lim_{u\to b^-} \int_a^u f(t)\,dt,$$

provided that this limit exists.

Example Let α satisfy $0 < \alpha < 1$. Then

$$\begin{aligned}\int_0^1 t^{-\alpha}\,dt &= \lim_{v\to 0+} \int_v^1 t^{-\alpha}\,dt \\ &= \frac{1}{1-\alpha} \lim_{v\to 0+} (1 - v^{1-\alpha}) \\ &= \frac{1}{1-\alpha}.\end{aligned}$$

7 Differentiation of Functions of Several Real Variables

7.1 Linear Transformations

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

then

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

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

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

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

$$|T\mathbf{x} - T\mathbf{y}| \leq M|\mathbf{x} - \mathbf{y}|$$

for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, where

$$M^2 = \sum_{i=1}^m \sum_{j=1}^n T_{ij}^2$$

(where (T_{ij}) is the $m \times n$ matrix which represents the linear transformation $T: \mathbb{R}^n \to \mathbb{R}^m$).

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

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

for all integers i between 1 and m. It follows from Schwarz' Inequality (Lemma 1.1) that

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

Hence

$$|\mathbf{v}^2| = \sum_{i=1}^m v_i^2 \leq \left(\sum_{i=1}^m \sum_{j=1}^n T_{ij}^2 \right) |\mathbf{x} - \mathbf{y}|^2 = M^2 |\mathbf{x} - \mathbf{y}|^2.$$

Thus $|T\mathbf{x} - T\mathbf{y}| \leq M|\mathbf{x} - \mathbf{y}|$. It follows from this that T is uniformly continuous. Indeed if we are given $\varepsilon > 0$ let us define $\delta > 0$ by $\delta = \varepsilon/M$. If $\mathbf{x}$ and $\mathbf{y}$ are elements of $\mathbb{R}^n$ which satisfy the condition $|\mathbf{x} - \mathbf{y}| < \delta$ then $|T\mathbf{x} - T\mathbf{y}| < \varepsilon$. This shows that $T: \mathbb{R}^n \to \mathbb{R}^m$ is uniformly continuous on $\mathbb{R}^n$, as required. ∎

7.2 Differentiability for Functions of One Real Variable

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

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

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

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

Lemma 7.2 *Let $f: I \to \mathbb{R}$ be a real-valued function defined on some open interval I in $\mathbb{R}$. Let a be an element of I. The function f is differentiable at a with derivative $f'(a)$ (where $f'(a)$ is some real number) if and only if*

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

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

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

if and only if

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

But

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

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

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

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

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

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

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

where

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

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

$$\mathbf{x} \mapsto f(\mathbf{a}) + f'(\mathbf{a})(\mathbf{x} - \mathbf{a})$$

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

7.3 Differentiation of Functions of Several Variables

Definition Let D be an open subset of $\mathbb{R}^n$ and let $f\colon D \to \mathbb{R}^m$ be a map from D into $\mathbb{R}^m$. Let $\mathbf{a}$ be a point of D. The function f is said to be *differentiable* at $\mathbf{a}$ if and only if there exists a linear transformation $f'(\mathbf{a})\colon R^n \to R^m$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ with the property that

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

If f is differentiable at $\mathbf{a}$ then the linear transformation $f'(\mathbf{a})$ is referred to as the *derivative* of f at $\mathbf{a}$.

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

Observe that if f is differentiable at $\mathbf{a}$ with derivative $f'(\mathbf{a})$ then

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + f'(\mathbf{a})\mathbf{h} + e(\mathbf{a},\mathbf{h}),$$

where

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

Thus if f is differentiable at $\mathbf{a}$ then the map $l\colon D \to \mathbb{R}$ defined by

$$l(\mathbf{x}) = f(\mathbf{a}) + f'(\mathbf{a})(\mathbf{x}-\mathbf{a})$$

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

Lemma 7.3 *Let $f\colon D \to \mathbb{R}^m$ be a function which maps an open subset D of $\mathbb{R}^n$ into $\mathbb{R}^m$ which is differentiable at some point $\mathbf{a}$ of D. Then f is continuous at $\mathbf{a}$.*

Proof If we define

$$e(\mathbf{a},\mathbf{h}) = f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - f'(\mathbf{a})\mathbf{h}$$

then

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

(because f is differentiable at $\mathbf{a}$), and hence

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

But

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

since

$$\lim_{\mathbf{h}\to\mathbf{0}} f'(\mathbf{a})\mathbf{h} = f'(\mathbf{a})\left(\lim_{\mathbf{h}\to\mathbf{0}} \mathbf{h}\right) = \mathbf{0}$$

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

$$\lim_{\mathbf{h}\to\mathbf{0}} f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}),$$

showing that f is continuous at $\mathbf{a}$. ∎

Lemma 7.4 *Let $f\colon D \to \mathbb{R}^m$ be a function which maps an open subset D of $\mathbb{R}^n$ into $\mathbb{R}^m$ which is differentiable at some point $\mathbf{a}$ of D. Let $f'(\mathbf{a})\colon \mathbb{R}^n \to \mathbb{R}^m$ be the derivative of f at $\mathbf{a}$. Let $\mathbf{u}$ be an element of $\mathbb{R}^n$. Then*

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

Thus the derivative $f'(\mathbf{a})$ of f at $\mathbf{a}$ is uniquely determined by the map f.

Proof It follows from the differentiability of f at $\mathbf{a}$ that

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

In particular, if we set $\mathbf{h} = \lambda \mathbf{u}$, where $\lambda > 0$ then we deduce that

$$\lim_{\lambda \to 0^+} \frac{1}{\lambda} \left(f(\mathbf{a} + \lambda \mathbf{u}) - f(\mathbf{a}) - \lambda f'(\mathbf{a})\mathbf{u} \right) = \mathbf{0},$$

showing that

$$\lim_{\lambda \to 0^+} \frac{1}{\lambda} \left(f(\mathbf{a} + \lambda \mathbf{u}) - f(\mathbf{a}) \right) = f'(\mathbf{a})\mathbf{u}.$$

Similarly if we set $\mathbf{h} = \lambda \mathbf{u}$, where $\lambda < 0$, when we deduce that

$$\lim_{\lambda \to 0^-} \frac{-1}{\lambda} \left(f(\mathbf{a} + \lambda \mathbf{u}) - f(\mathbf{a}) - \lambda f'(\mathbf{a})\mathbf{u} \right) = \mathbf{0},$$

showing that

$$\lim_{\lambda \to 0^-} \frac{1}{\lambda} \left(f(\mathbf{a} + \lambda \mathbf{u}) - f(\mathbf{a}) \right) = f'(\mathbf{a})\mathbf{u}.$$

We conclude therefore that

$$\lim_{\lambda \to 0} \frac{1}{\lambda} \left(f(\mathbf{a} + \lambda \mathbf{u}) - f(\mathbf{a}) \right) = f'(\mathbf{a})\mathbf{u},$$

as required. ∎

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

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

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

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

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

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

We see therefore that if f is differentiable at $\mathbf{a}$ then

$$f'(\mathbf{a})\mathbf{e}_j = \left(\frac{\partial f_1}{\partial x_j}, \frac{\partial f_2}{\partial x_j}, \ldots, \frac{\partial f_m}{\partial x_j} \right).$$

Thus the linear transformation $f'(\mathbf{a}) \colon \mathbb{R}^n \to \mathbb{R}^m$ is represented by the $m \times n$ matrix

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

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

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

$$f\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) = \begin{pmatrix} x_1^2 - x_2^2 \\ 2x_1 x_2 \end{pmatrix}.$$

Let (a_1, a_2) be a point of $\mathbb{R}^2$. Now

$$f\left(\begin{pmatrix} a_1 + h_1 \\ a_2 + h_2 \end{pmatrix} \right) = \begin{pmatrix} a_1^2 - a_2^2 + 2a_1h_1 - 2a_2h_2 + h_1^2 - h_2^2 \\ 2a_1a_2 + 2a_2h_1 + 2a_1h_2 + 2h_1h_2 \end{pmatrix}.$$

Thus

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + T\mathbf{h} + e(\mathbf{a}, \mathbf{h})$$

where

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \qquad \mathbf{h} = \begin{pmatrix} h_1 \\ h_2 \end{pmatrix},$$

$$T\mathbf{h} = \begin{pmatrix} 2a_1h_1 - 2a_2h_2 \\ 2a_2h_1 + 2a_1h_2 \end{pmatrix} = \begin{pmatrix} 2a_1 & -2a_2 \\ 2a_2 & 2a_1 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix},$$

and

$$e(\mathbf{a}, \mathbf{h}) = \begin{pmatrix} h_1^2 - h_2^2 \\ 2h_1h_2 \end{pmatrix}.$$

Now $|\mathbf{h}|^2 = h_1^2 + h_2^2$, and

$$\begin{aligned} |e(\mathbf{a}, \mathbf{h})|^2 &= (h_1^2 - h_2^2)^2 + (2h_1h_2)^2 \\ &= h_1^4 + h_2^4 - 2h_1^2h_2^2 + 4h_1^2h_2^2 \\ &= h_1^4 + h_2^4 + 2h_1^2h_2^2 \\ &= (h_1^2 + h_2^2)^2 = |\mathbf{h}|^4. \end{aligned}$$

Thus

$$\left|\frac{e(\mathbf{a}, \mathbf{h})}{|\mathbf{h}|}\right| = |\mathbf{h}|,$$

so that

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

This proves that f is differentiable at $\mathbf{a}$ and that the derivative $f'(\mathbf{a})$ of f at $\mathbf{a}$ is given by the linear transformation $T\colon \mathbb{R}^2 \to \mathbb{R}^2$ and is thus represented by the 2×2 matrix

$$f'(\mathbf{a}) = \begin{pmatrix} 2a_1 & -2a_2 \\ 2a_2 & 2a_1 \end{pmatrix}.$$

Note that

$$\begin{aligned} 2a_1 &= \left.\frac{\partial(x_1^2 - x_2^2)}{\partial x^1}\right|_{\mathbf{x}=\mathbf{a}} \\ 2a_2 &= \left.\frac{\partial(2x_1x_2)}{\partial x^1}\right|_{\mathbf{x}=\mathbf{a}} \\ -2a_2 &= \left.\frac{\partial(x_1^2 - x_2^2)}{\partial x^2}\right|_{\mathbf{x}=\mathbf{a}} \\ 2a_1 &= \left.\frac{\partial(2x_1x_2)}{\partial x^2}\right|_{\mathbf{x}=\mathbf{a}} \end{aligned}$$

Thus the derivative $f'(\mathbf{a})$ of f at $\mathbf{a}$ is indeed represented by the value at $\mathbf{a}$ of the Jacobian matrix

$$\begin{pmatrix} \dfrac{\partial(x_1^2 - x_2^2)}{\partial x^1} & \dfrac{\partial(x_1^2 - x_2^2)}{\partial x^2} \\ \dfrac{\partial(2x_1x_2)}{\partial x^1} & \dfrac{\partial(2x_1x_2)}{\partial x^2} \end{pmatrix}.$$

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

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

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

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

cannot possibly exist.) Because f is not continuous at $(0,0)$ we conclude from Lemma 7.3 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 $x, y \in \mathbb{R}$.

Remark This last example exhibits an important point. It shows 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 the next theorem shows that if the first order partial derivatives of the components of a function exist *and are continuous* throughout some neighbourhood of a given point then the function is differentiable at that point.

Theorem 7.5 *Let D be an open subset of $\mathbb{R}^m$ and let Let $f: D \to \mathbb{R}$ be a function mapping D into $\mathbb{R}$. Suppose that the first order partial derivatives of the components of f exist and are continuous on D. Then f is differentiable at each point of D.*

Proof We denote by $\partial_j f: D \to \mathbb{R}$ the partial derivative of f with respect to the jth coordinate. Thus

$$\partial_j f \equiv \frac{\partial f}{\partial x_j}$$

Let $\mathbf{a}$ be a point of D. If f is differentiable at $\mathbf{a}$ then the derivative of f at $\mathbf{a}$ would be represented by the Jacobian matrix, and would thus be the linear transformation sending $(h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$ to

$$\sum_{j=1}^{n} (\partial_j f)(\mathbf{a}) h_j.$$

Thus in order to show that f is differentiable at $\mathbf{a}$ we must prove that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|}\left(f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \sum_{j=1}^{n} (\partial_j f)(\mathbf{a}) h_j \right) = 0,$$

where $\mathbf{h} = (h_1, h_2, \ldots, h_n)$.

Let $\varepsilon > 0$ be given. Using the fact that D is open, and using the continuity of the functions $\partial_1 f, \partial_2 f, \ldots, \partial_n f$, we see that there exists some $\delta > 0$ such that the open ball $B(\mathbf{a}, \delta)$ of radius δ about $\mathbf{a}$ is contained in D, where

$$D = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{a}| < \delta\},$$

and such that, for all $\mathbf{x} \in B(\mathbf{a}, \delta)$,

$$|(\partial_j f)(\mathbf{x}) - (\partial_j f)(\mathbf{a})| < \frac{\varepsilon}{n} \qquad (j = 1, 2, \ldots, n).$$

Given $\mathbf{h}$ satisfying $|\mathbf{x}| < \delta$, let us define

$$c_j = f(a_1 + h_1, \ldots, a_j + h_j, a_{j+1}, \ldots, a_n) \qquad (j = 0, 1, \ldots, n).$$

Thus

$$c_0 = f(\mathbf{a}), \qquad c_n = f(\mathbf{a} + \mathbf{h}).$$

Also

$$c_j - c_{j-1} = g_j(a_j + h_j) - g_j(a_j) \qquad (j = 1, 2, \ldots, n),$$

where

$$g_j(t) = f(a_1 + h_1, \ldots, a_{j-1} + h_{j-1}, t, a_{j+1}, \ldots, a_n).$$

Now there exist $\theta_1, \theta_2, \ldots, \theta_n \in (0, 1)$ with the property that

$$g_j(a_j + h_j) - g_j(a_j) = h_j g_j'(a_j + \theta_j h_j) \qquad (j = 1, 2, \ldots, n),$$

by the Mean Value Theorem. Moreover

$$g_j'(a_j + \theta_j h_j) = (\partial_j f)(a_1 + h_1, \ldots, a_{j-1} + h_{j-1}, a_j + \theta_j h_j, a_{j+1}, \ldots, a_n),$$

and hence

$$\left| g_j'(a_j + \theta_j h_j) - (\partial f)(\mathbf{a}) \right| < \frac{\varepsilon}{n}$$

(on account of the fact that the point

$$(a_1 + h_1, \ldots, a_{j-1} + h_{j-1}, a_j + \theta_j h_j, a_{j+1}, \ldots, a_n)$$

lies within the open ball $B(\mathbf{a},\delta)$ of radius δ about $\mathbf{a}$). We see therefore that

$$|c_j - c_{j-1} - (\partial_j f)(\mathbf{a})h_j| < \frac{|h_j|\varepsilon}{n} \qquad (j = 1, 2, \ldots, n),$$

so that

$$\begin{aligned} \left| f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a} - \sum_{j=1}^{n} (\partial_j f)(\mathbf{a})h_j \right| &= \left| c_n - c_0 - \sum_{j=1}^{n} (\partial_j f)(\mathbf{a})h_j \right| \\ &\leq \sum_{j=1}^{n} |c_j - c_{j-1} - (\partial_j f)(\mathbf{a})h_j| \\ &< \sum_{j=1}^{n} \frac{|h_j|\varepsilon}{n} \leq \varepsilon|\mathbf{h}|. \end{aligned}$$

Thus

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

whenever $|\mathbf{h}| < \delta$. This shows that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} \left(f(\mathbf{a}+\mathbf{h}) - f(\mathbf{a}) - \sum_{j=1}^{n} (\partial_j f)(\mathbf{a})h_j \right) = 0,$$

thus proving that f is differentiable at $\mathbf{a}$. ∎

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

Corollary 7.6 *Let D be an open subset of $\mathbb{R}^n$ and let $u: D \to \mathbb{R}^m$ be a function mapping D into $\mathbb{R}^m$. Suppose that the Jacobian matrix*

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

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

Lemma 7.7 *Let $T\colon \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$. Then T is differentiable at each point $\mathbf{a}$ of $\mathbb{R}^n$, and $T'(\mathbf{a}) = T$.*

Proof This follows immediately from the fact that $T(\mathbf{a}+\mathbf{h}) = T\mathbf{a}+T\mathbf{h}$. ∎

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

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

$$\begin{aligned} (f+g)'(\mathbf{a}) &= f'(\mathbf{a}) + g'(\mathbf{a}), \\ (f-g)'(\mathbf{a}) &= f'(\mathbf{a}) - g'(\mathbf{a}), \\ (fg)'(\mathbf{a}) &= g(\mathbf{a})f'(\mathbf{a}) + f(\mathbf{a})g'(\mathbf{a}). \end{aligned}$$

Proof We can write

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) &= f(\mathbf{a}) + f'(\mathbf{a})\mathbf{h} + e_1(\mathbf{a},\mathbf{h}), \\ g(\mathbf{a}+\mathbf{h}) &= g(\mathbf{a}) + g'(\mathbf{a})\mathbf{h} + e_2(\mathbf{a},\mathbf{h}), \end{aligned}$$

for all sufficiently small h, where

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

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

$$\begin{aligned} &\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} \left| f(\mathbf{a}+\mathbf{h}) + g(\mathbf{a}+\mathbf{h}) - (f(\mathbf{a}) + g(\mathbf{a})) - (f'(\mathbf{a}) + g'(\mathbf{a}))\,\mathbf{h} \right| \\ &\qquad = \lim_{\mathbf{h}\to\mathbf{0}} \frac{e_1(\mathbf{a},\mathbf{h}) + e_2(\mathbf{a},\mathbf{h})}{|\mathbf{h}|} = 0, \\ &\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} \left| f(\mathbf{a}+\mathbf{h}) - g(\mathbf{a}+\mathbf{h}) - (f(\mathbf{a}) - g(\mathbf{a})) - (f'(\mathbf{a}) - g'(\mathbf{a}))\,\mathbf{h} \right| \\ &\qquad = \lim_{\mathbf{h}\to\mathbf{0}} \frac{e_1(\mathbf{a},\mathbf{h}) - e_2(\mathbf{a},\mathbf{h})}{|\mathbf{h}|} = 0. \end{aligned}$$

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

$$f(\mathbf{a}+\mathbf{h})g(\mathbf{a}+\mathbf{h}) = f(\mathbf{a})g(\mathbf{a}) + g(\mathbf{a})f'(\mathbf{a})\mathbf{h} + f(\mathbf{a})g'(\mathbf{a})\mathbf{h} + e(\mathbf{a},\mathbf{h}),$$

where

$$\begin{aligned} e(\mathbf{a}, \mathbf{h}) &= (f(\mathbf{a}) + f'(\mathbf{a})\mathbf{h})e_2(\mathbf{a}, \mathbf{h}) + (g(\mathbf{a}) + g'(\mathbf{a})\mathbf{h})e_1(\mathbf{a}, \mathbf{h}) \\ &\quad + (f'(\mathbf{a})\mathbf{h})(g'(\mathbf{a})\mathbf{h}) + e_1(\mathbf{a}, \mathbf{h})e_2(\mathbf{a}, \mathbf{h}). \end{aligned}$$

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

$$|f'(\mathbf{a})\mathbf{h}| \leq M_1|\mathbf{h}|, \qquad |g'(\mathbf{a})\mathbf{h}| \leq M_2|\mathbf{h}|.$$

Therefore

$$|(f'(\mathbf{a})\mathbf{h})(g'(\mathbf{a})\mathbf{h})| \leq M_1 M_2 |\mathbf{h}|^2,$$

so that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|}(f'(\mathbf{a})\mathbf{h})(g'(\mathbf{a})\mathbf{h}) = 0.$$

Also

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

Therefore

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

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

$$(fg)'(\mathbf{a}) = g(\mathbf{a})f'(\mathbf{a}) + f(\mathbf{a})g'(\mathbf{a}). \quad \blacksquare$$

Let D be an open set in $\mathbb{R}^n$, and let $f\colon D \to \mathbb{R}^m$ be a function mapping D into $\mathbb{R}^m$. Let E be an open set in $\mathbb{R}^m$ which contains $f(D)$, and let $g\colon E \to \mathbb{R}^p$ be a function mapping E into $\mathbb{R}^p$. The Chain Rule states that if f is differentiable at some point $\mathbf{a}$ of D and if g is differentiable at $f(\mathbf{a})$ then the composition $g \circ f$ is differentiable at $\mathbf{a}$ and

$$(g \circ f)'(\mathbf{a}) = g'(f(\mathbf{a})) \circ f'(\mathbf{a}).$$

Thus a composition of differentiable functions is differentiable. Moreover the derivative of a composition of two functions is the composition of the derivatives of those functions. The following lemma will be used in the proof of the Chain Rule.

Lemma 7.9 *Let D be an open set in $\mathbb{R}^n$ and let $\mathbf{a}$ be a point of D. Let $\varphi\colon D \to \mathbb{R}^m$ be a mapping from D to $\mathbb{R}^n$ and let $\lambda\colon D \to \mathbb{R}$ be a real-valued function on D. Suppose that*

$$\lim_{\mathbf{x}\to\mathbf{a}} \varphi(\mathbf{x}) = \mathbf{0}$$

and that $|\lambda(\mathbf{x})| \leq K$ for all $x \in D$, where K is a suitable constant. Then

$$\lim_{\mathbf{x}\to\mathbf{a}} \lambda(\mathbf{x})\varphi(\mathbf{x}) = \mathbf{0}.$$

Proof Let $\varepsilon > 0$ be given. Then there exists some $\delta > 0$ such that $|\varphi(\mathbf{x})| < \varepsilon/K$ whenever $\mathbf{x} \in D$ satisfies $|\mathbf{x}-\mathbf{a}| < \delta$, since $\varphi(\mathbf{x}) \to \mathbf{0}$ as $\mathbf{x} \to \mathbf{a}$. But then $|\lambda(\mathbf{x})\varphi(\mathbf{x})| < \varepsilon$ whenever $\mathbf{x} \in D$ satisfies $|\mathbf{x} - \mathbf{a}| < \delta$. Thus $\lambda(\mathbf{x})\varphi(\mathbf{x}) \to \mathbf{0}$ as $\mathbf{x} \to \mathbf{a}$. ∎

We now state and prove the Chain Rule.

Theorem 7.10 (Chain Rule) *Let D be an open set in $\mathbb{R}^n$, and let $f\colon D \to \mathbb{R}^m$ be a function mapping D into $\mathbb{R}^m$. Let E be an open set in $\mathbb{R}^m$ which contains $f(D)$, and let $g\colon E \to \mathbb{R}^p$ be a function mapping E into $\mathbb{R}^p$. Let $\mathbf{a}$ be a point of D. Suppose that f is differentiable at $\mathbf{a}$ and that g is differentiable at $f(\mathbf{a})$. Then the composition $g \circ f$ (i.e., f followed by g) is differentiable at $\mathbf{a}$. Moreover*

$$(g \circ f)'(\mathbf{a}) = g'(f(\mathbf{a})) \circ f'(\mathbf{a}).$$

Proof We can write

$$f(\mathbf{a} + \mathbf{h}) = f(\mathbf{a}) + f'(\mathbf{a})\mathbf{h} + e_1(\mathbf{a}, \mathbf{h})$$

for all sufficiently small h, where

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

on account of the fact that f is differentiable at $\mathbf{a}$. Define $\mathbf{b} = f(\mathbf{a})$. We can write

$$g(\mathbf{b} + \mathbf{k}) = g(\mathbf{b}) + g'(\mathbf{b})\mathbf{k} + e_2(\mathbf{b}, \mathbf{k})$$

for all sufficiently small h, where

$$\lim_{\mathbf{k}\to\mathbf{0}} \frac{e_2(\mathbf{b}, \mathbf{k})}{|\mathbf{k}|} = \mathbf{0},$$

on account of the fact that g is differentiable at $\mathbf{b}$. Thus

$$\begin{aligned} g(f(\mathbf{a}+\mathbf{h})) &= g(\mathbf{b}+f'(\mathbf{a})\mathbf{h}+e_1(\mathbf{a},\mathbf{h})) \\ &= g(\mathbf{b})+g'(\mathbf{b})(f'(\mathbf{a})\mathbf{h}+e_1(\mathbf{a},\mathbf{h}))+e_2(\mathbf{b},(f'(\mathbf{a})\mathbf{h}+e_1(\mathbf{a},\mathbf{h}))) \\ &= g(f(\mathbf{a}))+g'(f(\mathbf{a}))f'(\mathbf{a})\mathbf{h}+e(\mathbf{a},\mathbf{h}), \end{aligned}$$

where

$$e(\mathbf{a},\mathbf{h}) = g'(\mathbf{b})e_1(\mathbf{a},\mathbf{h})+e_2(\mathbf{b},(f'(\mathbf{a})\mathbf{h}+e_1(\mathbf{a},\mathbf{h}))).$$

Thus in order to show that $g \circ f$ is differentiable at $\mathbf{a}$ with derivative $g'(f(\mathbf{a})) \circ f'(\mathbf{a})$ we must show that

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

Using the fact that $g'(\mathbf{b})$ is a linear transformation from $\mathbb{R}^m$ to $\mathbb{R}^p$, and is thus continuous (by Lemma 7.1) we see that

$$\begin{aligned} \lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} g'(\mathbf{b})e_1(\mathbf{a},\mathbf{h}) &= \lim_{\mathbf{h}\to\mathbf{0}} g'(\mathbf{b}) \left(\frac{e_1(\mathbf{a},\mathbf{h})}{|\mathbf{h}|}\right) \\ &= g'(\mathbf{b}) \left(\lim_{\mathbf{h}\to\mathbf{0}} \frac{e_1(\mathbf{a},\mathbf{h})}{|\mathbf{h}|}\right) \\ &= g'(\mathbf{b})\mathbf{0} = \mathbf{0}. \end{aligned}$$

Thus it only remains to show that

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} e_2(\mathbf{b}, f'(\mathbf{a})\mathbf{h}+e_1(\mathbf{a},\mathbf{h})) = \mathbf{0}.$$

Let $\varphi(\mathbf{k})$ be defined by

$$\varphi(\mathbf{k}) = \begin{cases} \dfrac{e_2(\mathbf{b},\mathbf{k})}{|\mathbf{k}|} & \text{if } \mathbf{k} \neq \mathbf{0} \text{ and } \mathbf{b}+\mathbf{k} \in E; \\ \mathbf{0} & \text{if } \mathbf{k} = \mathbf{0}. \end{cases}$$

Then φ is continuous at $\mathbf{0}$, since

$$\lim_{\mathbf{k}\to\mathbf{0}} \frac{e_2(\mathbf{b},\mathbf{k})}{|\mathbf{k}|} = \mathbf{0}.$$

Now

$$\frac{1}{|\mathbf{h}|} e_2(\mathbf{b}, f'(\mathbf{a})\mathbf{h}+e_1(\mathbf{a},\mathbf{h})) = \varphi(f'(\mathbf{a})\mathbf{h}+e_1(\mathbf{a},\mathbf{h})) \frac{|f'(\mathbf{a})\mathbf{h}+e_1(\mathbf{a},\mathbf{h})|}{|\mathbf{h}|}.$$

Now it follows from Lemma 7.1 that there exists a constant M such that $|f'(\mathbf{a})\mathbf{h}| \leq M|\mathbf{h}|$ for all $\mathbf{h}$. also $e_1(\mathbf{a},\mathbf{h}) \to \mathbf{0}$ as $\mathbf{h} \to \mathbf{0}$. Thus there exists a sufficiently small neighbourhood of $\mathbf{0}$ in $\mathbb{R}^n$ such that

$$|f'(\mathbf{a})\mathbf{h} + e_1(\mathbf{a},\mathbf{h})| \leq (M+1)|\mathbf{h}|.$$

for all $\mathbf{h} \in N$. Now

$$\begin{aligned} \lim_{\mathbf{h}\to\mathbf{0}} \varphi(f'(\mathbf{a})\mathbf{h} + e_1(\mathbf{a},\mathbf{h})) &= \varphi\left(\lim_{\mathbf{h}\to\mathbf{0}}(f'(\mathbf{a})\mathbf{h} + e_1(\mathbf{a},\mathbf{h}))\right) \\ &= \varphi(\mathbf{0}) = \mathbf{0} \end{aligned}$$

and

$$\frac{|f'(\mathbf{a})\mathbf{h} + e_1(\mathbf{a},\mathbf{h})|}{|\mathbf{h}|} \leq M+1$$

for all sufficiently small $\mathbf{h}$. Therefore

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{1}{|\mathbf{h}|} e_2(\mathbf{b}, f'(\mathbf{a})\mathbf{h} + e_1(\mathbf{a},\mathbf{h})) = \mathbf{0},$$

by Lemma 7.9. It follows therefore that

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

so that $g \circ f$ is differentiable at $\mathbf{a}$ with derivative $g'(f(\mathbf{a})) \circ f'(\mathbf{a})$, as required. ∎

We can use Theorem 7.8 and Theorem 7.10 to produce many examples of differentiable functions.

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

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

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

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

is differentiable everywhere on $\mathbb{R}$, though its derivative $h'\colon \mathbb{R} \to \mathbb{R}$ is not continuous at 0. Also the functions $(x,y) \mapsto x$ and $(x,y) \mapsto y$ are differentiable everywhere on $\mathbb{R}$ (by Lemma 7.7). Now $f(x,y) = y^2 h(x)$. Using Theorem 7.8 and Theorem 7.10, we conclude that f is differentiable everywhere on $\mathbb{R}^2$.

We now summarize the main conclusions of this section. They are as follows.

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

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

The linear transformation $f'(\mathbf{a})$ (if it exists) is unique and is known as the *derivative* (or *total derivative*) of f at $\mathbf{a}$.

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

(iii) There exist functions $f: D \to \mathbb{R}^m$ whose first order partial derivatives are well-defined at a particular point of D but which are not differentiable at that point. Indeed there exist such functions whose first order partial derivatives are well-defined which are not 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 $f: D \to \mathbb{R}^m$ exist *and are continuous* throughout some neighbourhood of a given point then the function is differentiable at that point. (However the converse does not hold: there exist functions which are differentiable whose first order partial derivatives are not continuous.)

(v) Linear transformations are everywhere differentiable.

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

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

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

8 Second Order Partial Derivatives

Let D be an open subset of $\mathbb{R}^n$ and let $f\colon D \to \mathbb{R}$ be a real-valued function on D. We consider the second order partial derivatives of the function f defined by

$$\frac{\partial^2 f}{\partial x_i \partial x_j} \equiv \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right).$$

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

$$\begin{aligned} f_x(x,y) &\equiv \tfrac{\partial f(x,y)}{\partial x}, \\ f_y(x,y) &\equiv \tfrac{\partial f(x,y)}{\partial y}, \\ f_{xy}(x,y) &\equiv \tfrac{\partial^2 f(x,y)}{\partial x \partial y}, \\ f_{yx}(x,y) &\equiv \tfrac{\partial^2 f(x,y)}{\partial y \partial x}. \end{aligned}$$

If $(x,y) \neq (0,0)$ then

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

Similarly

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

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

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

Note that

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

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

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

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

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

does not exist. Indeed

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

Next we show that f_x, f_y, f_{xy} and f_{yx} all exist at $(0, 0)$, and thus exist everywhere on $\mathbb{R}^2$. Now $f(x, 0) = 0$ for all x, hence $f_x(0, 0) = 0$. Also $f(0, y) = 0$ for all y, hence $f_y(0, 0) = 0$. Thus

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

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

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

Thus

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

at $(0,0)$.

Observe that in this example the functions f_{xy} and f_{yx} are continuous throughout $\mathbb{R}^2 \setminus \{(0,0\}$ and are equal to one another there. Although the functions f_{xy} and f_{yx} are well-defined at $(0,0)$, they are not continuous at $(0,0)$ and $f_{xy}(0,0) \neq f_{yx}(0,0)$.

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

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

Theorem 8.1 *Let D be an open set in $\mathbb{R}^2$ and let $f\colon D \to \mathbb{R}$ be a real-valued function on D. Suppose that the partial derivatives*

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

exist and are continuous on D. Then

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

Proof For convenience, we shall denote the values of

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

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

Let (a,b) be a point of D. Then there exists some $\mathbb{R} > 0$ with the property that the open ball of radius R about (a,b) is contained in D, because D is open. Let h and k be real numbers satisfying $h^2 + k^2 < R^2$.

Let us define a differentiable function u by

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

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

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

But

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

and

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

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

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

Thus

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

Now let $\varepsilon > 0$ be given. Then there exists some $\delta_1 > 0$ (where $\delta_1 \leq R$) such that

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

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

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

We can repeat the above argument with the roles of the variables x and y interchanged. Thus on applying the Mean Value Theorem to the function v defined by

$$v(t) = f(a+h, t) - f(a, t)$$

we see that, given sufficiently small h and k, there exists φ_2, where $0 < \varphi_2 < 1$ such that

$$v(b+k) - v(b) = k v'(b+\varphi_2 k),$$

so that

$$\begin{aligned} & f(a+h, b+k) - f(a+h, b) - f(a, b+k) + f(a, b) \\ &\qquad = \quad k\left(f_y(a+h, b+\varphi_2 k) - f_y(a, b+\varphi_2 k)\right). \end{aligned}$$

Applying the Mean Value Theorem to the function $x \mapsto f_y(x, b+\varphi_2 k)$ on the interval $[a, a+h]$ we conclude that there exists some φ_1, where $0 < \varphi_1 < 1$ such that

$$f(a+h, b+k) - f(a+h, b) - f(a, b+k) + f(a, b) = hk\, f_{xy}(a+\varphi_1 h, b+\varphi_2 k).$$

Using the continuity of f_{xy}, we conclude that there exists some $\delta_2 > 0$ such that

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

whenever $h^2 + k^2 < \delta_2^2$.

Take δ to be the minimum of δ_1 and δ_2. If $h^2 + k^2 < \delta_2^2$ then

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

Using the triangle inequality we conclude that

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

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

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

We conclude therefore that

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

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

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

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

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

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

exists at (a,b) and

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

Corollary 8.2 *Let D be an open set in $\mathbb{R}^n$ and let $f: D \to \mathbb{R}$ be a real-valued function on D. 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 D 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.

8.1 Taylor's Theorem for Functions of Several Variables

Let $f: D \to \mathbb{R}$ be a smooth real-valued function defined on an open subset D of $\mathbb{R}^n$. The function f is said to be C^k if and only the partial derivatives

$$\frac{\partial^r f}{\partial x_{i_1} \partial x_{i_2} \cdots \partial x_{i_r}}$$

of order r exist and are continuous for all $r \leq k$. The function f is said to be *smooth* if and only if it is C^k for all positive integers k.

Let $f: D \to \mathbb{R}$ be a C^{k+1} function defined on an open subset D of $\mathbb{R}^n$. Let $\mathbf{a}$ be a point of D. There exists some $R > 0$ such that the open ball of radius R about $\mathbf{a}$ is contained in D on account of the fact that D is open. Let $\mathbf{h}$ be an element of $\mathbb{R}^n$ satisfying $|\mathbf{h}| < R$. Then the line segment joining the points $\mathbf{a}$ and $\mathbf{a} + \mathbf{h}$ is contained in D.

Consider the function $g: [0, 1] \to \mathbb{R}$ defined by $g(t) = f(\mathbf{a} + t\mathbf{h})$. On applying Taylor's Theorem to g, we see that, given any $t \in [0, 1]$ and given any non-negative integer k, there exists some θ satisfying $0 \leq \theta < 1$ such that

$$g(t) = \sum_{j=0}^{k} \frac{t^j}{j!} g^{(j)}(0) + \frac{t^{k+1}}{(k+1)!} g^{(k+1)}(\theta t).$$

In particular, if we set $t = 1$ we see that

$$f(\mathbf{a} + \mathbf{h}) = \sum_{j=0}^{k} \frac{g^{(j)}(0)}{j!} + \frac{g^{(k+1)}(\theta)}{(k+1)!}$$

for some θ satisfying $0 < \theta < 1$. Now

$$\begin{aligned}
g(t) &= f(\mathbf{a} + t\mathbf{h}), \\
g'(t) &= \frac{d}{dt} f(\mathbf{a} + t\mathbf{h}) \\
&= \sum_{i=1}^{n} h_i \left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\mathbf{a}+t\mathbf{h}}, \\
g''(t) &= \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{a}+t\mathbf{h}}, \\
&\vdots \\
g^{(k)}(t) &= \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_k=1}^{n} h_{i_1} h_{i_2} \ldots h_{i_k} \left. \frac{\partial^k f}{\partial x^{i_1} \partial x^{i_2} \cdots \partial x^{i_k}} \right|_{\mathbf{x}=\mathbf{a}+t\mathbf{h}},
\end{aligned}$$

where $\mathbf{h} = (h_1, h_2, \ldots, h_n)$. Thus

$$\begin{aligned}
&f(\mathbf{a} + \mathbf{h}) \\
&= f(\mathbf{a}) + \sum_{i=1}^{n} h_i \left. \frac{\partial f}{\partial x_i} \right|_{\mathbf{x}=\mathbf{a}} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} h_i h_j \left. \frac{\partial^2 f}{\partial x_i \partial x_j} \right|_{\mathbf{x}=\mathbf{a}} \\
&\quad + \cdots + \frac{1}{k!} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_k=1}^{n} h_{i_1} h_{i_2} \ldots h_{i_k} \left. \frac{\partial^k f}{\partial x^{i_1} \partial x^{i_2} \cdots \partial x^{i_k}} \right|_{\mathbf{x}=\mathbf{a}} \\
&\quad + \frac{1}{(k+1)!} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_{(k+1)}=1}^{n} h_{i_1} h_{i_2} \ldots h_{i_{(k+1)}} \left. \frac{\partial^{(k+1)} f}{\partial x^{i_1} \partial x^{i_2} \cdots \partial x^{i_{(k+1)}}} \right|_{\mathbf{x}=\mathbf{a}+\theta\mathbf{h}}
\end{aligned}$$

for some θ satisfying $0 < \theta < 1$. This is the form of Taylor's Theorem applicable to functions defined over open subsets of $\mathbb{R}^n$.

In particular, if $f \colon D \to \mathbb{R}$ is a C^2 function then

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

for some θ satisfying $0 < \theta < 1$.

8.2 Maxima and Minima

Let $f \colon D \to \mathbb{R}$ be a C^2 real-valued function defined over some open subset D of $\mathbb{R}^n$. Suppose that f has a local minimum at some point $\mathbf{a}$ of D, where $\mathbf{a} = (a_1, a_2, \ldots, a_n)$. Now for each integer i between 1 and n the map

$$t \mapsto f(a_1, \ldots, a_{i-1}, t, a_{i+1}, \ldots, a_n)$$

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

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

Let $R > 0$ be chosen such that the open ball of radius R about $\mathbf{a}$ is contained in D. It follows from Taylor's theorem that if $|\mathbf{h}| < R$ then

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

for some θ satisfying $0 < \theta < 1$. Let us denote by $(H_{ij}(\mathbf{a}))$ the *Hessian matrix* at the point $\mathbf{a}$, defined by

$$H_{ij}(\mathbf{a}) = \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\mathbf{x}=\mathbf{a}}.$$

If f is C^2 then the second order partial derivatives of f are continuous, hence $H_{ij}(\mathbf{a}) = Hji(\mathbf{a})$ for all i and j, by Corollary 8.2. Thus the Hessian matrix is symmetric.

We now recall some facts concerning symmetric matrices.

Let (c_{ij}) be a symmetric $n \times n$ matrix.

The matrix (c_{ij}) is said to be *positive semi-definite* if $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} h_i h_j \geq 0$ for all $(h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$.

The matrix (c_{ij}) is said to be *positive definite* if $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} h_i h_j > 0$ for all non-zero $(h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$.

The matrix (c_{ij}) is said to be *negative semi-definite* if $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} h_i h_j \leq 0$ for all $(h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$.

The matrix (c_{ij}) is said to be *negative definite* if $\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} h_i h_j < 0$ for all non-zero $(h_1, h_2, \ldots, h_n) \in \mathbb{R}^n$.

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

Lemma 8.3 *Let (c_{ij}) 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_{ij}) satisfy the inequality $|b_{ij} - c_{ij}| < \varepsilon$ then the matrix (b_{ij}) 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_{ij}) is positive definite if and only if

$$\sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} h_i h_j > 0$$

for all $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. Now the matrix (c_{ij}) is positive definite, by assumption. Therefore

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} h_i h_j > 0$$

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 3.2 that there exists some $(k_1, k_2, \ldots, k_n) \in S^{n-1}$ with the property that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} h_i h_j \geq \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} k_i k_j$$

for all $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. Thus there exists a strictly positive constant $A > 0$ with the property that

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

for all $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. Set $\varepsilon = A/n^2$. If (b_{ij}) is a symmetric $n \times n$ matrix all of whose components satisfy $|b_{ij} - c_{ij}| < \varepsilon$ then

$$\left| \sum_{i=1}^{n} \sum_{j=1}^{n} (b_{ij} - c_{ij}) h_i h_j \right| < \varepsilon n^2 = A,$$

for all $(h_1, h_2, \ldots, h_n) \in S^{n-1}$, hence

$$\sum_{i=1}^{n} \sum_{j=1}^{n} b_{ij} h_i h_j > \sum_{i=1}^{n} \sum_{j=1}^{n} c_{ij} h_i h_j - A \geq 0$$

for all $(h_1, h_2, \ldots, h_n) \in S^{n-1}$. Thus the matrix (b_{ij}) is positive-definite, as required. ∎

Using the fact that a symmetric $n \times n$ matrix (c_{ij}) is negative definite if and only if the matrix $(-c_{ij})$ is positive-definite, we see that if (c_{ij}) 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_{ij}) satisfy the inequality $|b_{ij} - c_{ij}| < \varepsilon$ then the matrix (b_{ij}) is negative definite.

Let $f: D \to \mathbb{R}$ be a C^2 function defined on an open subset D of $\mathbb{R}^n$. Let $\mathbf{a}$ be a point of D. We have already observed that if the function f has a local maximum or a local minimum at $\mathbf{a}$ then

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

We now apply Taylor's theorem to study the behaviour of the function f around a point $\mathbf{a}$ at which the first order partial derivatives vanish. We consider the Hessian matrix $(H_{ij}(\mathbf{a})$ defined by

$$H_{ij}(\mathbf{a}) = \left.\frac{\partial^2 f}{\partial x_i \partial x_j}\right|_{\mathbf{x}=\mathbf{a}}.$$

Lemma 8.4 *Let $f: D \to \mathbb{R}$ be a C^2 function defined on an open subset D of $\mathbb{R}^n$. Let $\mathbf{a}$ be a point of D at which*

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

If f has a local minimum at a point $\mathbf{a}$ of D then the Hessian matrix $(H_{ij}(\mathbf{a}))$ at $\mathbf{a}$ is positive semi-definite.

Proof The first order partial derivatives of f vanish at $\mathbf{a}$. 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

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

where

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

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

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

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

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

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

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

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

Let $f\colon D \to \mathbb{R}$ be as C^2 function defined on an open set D in $\mathbb{R}^n$ and let $\mathbf{a}$ 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 $\mathbf{a}$. However the fact that the Hessian matrix of f is positive semi-definite at $\mathbf{a}$ is noit sufficient to ensure that f is has a local minimum at $\mathbf{a}$, as the following example shows.

Example Consider the function $f\colon \mathbb{R}^2 \to \mathbb{R}$ 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 8.5 *Let $f\colon D \to \mathbb{R}$ be a C^2 function defined on an open subset D of $\mathbb{R}^n$. Let $\mathbf{a}$ be a point of D at which*

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

Suppose that the Hessian matrix $H_{ij}(\mathbf{a})$ at $\mathbf{a}$ is positive definite. Then f has a local minimum at $\mathbf{a}$.

Proof The first order partial derivatives of f vanish at $\mathbf{a}$. 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

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

where

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

Suppose that the Hessian matrix $(H_{ij}(\mathbf{a}))$ is positive definite. It follows from Lemma 8.3 that there exists some $\varepsilon > 0$ such that if $|H_{ij}(\mathbf{x}) - H_{ij}(\mathbf{a})| < \varepsilon$ for all i and j then $(H_{ij}(\mathbf{x}))$ 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 $|H_{ij}(\mathbf{x}) - H_{ij}(\mathbf{a})| < \varepsilon$ whenever $|\mathbf{x} - \mathbf{a}| < \delta$. Thus if $|\mathbf{h}| < \delta$ then $(H_{ij}(\mathbf{a}+\theta\mathbf{h}))$ is positive definite for all $\theta \in (0,1)$ so that $f(\mathbf{a}+\mathbf{h}) > f(\mathbf{a})$. Thus $\mathbf{a}$ is a local minimum of f. ∎

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

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

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

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

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

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

9 Differential Forms on Euclidean Space

9.1 Permutations

We recall some basic facts concerning permutations. Let A be a finite set. A *permutation* π of A is a bijection $\pi\colon A \to A$ which maps A onto itself. The collection of all permutations of the set A forms a group under the operation of composition. We denote this group by S_A.

A permutation $\pi\colon A \to A$ of a finite set A is said to be a *transposition* if there exist elements a_i and a_j of A such that

$$\pi(a) = \begin{cases} a_j & \text{if } a = a_i; \\ a_i & \text{if } a = a_j; \\ a & \text{if } a \in A \setminus \{a_i, a_j\}. \end{cases}$$

Every permutation of a finite set A can be expressed as a composition of transpositions (i.e., the group S_A is generated by the set of all transpositions).

Let π be a permutation of the finite set A. Suppose that π can be expressed as the product of r transpositions and also as a product of s transpositions, where r and s are non-negative integers. A basic result in the theory of permutations then states that $r - s$ is an even integer, so that either r and s are both even or else r and s are both odd. We say that the permutation π of A is *even* if π can be expressed as the composition of an even number of transpositions. Similarly a permutation π of A is said to be *odd* if π can be expressed as the composition of an odd number of transpositions. Every permutation of a finite set A is either even or odd, and no permutation can be both even and odd.

Given a permutation π of a finite set A, let us define the *parity* ϵ_π of π by

$$\epsilon_\pi = \begin{cases} +1 & \text{if } \pi \text{ is even}; \\ -1 & \text{if } \pi \text{ is odd}. \end{cases}$$

Observe that if π can be expressed as the composition of p transpositions then $\epsilon_\pi = (-1)^p$.

If π and σ are permutations of A then $\epsilon_{\pi\circ\sigma} = \epsilon_\pi \epsilon_\sigma$. Thus if we regard $\{1, -1\}$ as a group under the operation of multiplication then the function sending $\pi \in S_A$ to $\epsilon_\pi \in \{1, -1\}$ is a homomorphism from the group S_A of permutations of A to the group $\{1, -1\}$.

Let A and B be finite sets with $A \cap B = \emptyset$. Let σ be a permutation of A and let τ be a permutation of B. Then these permutations induce a permutation π of $A \cup B$, where

$$\pi(x) = \begin{cases} \sigma(x) & \text{if } x \in A; \\ \tau(x) & \text{if } x \in B. \end{cases}$$

It follows directly from the definition of even and odd permutations that $\epsilon_\pi = \epsilon_\sigma \epsilon_\tau$, for if σ can be expressed as a product of s transpositions of the set A and τ can be expressed as a product of t transpositions of the set B then clearly π can be expressed as a product of $s + t$ transpositions of the set $A \cup B$, so that

$$\epsilon_\pi = (-1)^{s+t} = (-1)^s (-1)^t = \epsilon_\sigma \epsilon_\tau.$$

9.2 Differential Forms on n-dimensional Euclidean Space

Definition Let D be an open subset of $\mathbb{R}^n$. Let $(x_1, x_2, \ldots, x_n)$ denote the standard coordinate system on $\mathbb{R}^n$. Let p be a positive integer. We define a *differential form* of degree p to be a sum of expressions of the form

$$f(\mathbf{x})\, dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p},$$

where $i_1, i_2, \ldots, i_p$ are integers between 1 and n, where $f \colon D \to \mathbb{R}$ is a real-valued function on D and where these expressions are subject to the following rules:

Rule I:

$$f(\mathbf{x})\, dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p} = 0$$

unless $i_1, i_2, \ldots, i_p$ are all distinct integers between 1 and n.

Rule II: if $j_m = \pi(i_m)$ for $m = 1, 2, \ldots, p$, where π is a permutation of the set $\{i_1, i_2, \ldots, i_p\}$ then

$$f(\mathbf{x})\, dx_{j_1} \wedge dx_{j_2} \wedge \cdots \wedge dx_{j_p} = \epsilon_\pi f(\mathbf{x})\, dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p},$$

where ϵ_π is the *parity* of the permutation π, defined by

$$\epsilon_\pi = \begin{cases} +1 & \text{if } \pi \text{ is an even permutation;} \\ -1 & \text{if } \pi \text{ is an odd permutation.} \end{cases}$$

Two differential forms of degree p are regarded as being equivalent if and only if the expressions representing these differential forms can be transformed into each other by repeated applications of Rule I and Rule II.

By convention, we define a *differential form of degree 0* on D to be a real-valued function from D to $\mathbb{R}$.

The sum $\omega + \eta$ two differential forms ω and η of degree p is well-defined and is a differential form of degree p. Also if ω is a p-form on D and f is a real-valued function on D then we can multiply the p-form ω by the function f to get a p-form $f\omega$ on D.

If ω is a differential form of degree p for some non-negative integer p then we say that ω is a *p-form*.

Example *(Differential forms on $\mathbb{R}$).* Let I be an open interval in $\mathbb{R}$. A 0-form on I is a real-valued function on I, and a 1-form on I is an expression of the form $f(x)\,dx$, where f is a real-valued function on I. There are no non-zero p-forms on I for $p > 1$ on account of Rule I (since Rule I implies that $dx \wedge dx = 0$, $dx \wedge dx \wedge dx = 0$ etc.).

Example *(Differential forms on $\mathbb{R}^2$).* Let D be an open set in $\mathbb{R}^2$. Let (x, y) denote the standard coordinate system on $\mathbb{R}^2$. A 0-form on D is a real-valued function on D. A 1-form on D is an expression of the form

$$f(x, y)\,dx + g(x, y)\,dy,$$

where f and g are real-valued functions on D. A 2-form on D can be expressed in the form $h(x, y)\,dx \wedge dy$, where h is a real-valued function on D. Rule II tells us that

$$h(x, y)\,dy \wedge dx = -h(x, y)\,dx \wedge dy.$$

There are no non-zero p-forms on D for $p > 2$ on account of Rule I (since Rule I implies that $dx \wedge dx \wedge dx = 0$, $dx \wedge dx \wedge dy = 0$ etc.).

Example *(Differential forms on $\mathbb{R}^3$).* Let D be an open set in $\mathbb{R}^3$. Let (x, y, z) denote the standard coordinate system on $\mathbb{R}^3$. A 0-form on D is a real-valued function on D. A 1-form on D is an expression of the form

$$f(x, y, z)\,dx + g(x, y, z)\,dy + h(x, y, z)\,dz,$$

where f, g and h are real-valued functions on D. A 2-form on D can be expressed in the form

$$f(x, y, z)\,dy \wedge dz + g(x, y, z)\,dx \wedge dz + h(x, y, z)\,dx \wedge dy,$$

where f, g and h are real-valued functions on D. Note that

$$dx \wedge dy = -dy \wedge dx. \quad dy \wedge dz = -dz \wedge dy. \quad dz \wedge dx = -dx \wedge dz.$$

A 3-form on D can be expressed in the form $h(x, y, z)\, dx \wedge dy \wedge dz$, where h is a real-valued function on D. Rule II tells us that

$$\begin{aligned} dx \wedge dy \wedge dz &= dy \wedge dz \wedge dx = dz \wedge dx \wedge dy \\ &= -dz \wedge dy \wedge dx = -dx \wedge dz \wedge dy = -dy \wedge dx \wedge dz. \end{aligned}$$

There are no non-zero p-forms on D for $p > 3$ on account of Rule I.

Example *(Differential forms on $\mathbb{R}^4$).* Let D be an open set in $\mathbb{R}^4$. Let (x, y, z, t) denote the standard coordinate system on $\mathbb{R}^4$. A 0-form on D is a real-valued function on D. A 1-form on D is an expression of the form

$$f\, dx + g\, dy + h\, dz + k\, dt,$$

where f, g h and k are real-valued functions on D. A 2-form on D can be expressed in the form

$$f\, dy \wedge dz + g\, dx \wedge dz + h\, dx \wedge dy + k\, dx \wedge dt + l\, dy \wedge dt + m\, dz \wedge dt,$$

where f, g, h, k, l and m are real-valued functions on D. Note that

$$dx \wedge dy = -dy \wedge dx. \quad dy \wedge dz = -dz \wedge dy. \quad dz \wedge dx = -dx \wedge dz.$$

$$dt \wedge dx = -dx \wedge dt. \quad dt \wedge dy = -dy \wedge dt. \quad dt \wedge dz = -dz \wedge dt.$$

A 3-form on D can be expressed in the form

$$f\, dy \wedge dz \wedge dt + g\, dx \wedge dz \wedge dt + h\, dx \wedge dy \wedge dt + k\, dx \wedge dy \wedge dz,$$

where f, g h and k are real-valued functions on D. A 4-form on D can be expressed in the form $h\, dx \wedge dy \wedge dz \wedge dt$, where h is a real-valued function on D. There are no non-zero p-forms on D for $p > 4$ on account of Rule I.

The following two results follow immediately on applying Rules I and II given above.

Lemma 9.1 *Let D be an open subset of $\mathbb{R}^n$. Then there are no non-zero p-forms on D for $p > n$. Every n-form on D is of the form*

$$f(x_1, x_2, \ldots, x_n)\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n,$$

where $(x_1, x_2, \ldots, x_n)$ is the standard coordinate system on $\mathbb{R}^n$.

Lemma 9.2 *Let D be an open subset of $\mathbb{R}^n$ and let p be an integer between 1 and n. Let $(x_1, x_2, \ldots, x_n)$ denote the standard coordinate system on $\mathbb{R}^n$. Then every p-form on D can be expressed uniquely in the form*

$$\sum_{1 \le i_1 < i_2 < \cdots < i_p \le n} f_{i_1 i_2 \ldots i_p}\, dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p},$$

where $f_{i_1 i_2 \ldots i_p}$ is a real-valued function on D for each p-tuple $(i_1, i_2, \ldots, i_p)$ of integers between 1 and n which satisfy $1 \le i_1 < i_2 < \cdots < i_p \le n$.

9.3 The Wedge Product

We shall define the *wedge product* (or *exterior product*) of two differential forms on $\mathbb{R}^n$. Let $(x_1, x_2, \ldots x_n)$ denote the standard coordinate system on $\mathbb{R}^n$. Let ω be a p-form and let η be a q-form on some open set D in $\mathbb{R}^n$, where $p > 0$ and $q > 0$. Then ω and η can be expressed as finite sums of the form

$$\begin{aligned} \omega &= \sum_{(i_1,i_2,\ldots,i_p)\in I} f_{i_1 i_2 \ldots i_p}\, dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}, \\ \eta &= \sum_{(j_1,j_2,\ldots,j_q)\in J} g_{j_1 j_2 \ldots j_q}\, dx_{j_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{j_q}, \end{aligned}$$

where each $f_{i_1 i_2 \ldots i_p}$ and $g_{j_1 j_2 \ldots j_q}$ is a real valued function on D. (Here the p-tuple $(i_1, i_2, \ldots, i_p)$ and the q-tuple $(j_1, j_2, \ldots, j_q)$ of integers between 1 and n range over suitable indexing sets I and J respectively.) The wedge product $\omega \wedge \eta$ of ω and η is defined by

$$\omega \wedge \eta = \sum_{(i_1,i_2,\ldots,i_p)\in I} \sum_{(j_1,j_2,\ldots,j_q)\in J} f_{i_1 i_2 \ldots i_p} g_{j_1 j_2 \ldots j_q}\, dx_{i_1} \wedge \cdots \wedge dx_{i_p} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_q}.$$

This defines the wedge product $\omega \wedge \eta$ of a p-form ω and a q-form η in the case when $p > 0$ and $q > 0$. If f and g are real-valued functions on D then we define $f \wedge g$ to be the product $f.g$ of f and g. This defines the wedge product of two 0-forms on D. (Recall that the 0-forms on D are by definition the real-valued functions on D.) Similarly if f is a real-valued function on D and if ω is a p-form for some $p > 0$ then we define

$$f \wedge \omega = \omega \wedge f = f\omega.$$

However, in order to check that the wedge product of two differential forms is indeed well-defined, we must verify that the definition of the wedge product that we have given is consistent with Rule I and Rule II in the definition of differential forms.

Let $i_1, i_2, \ldots, i_p$ and $j_1, j_2, \ldots, j_q$ be integers between 1 and n. It follows from Rule I in the definition of differential forms given above that

$$dx_{i_1} \wedge \cdots \wedge dx_{i_p} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_q} = 0$$

unless $i_1, i_2, \ldots, i_p$ and $j_1, j_2, \ldots, j_q$ are all distinct. Thus the definition of the wedge product of differential forms is consistent with Rule I. Now suppose that the integers $i_1, i_2, \ldots, i_p$ and $j_1, j_2, \ldots, j_q$ are all distinct. Let σ be a permutation of the set

$$\{i_1, i_2, \ldots, i_p\}$$

and let τ be a permutation of the set

$$\{j_1, j_2, \ldots, j_q\}.$$

Define $k_1, k_2, \ldots, k_p$ and $l_1, l_2, \ldots, l_q$ by

$$k_r = \sigma(i_r), \qquad (r = 1, 2, \ldots, p),$$

$$l_r = \tau(j_r), \qquad (r = 1, 2, \ldots, q),$$

Then

$$dx_{k_1} \wedge \cdots \wedge dx_{k_p} = \epsilon_\sigma \, dx_{i_1} \wedge \cdots \wedge dx_{i_p},$$
$$dx_{l_1} \wedge \cdots \wedge dx_{l_q} = \epsilon_\tau \, dx_{j_1} \wedge \cdots \wedge dx_{j_q},$$

where ϵ_σ and ϵ_τ are the parities of the permutations σ and τ. Consider the permutation π of the set

$$\{i_1, i_2, \ldots, i_p, j_1, \ldots, j_q\}$$

defined by

$$\pi(m) = \begin{cases} \sigma(m) & \text{if } m \in \{i_1, i_2, \ldots, i_p\}; \\ \tau(m) & \text{if } m \in \{j_1, j_2, \ldots, j_q\}. \end{cases}$$

The parity ϵ_π of this permutation is given by $\epsilon_\pi = \epsilon_\sigma \epsilon_\tau$. Therefore

$$dx_{k_1} \wedge \cdots \wedge dx_{k_p} \wedge dx_{l_1} \wedge \cdots \wedge dx_{l_q} = \epsilon_\sigma \epsilon_\tau \, dx_{i_1} \wedge \cdots \wedge dx_{i_p} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_q}.$$

This shows that the definition of the wedge product is consistent with Rule II. Thus the wedge product of two differential forms is indeed well-defined.

Example Let ω and η be 1-forms on an open subset D of $\mathbb{R}^2$. Let (x, y) denote the standard coordinate system on $\mathbb{R}^2$. We can write

$$\omega = f_1(x, y)\, dx + f_2(x, y)\, dy, \qquad \eta = g_1(x, y)\, dx + g_2(x, y)\, dy,$$

where f_1, f_2, g_1 and g_2 are real-valued functions on D. Then

$$\begin{aligned} \omega \wedge \eta &= (f_1(x, y)\, dx + f_2(x, y)\, dy) \wedge (g_1(x, y)\, dx + g_2(x, y)\, dy) \\ &= f_1(x, y) g_1(x, y)\, dx \wedge dx + f_1(x, y) g_2(x, y)\, dx \wedge dy \\ &\quad + f_2(x, y) g_1(x, y)\, dy \wedge dx + f_2(x, y) g_2(x, y)\, dy \wedge dy. \end{aligned}$$

We simplify these expressions using Rule I and Rule II in the definition of differential forms. Note that

$$dx \wedge dx = 0, \qquad dy \wedge dy = 0,$$

by Rule I, and

$$dy \wedge dx = -dx \wedge dy$$

by Rule II. Therefore

$$\omega \wedge \eta = (f_1(x,y) g_2(x,y) - f_2(x,y) g_1(x,y))\, dx \wedge dy.$$

Note that $\omega \wedge \omega = 0$.

Example Let ω be a 1-form and let σ be a 2-form on an open subset D of $\mathbb{R}^3$. Let (x,y,z) denote the standard coordinate system on $\mathbb{R}^2$. We can write

$$\begin{aligned} \omega &= f_1(x,y,z)\, dx + f_2(x,y,z)\, dy + f_3(x,y,z)\, dz \\ \sigma &= g_1(x,y,z)\, dy \wedge dz + g_2(x,y,z)\, dz \wedge dx + g_3(x,y,z)\, dx \wedge dy, \end{aligned}$$

where f_1, f_2, f_3, g_1, g_2 and g_3 are real-valued functions on D. Then

$$\begin{aligned} \omega \wedge \sigma &= f_1.g_1\, dx \wedge dy \wedge dz + f_1.g_2\, dx \wedge dz \wedge dx + f_1.g_3\, dx \wedge dx \wedge dy \\ &\quad + f_2.g_1\, dy \wedge dy \wedge dz + f_2.g_2\, dy \wedge dz \wedge dx + f_2.g_3\, dy \wedge dx \wedge dy \\ &\quad + f_3.g_1\, dz \wedge dy \wedge dz + f_3.g_2\, dz \wedge dz \wedge dx + f_3.g_3\, dz \wedge dx \wedge dy. \end{aligned}$$

We simplify this expression using Rule I and Rule II in the definition of differential forms. Now

$$dx \wedge dy \wedge dz = dy \wedge dz \wedge dx = dz \wedge dx \wedge dy,$$

by Rule II. Thus

$$\omega \wedge \sigma = (f_1.g_1 + f_2.g_2 + f_3.g_3)\, dx \wedge dy \wedge dz.$$

Similarly one can show that $\sigma \wedge \omega = 0$.

Lemma 9.3 *Let D be an open set in $\mathbb{R}^n$. Then*

(i) $(\omega_1 + \omega_2) \wedge \eta = \omega_1 \wedge \eta + \omega_2 \wedge \eta$ *for all p-forms ω_1 and ω_2 and q-forms η on D,*

(ii) $\omega \wedge (\eta_1 + \eta_2) = \omega \wedge \eta_1 + \omega \wedge \eta_2$ *for all p-forms ω and q-forms η_1 and η_2 on D,*

(iii) $(f\omega) \wedge \eta = f.(\omega \wedge \eta) = \omega \wedge (f\eta)$ *for all differential forms ω and η and real-valued functions f on D,*

(iv) $(\omega \wedge \eta) \wedge \sigma = \omega \wedge (\eta \wedge \sigma)$ *for all differential forms ω, η and σ on D,*

(v) $\omega \wedge \eta = (-1)^{pq} \eta \wedge \omega$ *for all p-forms* ω *and q-forms* η *on* D,

(vi) *if* p *is odd then* $\omega \wedge \omega = 0$ *for all p-forms* ω *on* D.

Proof Properties (i), (ii), (iii) and (iv) follow directly from the definition of the wedge product. To prove (v) we show that

$$dx_{j_1} \wedge \cdots \wedge dx_{j_q} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_p} = (-1)^{pq} dx_{i_1} \wedge \cdots \wedge dx_{i_p} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_q}.$$

Property (v) then follows directly from this identity and the definition of the wedge product.

Now

$$dx_{j_1} \wedge \cdots \wedge dx_{j_q} \wedge dx_{i_m} = (-1)^q dx_{i_m} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_q}$$

for each i_m, since we can interchange dx_{i_m} with $dx_{j_1} \wedge \cdots \wedge dx_{j_q}$ by making q transpositions. We have to perform this operation p times in order to swap each of dx_{i_1}, $dx_{1_2}, \ldots,$ dx_{i_p} past

$$dx_{j_1} \wedge \cdots \wedge dx_{j_q}.$$

Therefore

$$dx_{j_1} \wedge \cdots \wedge dx_{j_q} \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_p} = (-1)^{pq} dx_{i_1} \wedge \cdots \wedge dx_{i_p} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_q}.$$

Property (v) follows directly from this.

It follows immediately from (v) that $\omega \wedge \omega = -\omega \wedge \omega$ if the degree p of ω is odd. Thus if the degree of ω is odd then $\omega \wedge \omega = 0$. This proves (vi). ∎

9.4 The Exterior Derivative

Let ω be a p-form on an open subset D of $\mathbb{R}^n$. We say that ω is differentiable if and only if ω can be expressed as a sum of expressions of the form

$$f(\mathbf{x})\, dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p},$$

where the function f is differentiable on D. Similarly we say that ω is C^k or is smooth if and only if ω can be expressed as a sum of expressions of this form where the function f is C^k or is smooth. (Recall that a function f is said to be C^k on D if and only if all its partial derivatives of all orders less than or equal to k exist *and are continuous* on D. A function is said to be *smooth* if and only if it is C^k for all non-negative integers k.)

Definition Let $f\colon D \to \mathbb{R}$ be a smooth real-valued function on D. The *differential* df of f is defined to be the 1-form on D defined by

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} dx_i.$$

Observe that if f and g are differentiable real-valued functions on D then

$$d(f+g) = df + dg, \qquad d(f.g) = g\,df + f\,dg$$

(where we have used the Product Rule for differentiation in deriving the second of these identities).

Definition Let ω be a differentiable p-form for some $p > 0$, where ω is specified by a sum of the form

$$\omega = \sum_{(i_1,i_2,\ldots,i_p)\in I} f_{i_1 i_2 \ldots i_p}\, dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}$$

where $f_{i_1 i_2 \ldots i_n}$ is a differentiable function for each p-tuple $(i_1, i_2, \ldots, i_p)$ belonging to I. The *exterior derivative* $d\omega$ of ω is defined to be the $(p+1)$-form on D defined by

$$d\omega = \sum_{(i_1,i_2,\ldots,i_p)\in I} df_{i_1 i_2 \ldots i_p} \wedge dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p},$$

where $df_{i_1 i_2 \ldots i_p}$ is the differential of the function $f_{i_1 i_2 \ldots i_p}$. Thus

$$d\omega = \sum_{(i_1,i_2,\ldots,i_p)\in I} \sum_{j=1}^{n} \frac{\partial f_{i_1 i_2 \ldots i_p}}{\partial x_j}\, dx_j \wedge dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}.$$

If f is a differentiable 0-form on D (i.e. a differentiable real-valued function on D) then the exterior derivative df of f is simply defined to be the differential of f.

Example Let us consider differential forms on $\mathbb{R}^3$. Let (x, y, z) denote the standard coordinate system on $\mathbb{R}^3$. If $f\colon \mathbb{R}^3 \to \mathbb{R}$ is a differentiable function on $\mathbb{R}^3$ then

$$df \equiv \frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy + \frac{\partial f}{\partial z}\, dz.$$

Now any differentiable 1-form on ω on $\mathbb{R}^3$ can be expressed as a sum of the form

$$\omega = P\,dx + Q\,dy + R\,dz.$$

Then

$$
\begin{aligned}
d\omega &= \frac{\partial P}{\partial x}\,dx\wedge dx + \frac{\partial P}{\partial y}\,dy\wedge dx + \frac{\partial P}{\partial z}\,dz\wedge dx \\
&\quad + \frac{\partial Q}{\partial x}\,dx\wedge dy + \frac{\partial Q}{\partial y}\,dy\wedge dy + \frac{\partial Q}{\partial z}\,dz\wedge dy \\
&\quad + \frac{\partial R}{\partial x}\,dx\wedge dz + \frac{\partial R}{\partial y}\,dy\wedge dz + \frac{\partial R}{\partial z}\,dz\wedge dz.
\end{aligned}
$$

We now use the identities

$$dx\wedge dx = dy\wedge dy = dz\wedge dz = 0,$$

$$dx\wedge dy = -dy\wedge dx, \quad dy\wedge dz = -dz\wedge dy, \quad dz\wedge dx = -dx\wedge dz$$

to simplify the expression for $d\omega$. We conclude that

$$d\omega = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\wedge dy + \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy\wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dz\wedge dx.$$

Now let η be a differentiable 2-form on $\mathbb{R}^3$. We can write

$$\eta = U\,dy\wedge dz + V\,dz\wedge dx + W\,dx\wedge dy.$$

where U, V and W are differentiable functions on $\mathbb{R}^3$. Then

$$
\begin{aligned}
d\eta &= \frac{\partial U}{\partial x}\,dx\wedge dy\wedge dz + \frac{\partial U}{\partial y}\,dy\wedge dy\wedge dz + \frac{\partial U}{\partial z}\,dz\wedge dy\wedge dz \\
&\quad + \frac{\partial V}{\partial x}\,dx\wedge dz\wedge dx + \frac{\partial V}{\partial y}\,dy\wedge dz\wedge dx + \frac{\partial V}{\partial z}\,dz\wedge dz\wedge dx \\
&\quad + \frac{\partial W}{\partial x}\,dx\wedge dx\wedge dy + \frac{\partial W}{\partial y}\,dy\wedge dx\wedge dy + \frac{\partial W}{\partial z}\,dz\wedge dx\wedge dy. \\
&= \left(\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} + \frac{\partial W}{\partial z}\right) dx\wedge dy\wedge dz.
\end{aligned}
$$

Here we have used the fact that

$$dx\wedge dy\wedge dz = dy\wedge dz\wedge dx = dz\wedge dx\wedge dy.$$

If σ is a differentiable 3-form on $\mathbb{R}^3$ then $d\sigma$ is a 4-form on $\mathbb{R}^3$. But there are no non-zero 4-forms on $\mathbb{R}^3$ (Lemma 9.1). Therefore $d\sigma = 0$.

Remark Let $\mathbf{E}$ be a differentiable vector field on $\mathbb{R}^3$. We can write $\mathbf{E} = (E_1, E_2, E_3)$, where E_1, E_2 and E_3 are the components of E. Then

$$\begin{aligned} \operatorname{curl}\mathbf{E} &= \left(\frac{\partial E_3}{\partial y} - \frac{\partial E_2}{\partial z}, \frac{\partial E_1}{\partial z} - \frac{\partial E_3}{\partial x}, \frac{\partial E_2}{\partial x} - \frac{\partial E_1}{\partial y}\right), \\ \operatorname{div}\mathbf{E} &= \tfrac{\partial E_1}{\partial x} + \tfrac{\partial E_2}{\partial y} + \tfrac{\partial E_3}{\partial z}. \end{aligned}$$

Thus if ω is the 1-form on $\mathbb{R}^3$ defined by

$$\omega = E_1\, dx + E_2\, dy + E_3\, dz$$

then

$$d\omega = (\operatorname{curl}\mathbf{E})_1\, dy \wedge dz + (\operatorname{curl}\mathbf{E})_2\, dz \wedge dx + (\operatorname{curl}\mathbf{E})_3\, dx \wedge dy,$$

where $(\operatorname{curl}\mathbf{E})_i$ denotes the ith component of $\operatorname{curl}\mathbf{E}$ for $i = 1, 2, 3$. Also, if η is the 2-form on $\mathbb{R}^3$ defined by

$$\eta = E_1\, dy \wedge dz + E_2\, dz \wedge dx + E_3\, dx \wedge dy$$

then

$$d\eta = (\operatorname{div}\mathbf{E})\, dx \wedge dy \wedge dz.$$

Example Let ω be the 2-form on $\mathbb{R}^2$ defined by

$$\omega = \cos x\, \cos y\, dx - \sin x\, \sin y\, dy.$$

where (x, y) is the standard coordinate system on $\mathbb{R}^2$. Then

$$\begin{aligned} d\omega &= \frac{\partial}{\partial y}(\cos x\, \cos y)\, dy \wedge dx - \frac{\partial}{\partial x}(\sin x\, \sin y)\, dx \wedge dy \\ &= -\cos x\, \sin y\, dy \wedge dx - \cos x\, \sin y\, dx \wedge dy \\ &= +\cos x\, \sin y\, dx \wedge dy - \cos x\, \sin y\, dx \wedge dy \\ &= 0. \end{aligned}$$

Note that $\omega = df$, where $f(x, y) = \sin x\, \cos y$.

Lemma 9.4 *Let D be an open subset of $\mathbb{R}^n$, ω, ω_1 and ω_2 be differentiable p-forms on D, let η be a differentiable q-form on M, and let f be a differentiable real-valued function on M. Then*

(i) $d(\omega_1 + \omega_2) = d\omega_1 + d\omega_2$

(ii) $d(f\omega) = f\, d\omega + df \wedge \omega$,

(iii) $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^p \omega \wedge d\eta$, *where p is the degree of ω.*

Proof Property (i) is immediate from the definition of the exterior derivative. In order to verify (ii), it suffices to check that this identity holds in the particular case when ω is of the form

$$\omega = g\, dx_{i_1} \wedge \cdots \wedge dx_{i_p}$$

where g is a differentiable real-valued function on D (since any differentiable p-form on D is a sum of expressions of this form). But then

$$\begin{aligned} d(f\omega) &= d(f.g)\, dx_{i_1} \wedge \cdots \wedge dx_{i_p} \\ &= (f\, dg + g\, df) \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_p} \\ &= f\, d\omega + df \wedge \omega, \end{aligned}$$

as required.

In order to verify (iii), it suffices to check that this identity holds in the particular case when ω and η are of the form

$$\begin{aligned} \omega &= g\, dx_{i_1} \wedge \cdots \wedge dx_{i_p}, \\ \eta &= h\, dx_{j_1} \wedge \cdots \wedge dx_{j_q}. \end{aligned}$$

In this case we see that

$$\omega \wedge \eta = g.h\, dx_{i_1} \wedge \cdots \wedge dx_{i_p} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_q},$$

so that

$$d(\omega \wedge \eta) = (h\, dg + g\, dh) \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_p} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_q}.$$

But

$$dh \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_p} = (-1)^p dx_{i_1} \wedge \cdots \wedge dx_{i_p} \wedge dh,$$

hence

$$\begin{aligned} d(\omega \wedge \eta) &= (dg \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_p}) \wedge (h\, dx_{j_1} \wedge \cdots \wedge dx_{j_q}) \\ &\quad + (-1)^p (g\, dx_{i_1} \wedge \cdots \wedge dx_{i_p}) \wedge (dh \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_q}) \\ &= d\omega \wedge \eta + (-1)^p \omega \wedge d\eta, \end{aligned}$$

as required. This proves (iii). ∎

Lemma 9.5 *Let ω be a C^2 differential form defined over some open subset D of $\mathbb{R}^n$. Then $d(d\omega) = 0$.*

Proof First let us suppose $f\colon D \to \mathbb{R}$ is a C^2 function. Then

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\, dx_i,$$

hence

$$\begin{aligned} d(df) &= \sum_{i=1}^{n} d\left(\frac{\partial f}{\partial x_i}\right) \wedge dx_i \\ &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}\, dx_j \wedge dx_i. \end{aligned}$$

But

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

for all i and j by Corollary 8.2 (since $f\colon D \to \mathbb{R}$ is a C^2 function). Also $dx_j \wedge dx_i = -dx_j \wedge dx_i$ for all i and j. We deduce from these facts that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}\, dx_j \wedge dx_i = 0.$$

Indeed we see that

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}\, dx_j \wedge dx_i &= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\, dx_j \wedge dx_i \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}\, dx_i \wedge dx_j \\ &= -\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}\, dx_j \wedge dx_i \end{aligned}$$

(where the last of these equalities is obtained by interchanging the roles of the indices i and j). We conclude therefore that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_j \partial x_i}\, dx_j \wedge dx_i = 0,$$

and hence $d(df) = 0$.

Now suppose that ω is a p-form given by by an expression of the form

$$\omega = f\, dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p},$$

where $f\colon D \to \mathbb{R}$ is a C^2 function on D. Then

$$d\omega = df \wedge dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p},$$

so that

$$d(d\omega) = d(df) \wedge dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p} - df \wedge d\left(dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}\right),$$

(by Lemma 9.4). But we have shown that $d(df) = 0$. Also

$$d\left(dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}\right) = 0.$$

Therefore $d(d\omega) = 0$. The required result then follows directly from the fact that any p-form which is C^2 is a sum of expressions of the form

$$f\, dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p},$$

where $f\colon D \to \mathbb{R}$ is a C^2 function on D. ∎

Lemma 9.6 *Let D be an open subset of $\mathbb{R}^n$ and let $\tilde{d}$ be a differential operator mapping differentiable p-forms on D to $(p+1)$-forms on D. Suppose that the operator $\tilde{d}$ satisfies the following four conditions:*

(i) $\tilde{d}(\omega_1 + \omega_2) = \tilde{d}\omega_1 + \tilde{d}\omega_2$ *for all differentiable p-forms ω_1 and ω_2 on D,*

(ii) $\tilde{d}(\omega \wedge \eta) = \tilde{d}\omega \wedge \eta + (-1)^p \omega \wedge \tilde{d}\eta$ *for all differentiable p-forms ω and q-forms η on D,*

(iii) $\tilde{d}(\tilde{d}\omega) = 0$ *for all C^2 differential forms ω on D,*

(iv) $\tilde{d}f = df = \dfrac{\partial f}{\partial x_1}\, dx_1 + \dfrac{\partial f}{\partial x_2}\, dx_2 + \cdots + \dfrac{\partial f}{\partial x_n}\, dx_n$ *for all differentiable functions $f\colon D \to \mathbb{R}$ on D.*

Then $\tilde{d}\omega = d\omega$ for all differential forms ω on D that are differentiable on D. Thus these four conditions completely characterize the exterior derivative operator sending differentiable p-forms on D to $(p+1)$-forms on D.

Proof Let $\tilde{x}_i\colon D \to \mathbb{R}$ denote the function sending a point $(x_1, x_2, \ldots, x_n)$ of D to the ith coordinate x_i of that point. Then $\tilde{d}\tilde{x}_i = dx_i$, by (iv). Therefore

$$\tilde{d}(dx_i) = \tilde{d}(\tilde{d}\tilde{x}_i) = 0$$

by (iii). Thus if we apply (ii) then it follows by induction on p that

$$\tilde{d}(dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}) = 0$$

for all $i_1, i_2, \ldots, i_p$. Thus if the p-form ω is defined by

$$\omega = f\,(dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}$$

for some differentiable function $f\colon D \to \mathbb{R}$ then

$$\begin{aligned}\tilde{d}\omega &= \tilde{d}f \wedge dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p} + f\,\tilde{d}(dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}) \\ &= df \wedge dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p} \\ &= d\omega,\end{aligned}$$

by (ii) and (iv). But any differentiable p-form on D can be expressed as a sum of p-forms which are of the form

$$f\,dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}$$

for some differentiable function f on D. The required result therefore follows from (i). ∎

9.5 Pullbacks of Differentiable Forms along Smooth Maps

Let D be an open set in $\mathbb{R}^n$, let E be an open set in $\mathbb{R}^m$, and let $\varphi\colon D\colon E$ be a smooth map. Let us denote by $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_m)$ be the standard coordinate systems on $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. Let $\varphi_j\colon D \to \mathbb{R}$ denote the jth component of the map φ, for $j = 1, 2, \ldots, m$, so that

$$\varphi(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots, \varphi_m(\mathbf{x}))$$

for all $\mathbf{x} \in D$.

Let ω be a p-form on E for some positive integer p. Then ω can be expressed as a sum of the form

$$\omega = \sum_{(j_1, j_2, \ldots, j_p) \in I} f_{j_1 j_2 \ldots j_p}\, dy_{j_1} \wedge dy_{j_2} \wedge \cdots \wedge dy_{j_p}$$

where each $f_{j_1 j_2 \ldots j_n}$ is a real-valued function on E. We define the p-form $\varphi^*\omega$ on D by

$$\varphi^*\omega = \sum_{(j_1, j_2, \ldots, j_p) \in I} (f_{j_1 j_2 \ldots j_p} \circ \varphi)\, d\varphi_{j_1} \wedge d\varphi_{j_2} \wedge \cdots \wedge d\varphi_{j_p}.$$

If f is a real-valued function on E (i.e., a 0-form on E) then we define $\varphi^* f \equiv f \circ \varphi$ (i.e., $\varphi^* f$ is the composition of φ and f). The p-form $\varphi^*\omega$ is referred to as the *pullback* of ω under the map φ.

One must check that the pullback $\varphi^*\omega$ is well-defined. Note that

$$d\varphi_j \wedge d\varphi_j = 0, \qquad d\varphi_j \wedge d\varphi_k = -d\varphi_k \wedge d\varphi_j$$

for all integers j and k between 1 and m, by Lemma 9.3. It follows from this that

$$d\varphi_{j_1} \wedge d\varphi_{j_2} \wedge \cdots \wedge d\varphi_{j_p} = 0$$

unless the integers $j_1, j_2, \ldots, j_p$ are distinct. If these integers are distinct, if π is a permutation of the set $\{j_1, j_2, \ldots, j_p\}$, and if $k_r = \pi(j_r)$ for $r = 1, 2, \ldots, p$, then

$$d\varphi_{k_1} \wedge d\varphi_{k_2} \wedge \cdots \wedge d\varphi_{k_p} = \epsilon_\pi \, d\varphi_{j_1} \wedge d\varphi_{j_2} \wedge \cdots \wedge d\varphi_{j_p},$$

where ϵ_π is $+1$ if the permutation π is even and -1 if the permutation π is odd. (For any permutation of an indexed set can be accomplished by successively transposing adjacent elements of that set. The permutation is evan if it is obtained by an even number of such transpositions, and it is odd if it is obtained by an odd number of transpositions. Hence the result stated above follows from the fact that $d\varphi_j \wedge d\varphi_k = -d\varphi_k \wedge d\varphi_j$ for all j and k.) We conclude therefore that $\varphi^*\omega$ is indeed well-defined (i.e., the formula given above is consistent with Rule I and Rule II in the definition of differential forms).

Example Let (x, y) denote the standard coordinate system on $\mathbb{R}^2$ and let t denote the standard coordinate on $\mathbb{R}$. Consider the smooth curve $\gamma\colon \mathbb{R} \to \mathbb{R}^2$ in $\mathbb{R}^2$ defined by

$$\gamma(t) = (t^2, t^3)$$

Therefore

$$\begin{aligned} \gamma^*(dx) &= d(t^2) = 2t\,dt, \\ \gamma^*(dy) &= d(t^3) = 3t^2\,dt. \end{aligned}$$

Therefore

$$\gamma^*(2x\,dy - 3y\,dx) = 2t^2(3t^2\,dt) - 3t^3(2t\,dt) = 0.$$

Example Let (u, v) denote the standard coordinate system on $\mathbb{R}^2$ and let (x, y, z) denote the standard coordinate system on $\mathbb{R}^3$. Let $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^3$ be the smooth map defined by

$$\varphi(u, v) = (\sin u \, \cos v, \sin u \, \sin v, \cos u)$$

Let ω be the 1-form on $\mathbb{R}^3$ defined by

$$\omega = x\,dx + y\,dy + z\,dz.$$

Now

$$\begin{aligned}
\varphi^* dx &= d(\sin u\ \cos v)\\
&= \frac{\partial(\sin u\ \cos v)}{\partial u}\,du + \frac{\partial(\sin u\ \cos v)}{\partial v}\,dv\\
&= \cos u\ \cos v\,du - \sin u\ \sin v\,dv,
\end{aligned}$$

$$\begin{aligned}
\varphi^* dy &= d(\sin u\ \sin v)\\
&= \frac{\partial(\sin u\ \sin v)}{\partial u}\,du + \frac{\partial(\sin u\ \sin v)}{\partial v}\,dv\\
&= \cos u\ \sin v\,du + \sin u\ \cos v\,dv,
\end{aligned}$$

$$\begin{aligned}
\varphi^* dz &= d(\cos u)\\
&= \frac{\partial(\cos u)}{\partial u}\,du + \frac{\partial(\cos u)}{\partial v}\,dv\\
&= -\sin u\,du.
\end{aligned}$$

Therefore

$$\begin{aligned}
\varphi^*\omega &= \sin u\ \cos v\,(\cos u\ \cos v\,du - \sin u\ \sin v\,dv)\\
&\quad + \sin u\ \sin v\,(\cos u\ \sin v\,du + \sin u\ \cos v\,dv)\\
&\quad + \cos u\,(-\sin u\,du)\\
&= (\sin u\ \cos u\ \cos^2 v + \sin u\ \cos u\ \sin^2 v - \sin u\ \cos u)\,du\\
&\quad + (-\sin^2 u\ \sin v\ \cos v + \sin^2 u\ \sin v\ \cos v)\,dv\\
&= 0.
\end{aligned}$$

If η is the 2-form on $\mathbb{R}^3$ defined by

$$\eta = x\,dy \wedge dz + y\,dz \wedge dx + z\,dx \wedge dy$$

then

$$\begin{aligned}
&\varphi^*\eta\\
&= \sin u\ \sin v(-\sin u\,du) \wedge (\cos u\ \cos v\,du - \sin u\ \sin v\,dv)\\
&\quad + \cos u\,(\cos u\ \cos v\,du - \sin u\ \sin v\,dv) \wedge (\cos u\ \sin v\,du + \sin u\ \cos v\,dv)\\
&\quad + \sin u\ \cos v(\cos u\ \sin v\,du + \sin u\ \cos v\,dv) \wedge (-\sin u\,du)\\
&= \sin^3 u\ \sin^2 v\,du \wedge dv + \cos^2 u\ \sin u\ \cos^2 v\,du \wedge dv\\
&\quad - \cos^2 u\ \sin u\ \sin^2 v\,dv \wedge du - \sin^3 u\ \cos^2 v\,dv \wedge du\\
&= \left(\sin^3 u\,(\sin^2 v + \cos^2 v) + \cos^2 u\ \sin u(\cos^2 v + \sin^2 v)\right)\,du \wedge dv\\
&= \sin u\,(\sin^2 u + \cos^2 u)\,du \wedge dv\\
&= \sin u\,du \wedge dv.
\end{aligned}$$

Lemma 9.7 *Let D be an open set in $\mathbb{R}^n$, let E be an open set in $\mathbb{R}^m$, and let $\varphi\colon D\colon E$ be a smooth map. Let ω, ω_1 and η be p-forms on E and let η be a q form on E for some non-negative integers p and q. Then*

(i) $\varphi^*(\omega_1 + \omega_2) = \varphi^*\omega_1 + \varphi^*\omega_2$ *for all p-forms ω_1 and ω_2 on E,*

(ii) $\varphi^*(\omega \wedge \eta) = \varphi^*\omega \wedge \varphi^*\eta$ *for all differential forms ω and η on E.*

(iii) $\varphi^*(d\omega) = d(\varphi^*\omega)$ *for all differential forms ω on E that are differentiable on E.*

Proof Properties (i) and (ii) follow directly from the definition of the pull-back of a differential form under the smooth map $\varphi\colon D \to E$. In order to verify (iii) we first verify that (iii) holds for all real-valued functions (i.e., 0-forms) on E. Let f be a real-valued function on E. Then

$$df = \sum_{j=1}^{m} \frac{\partial f}{\partial y_j}\, dy_j,$$

where $(y_1, y_2, \ldots, y_m)$ denotes the standard coordinate system on $\mathbb{R}^m$. Therefore

$$\begin{aligned} \varphi^* df &= \sum_{j=1}^{m} \frac{\partial f}{\partial y_j} \circ \varphi\, d\varphi_j \\ &= \sum_{j=1}^{m} \sum_{i=1}^{n} \left(\frac{\partial f}{\partial y_j} \circ \varphi \right) \frac{\partial \varphi_j}{\partial x_i}\, dx_i. \end{aligned}$$

But it follows from the Chain Rule for differentiation (Theorem 7.10) that

$$\frac{\partial (f \circ \varphi)}{\partial x_i} = \sum_{j=1}^{m} \left(\frac{\partial f}{\partial y_j} \circ \varphi \right) \frac{\partial \varphi_j}{\partial x_i}.$$

(For the Chain Rule tells us that $(f \circ \varphi)'(\mathbf{x}) = f'(\varphi(\mathbf{x})) \circ \varphi'(\mathbf{x})$. This means that the Jacobian matrix for $f \circ \varphi$ at a point $\mathbf{x}$ of D is the product of the Jacobian matrices of f at $\varphi(\mathbf{x})$ and of φ at $\mathbf{x}$.) We conclude that

$$\varphi^* df = \sum_{i=1}^{n} \frac{\partial (f \circ \varphi)}{\partial x_i}\, dx_i = d(f \circ \varphi).$$

Now let ω be a p-form on E which is of the form

$$\omega = f\, dy_{j_1} \wedge dy_{j_2} \wedge \cdots \wedge dy_{j_p}$$

for some differentiable function $f\colon E \to \mathbb{R}$. Then

$$d\omega = df \wedge dy_{j_1} \wedge dy_{j_2} \wedge \cdots \wedge dy_{j_p},$$

so that

$$\varphi^*(d\omega) = \varphi^*(df) \wedge \varphi^*(dy_{j_1}) \wedge \varphi^*(dy_{j_2}) \wedge \cdots \wedge \varphi^*(dy_{j_p}),$$

by (ii). But $\varphi^*(df) = d(f \circ \varphi)$ and $\varphi^*(dy_{j_r}) = d\varphi_{j_r}$ for all r. Therefore

$$\varphi^*(d\omega) = d(f \circ \varphi) \wedge d\varphi_{j_1} \wedge d\varphi_{j_2} \wedge \cdots \wedge d\varphi_{j_p}.$$

But $d(d\varphi_j) = 0$ for all j by Lemma 9.5, hence

$$\begin{aligned} d(\varphi^*\omega) &= d\bigl((f \circ \varphi)\, d\varphi_{j_1} \wedge d\varphi_{j_2} \wedge \cdots \wedge d\varphi_{j_p}\bigr) \\ &= d(f \circ \varphi) \wedge d\varphi_{j_1} \wedge d\varphi_{j_2} \wedge \cdots \wedge d\varphi_{j_p} \end{aligned}$$

(where we have applied property (iii) of Lemma 9.4 to derive this identity). Therefore

$$\varphi^*(d\omega) = d(\varphi^*\omega).$$

Property (iii) now follows from this result, together with the fact that any differentiable p-form on E can be expressed as a sum of expressions of the form

$$f\, dy_{j_1} \wedge dy_{j_2} \wedge \cdots \wedge dy_{j_p},$$

where f is a differentiable function from E to $\mathbb{R}$. ∎

10 The Poincaré Lemma

Let D be an open subset of $\mathbb{R}^n$ and let $\mathbf{a}$ be a point of D. We say that D is *star-shaped* with respect to the point $\mathbf{a}$ if and only if the point $(1-t)\mathbf{x} + t\mathbf{a}$ is in D for all $\mathbf{x} \in D$ and $t \in [0,1]$. Thus D is star-shaped with respect to the point $\mathbf{a}$ if and only if, for all $\mathbf{x} \in D$, the line segment joining $\mathbf{a}$ to $\mathbf{x}$ is contained in D.

Let ω be a smooth p-form on an open subset D of $\mathbb{R}^n$. We say that ω is *closed* if and only if $d\omega = 0$. We say that ω is *exact* if and only if there exists a $(p-1)$-form η on D with the property that $\omega = d\eta$. Note that every exact differential form on D is closed, since $d(d\eta) = 0$ for all η.

Lemma 10.1 (The Poincaré Lemma) *Let D be a star-shaped open set in $\mathbb{R}^n$. Let ω be a smooth p-form on D, where $p \geq 1$. Suppose that $d\omega = 0$. Then there exists a $(p-1)$-form η on D such that $\omega = d\eta$.*

In order to prove the Poincaré Lemma, we consider differential forms on $[0,1] \times D$, where D is an open set in $\mathbb{R}^n$. Let $(x_1, x_2, \ldots, x_n)$ denote the standard Cartesian coordinates on D. We denote coordinates on $[0,1] \times D$ by $(t, x_1, x_2, \ldots, x_n)$, where $t \in [0,1]$ and $(x_1, x_2, \ldots, x_n) \in D$. For each $\tau \in [0,1]$, let $\beta_\tau \colon D \to [0,1] \times D$ be the map which sends $(x_1, x_2, \ldots, x_n)$ to $(\tau, x_1, x_2, \ldots, x_n)$. Observe that

$$\beta_\tau^*(dt) = 0, \quad \beta_\tau^*(dx_i) = dx_i \ \ (i = 1, 2, \ldots, n) \tag{1}$$

for all $\tau \in [0,1]$.

We define an operator I which takes (continuous) p-forms on $[0,1] \times D$ to $(p-1)$-forms on D. This operator I is characterized uniquely by the following four properties:

(i) $I(f) = 0$ for all continuous functions $f \colon [0,1] \times D \to \mathbb{R}$,

(ii) $I(g\,dt)(\mathbf{x}) = \displaystyle\int_0^1 g(t, \mathbf{x})\,dt$ for all continuous functions $g \colon [0,1] \times D \to \mathbb{R}$,

(iii) $I(\eta \wedge dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}) = I(\eta) \wedge dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_p}$ for all continuous differential forms η on $[0,1] \times D$,

(iv) If $\eta = \eta_1 + \eta_2$, where η_1 and η_2 are continuous p-forms on $[0,1] \times D$ then $I(\eta) = I(\eta_1) + I(\eta_2)$.

Note that the operator I is defined such that

$$I(f\,dx_{i_1} \wedge \cdots \wedge dx_{i_p}) = 0, \tag{2}$$

$$I(g\,dt \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_p}) = \left(\int_0^1 g(t, \mathbf{x})\,dt \right) \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_p}, \tag{3}$$

where f and g are continuous real-valued functions on $[0,1] \times D$.

We claim that if $p \geq 1$ then

$$dI(\eta) + I(d\eta) = \beta_1^*\eta - \beta_0^*\eta \tag{4}$$

for all smooth p-forms η on $[0,1] \times D$ (where $\beta_\tau(\mathbf{x}) = (\tau, \mathbf{x})$ for all $\tau \in [0,1]$ and $\mathbf{x} \in D$). We now show how the Poincaré Lemma can be derived from this result.

Proof of the Poincaré Lemma, assuming (4). Suppose that the open set D in $\mathbb{R}^n$ is star-shaped with respect to the point $\mathbf{a}$ of D. Define a smooth map $H \colon [0,1] \times D \to D$ by

$$H(t, \mathbf{x}) = t\mathbf{x} + (1-t)\mathbf{a} \qquad (t \in [0,1], \quad \mathbf{x} \in D).$$

Note that $H \circ \beta_1$ is the identity map of D and that $H \circ \beta_0$ maps the whole of D to a single point $\mathbf{a}$ of D. It follows from this that $(H \circ \beta_0)^*(dx_i) = 0$ for $i = 1, 2, \ldots, n$. Thus

$$(H \circ \beta_1)^*\omega = \omega, \qquad (H \circ \beta_0)^*\omega = 0$$

for all p-forms ω on D, where $p \geq 1$. Suppose that ω satisfies $d\omega = 0$. We apply Equation (4) to the pullback $H^*\omega$ of ω under the map H to deduce that

$$d(I(H^*\omega)) + I(d(H^*\omega)) = \beta_1^*(H^*\omega) - \beta_0^*(H^*\omega) = (H \circ \beta_1)^*\omega - (H \circ \beta_0)^*\omega = \omega$$

by Equation (4). Thus if $d\omega = 0$ then $d(H^*\omega) = H^*(d\omega)$ by Lemma 9.7, so that $\omega = d\eta$, where $\eta = I(H^*\omega)$. This proves the Poincaré Lemma, given Equation (4).

In order to complete the proof of the Poincaré Lemma we must prove Equation (4). Let $f\colon [0,1] \times D \to \mathbb{R}$ be smooth. Then

$$df = \frac{\partial f}{\partial t}\, dt + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\, dx_i.$$

On applying the four properties listed above which characterize the operation I, we see that

$$I(df)(\mathbf{x}) = \int_0^1 \frac{\partial f(t,\mathbf{x})}{\partial t}\, dt = f(1,\mathbf{x}) - f(0,\mathbf{x}),$$

so that

$$I(df) = f \circ \beta_1 - f \circ \beta_0 \tag{5}$$

We now show that (4) holds for all smooth 1-forms η on $[0,1] \times D$. We can express η as a sum of the form

$$\eta = g\, dt + \sum_{i=1}^{n} f_i\, dx_i$$

where $g, f_1, \ldots, f_n$ are smooth real-valued functions on $[0,1] \times D$. Then

$$I(\eta)(\mathbf{x}) = \int_0^1 g(t,\mathbf{x})\, dt,$$

so that

$$\begin{aligned} dI(\eta) &= \sum_{i=1}^n \frac{\partial I(\eta)}{\partial x_i}\, dx_i = \sum_{i=1}^n \frac{\partial}{\partial x_i} \left(\int_0^1 g\, dt \right) dx_i \\ &= \sum_{i=1}^n \left(\int_0^1 \frac{\partial g}{\partial x_i}\, dt \right) dx_i = \sum_{i=1}^n I \left(\frac{\partial g}{\partial x_i}\, dt \right) dx_i \\ &= I \left(\sum_{i=1}^n \frac{\partial g}{\partial x_i}\, dt \wedge dx_i \right) = -I \left(\sum_{i=1}^n \frac{\partial g}{\partial x_i}\, dx_i \wedge dt \right) \\ &= -I(dg \wedge dt). \end{aligned}$$

Thus

$$\begin{aligned} I(d\eta) + dI(\eta) &= I(dg \wedge dt) + \sum_{i=1}^n I(df_i \wedge dx_i) + dI(\eta) = \sum_{i=1}^n I(df_i) \wedge dx_i \\ &= \sum_{i=1}^n (f_i \circ \beta_1 - f_i \circ \beta_0)\, dx_i = (\beta_1^* \eta) - (\beta_0^* \eta). \end{aligned}$$

(where we have used Equation (5)). This verifies (4) in the particular case when η is a smooth 1-form on $[0,1] \times D$.

In order to prove (4) when $p > 1$, we simply use the fact that every smooth p-form on $[0,1] \times D$ can be expressed as a sum of p-forms of the form

$$\sigma \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_{p-1}}$$

where σ is a smooth 1-form on $[0,1] \times D$. (This follows from the fact that in any non-zero wedge product of terms of the form dt or dx_i, at most one of the factors of the wedge product will be dt, and we may suppose that this is the first factor of the wedge product.) If $\eta = \sigma \wedge dx_I$, where σ is a smooth 1-form and dx_I is of the form

$$dx_I = dx_{i_1} \wedge \cdots \wedge dx_{i_{p-1}}$$

then

$$d\eta = d\sigma \wedge dx_I, \qquad I(\eta) = I(\sigma) \wedge dx_I,$$

so that

$$\begin{aligned} dI(\eta) + I(d\eta) &= (dI(\sigma) + I(d\sigma)) \wedge dx_I = (\beta_1^* \sigma - \beta_0^* \sigma) \wedge dx_I \\ &= \beta_1^* \eta - \beta_0^* \eta, \end{aligned}$$

as required. We have therefore verified (4). This completes the proof of the Poincaré Lemma.

Remark One can generalize the Poincaré Lemma to apply to smooth forms on 'smoothly contractible' domains. An open set D in $\mathbb{R}^n$ is said to be *smoothly contractible* if and only if there exists a point a of D and a smooth map $H\colon [0,1] \times D \to D$ mapping $[0,1] \times D$ from $[0,1] \times D$ to D with the property that $H(0, \mathbf{x}) = \mathbf{a}$ and $H(1, \mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in D$. Note that star-shaped domains are smoothly contractible. The proof of the Poincaré Lemma applies can obviously be generalized to contractible open sets.

Remark Consider the open set D in $\mathbb{R}^2$ defined by $D = \mathbb{R}^2 \setminus \{(0,0)\}$. Let (x, y) denote the standard Cartesian coordinates on D and let ω be the 1-form on D defined by

$$\omega = \frac{x}{x^2 + y^2}\, dy - \frac{y}{x^2 + y^2}\, dx.$$

A simple calculation shows that $d\omega = 0$ everywhere on D. However it can be shown that there does not exist any real-valued function f *defined over the whole of* D with the property that $df = \omega$. Note that the open set D is not star shaped.

One can deduce certain useful results concerning vector fields in $\mathbb{R}^3$ from the Poincaré Lemma. Let D be an open set in $\mathbb{R}^3$ which is star-shaped with respect to some point of D, and let $\mathbf{V}\colon D \to \mathbb{R}^3$ be a smooth vector field on D. Let V_1, V_2 and V_3 be the three components of V. If we apply the Poincaré Lemma to the 1-form

$$V_1\, dx + V_2\, dy + V_3\, dz$$

we see that if $\operatorname{curl} \mathbf{V} = 0$ then there exists some function f on D such that $\mathbf{V} = \operatorname{grad} f$. Such a function f is referred to as a *scalar potential* for the vector field $\mathbf{V}$.

Similarly if we apply the Poincaré Lemma to the 2-form

$$V_1\, dy \wedge dz + V_2\, dz \wedge dx + V_3\, dx \wedge dy$$

we see that if $\operatorname{div} \mathbf{V} = 0$ then there exists some vector field $\mathbf{A}$ on D such that $\mathbf{V} = \operatorname{curl} \mathbf{A}$. Such a vector field $\mathbf{A}$ is referred to as a *vector potential* for the vector field $\mathbf{V}$.

11 The Riemann Integral in n Dimensions

Definition We define a (closed) *n-rectangle* in $\mathbb{R}^n$ to be a set of the form

$$\{\mathbf{x} \in \mathbb{R}^n : a_i \le x_i \le b_i \text{ for } i = 1, 2, \ldots, n\},$$

where $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ are real numbers such that $a_i \leq b_i$ for all i. The *interior* of the above n-rectangle is defined to be the open set

$$\{\mathbf{x} \in \mathbb{R}^n : a_i < x_i < b_i \text{ for } i = 1, 2, \ldots, n\}.$$

The *volume* (or *n-volume*) of the above n-rectangle is defined to be the product

$$(b_1 - a_1)(b_2 - a_2) \ldots (b_n - a_n)$$

of the quantities $b_i - a_i$ for $i = 1, 2, \ldots, n$. We denote the volume of an n-rectangle S by $\mathrm{vol}(S)$.

We define the the notion of Riemann-integrability for a bounded function defined on a (closed) n-rectangle in $\mathbb{R}^n$. First we must discuss partitions of n-rectangles.

Definition Let C be the n-rectangle in $\mathbb{R}^n$ given by

$$C = \{\mathbf{x} \in \mathbb{R}^n : a_i \leq x_i \leq b_i \quad \text{for} \quad i = 1, 2, \ldots, n\},$$

where $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$ are real numbers satisfying $a_i \leq b_i$ for $i = 1, 2, \ldots, n$. Suppose that, for each i, we have a partition P_i of the closed interval $[a_i, b_i]$ given by $P_i = \{t_{i,0}, t_{i,1}, \ldots, t_{i,s(i)}\}$, where

$$a_i = t_{i,0} < t_{i,1} < \ldots < t_{i,s(i)} = b_i.$$

Then these partitions of the closed intervals $[a_1, b_1], [a_2, b_2], \ldots, [a_n, b_n]$ determine a partition of the n-rectangle C into subrectangles, these subrectangles being of the form

$$\{\mathbf{x} \in \mathbb{R}^n : t_{i,r(i)-1} \leq x_i \leq t_{i,r(i)}\},$$

for some choice of $r(1), r(2), \ldots, r(n)$ (where each $r(i)$ is an integer between 1 and $s(i)$). If P denotes the partition of the n-rectangle C into subrectangles determined by the partitions P_i of $[a_i, b_i]$ for $i = 1, 2, \ldots, n$, then we write $P = P_1 \times P_2 \times \cdots \times P_n$.

Given a partition P of the n-rectangle C, let us denote by $\mathcal{R}(P)$ the collection consisting of all the subrectangles of C arising from the partition P of C. Thus if

$$C = \{\mathbf{x} \in \mathbb{R}^n : a_i \leq x_i \leq b_i \quad \text{for} \quad i = 1, 2, \ldots, n\},$$

and if $P = P_1 \times P_2 \times \cdots \times P_n$, where $P_i = \{t_{i,0}, t_{i,1}, \ldots, t_{i,s(i)}\}$, then $\mathcal{R}(P)$ consists of all n-rectangles of the form

$$\{\mathbf{x} \in \mathbb{R}^n : t_{i,r(i)-1} \leq x_i \leq t_{i,r(i)}\},$$

for all choices of $r(1), r(2), \ldots, r(n)$ (where each $r(i)$ is an integer between 1 and $s(i)$). Given a bounded real-valued function f defined on an n-rectangle C and given a partition P of C, we now define the lower sum $L(P, f)$ and the upper sum $U(P, f)$ of f with respect to this partition. These quantities are the analogues (in n-dimensions) of the corresponding quantities in the one-dimensional case.

Definition Let $f: C \to \mathbb{R}$ be a real-valued function on an n-rectangle C which is bounded above and below on C. We define

$$L(P, f) = \sum_{S \in \mathcal{R}(P)} m_S(f) \operatorname{vol}(S), \qquad U(P, f) = \sum_{S \in \mathcal{R}(P)} M_S(f) \operatorname{vol}(S),$$

where

$$m_S(f) = \inf\{f(\mathbf{x}) : \mathbf{x} \in S\}, \qquad M_S(f) = \sup\{f(\mathbf{x}) : \mathbf{x} \in S\}.$$

Remark Suppose that the partition P of the n-rectangle C subdivides C into the subrectangles $S_1, S_2, \ldots, S_p$ (i.e., $\mathcal{R}(P) = \{S_1, S_2, \ldots, S_p\}$). Then

$$\begin{aligned} \operatorname{vol}(C) &= \sum_{j=1}^{p} \operatorname{vol}(S_j), \\ L(P, f) &= \sum_{j=1}^{p} m_{S_j}(f) \operatorname{vol}(S_j), \\ U(P, f) &= \sum_{j=1}^{p} M_{S_j}(f) \operatorname{vol}(S_j), \end{aligned}$$

where $m_{S_j}(f)$ is the infimum of the values of the function f on S_j and $M_{S_j}(f)$ is the supremum of the values of f on S_j.

Observe that if $\alpha \leq f(\mathbf{x}) \leq \beta$ for all $\mathbf{x} \in C$, where α and β are suitable constants, then

$$\alpha \operatorname{vol}(C) \leq L(P, f) \leq U(P, f) \leq \beta \operatorname{vol}(C).$$

Definition Let C be an n-rectangle in $\mathbb{R}^n$ and let P and R be partitions of C. The partition R is said to be a refinement of the partition P if and only if each of the subrectangle of C arising from the partition R of C is contained within some subrectangle arising from the partition P of C.

Suppose that the n-rectangle C is given by

$$C = \{\mathbf{x} \in \mathbb{R}^n : a_i \leq x_i \leq b_i \text{ for } i = 1, 2, \ldots, n\},$$

and that the partitions P and R of C are given by

$$P = P_1 \times P_2 \times \cdots \times P_n, \qquad R = R_1 \times R_2 \times \cdots \times R_n,$$

where P_i and R_i are partitions of the closed interval $[a_i, b_i]$ for each integer i between 1 and n. One can easily verify that the partition R of C is a refinement of the partition P if and only if R_i is a refinement of P_i for each i.

Let P be a partition of the n-rectangle C and let R be a refinement of P. Let S be a subrectangle of C which arises from the partition P of C. Then those subrectangles of C arising from the partition R which are contained wholly within S constitute a partition of S.

Let P and Q be partitions of the n-rectangle C. Then there exists a partition R of C which is a refinement of both P and Q. Indeed if

$$P = P_1 \times P_2 \times \cdots \times P_n, \qquad Q = Q_1 \times Q_2 \times \cdots \times Q_n$$

then we choose a partition R_i of $[a_i, b_i]$ for each i such that R_i is a common refinement of P_i and Q_i. We say that the partition R of C is a common refinement of the partitions P and Q.

Lemma 11.1 *Let C be an n-rectangle in $\mathbb{R}^n$. Let $f: C \to \mathbb{R}$ be a bounded real-valued function defined on C. Let P be a partition of C and let R be a refinement of P. Then*

$$L(R, f) \geq L(P, f), \qquad U(R, f) \leq U(P, f).$$

Proof Let $S_1, S_2, \ldots, S_p$ be an enumeration of the subrectangles of C which arise from the partition P. Let S_j be any one of these subrectangles. Let $T_{j,1}, T_{j,2}, \ldots, T_{j,r_j}$ be an enumeration of those subrectangles of C arising from the partition R which are contained wholly within S_j. Now those subrectangles arising from the partition R which are contained wholly within the n-rectangle S_j constitute a partition of S_j, since the partition R of C is a refinement of the partition P. Thus, for each integer j between 1 and p, $S_j = T_{j,1} \cup T_{j,2} \cup \ldots \cup T_{j,r_j}$ and

$$\operatorname{vol}(S_j) = \sum_{k=1}^{r_j} \operatorname{vol}(T_{j,k}).$$

Now

$$L(P,f) = \sum_{j=1}^{p} m_{S_j}(f)\,\mathrm{vol}(S_j), \quad U(P,f) = \sum_{j=1}^{p} M_{S_j}(f)\,\mathrm{vol}(S_j),$$

and

$$L(R,f) = \sum_{j=1}^{p}\sum_{k=1}^{r_j} m_{T_{j,k}}(f)\,\mathrm{vol}(T_{j,k}), \quad U(R,f) = \sum_{j=1}^{p}\sum_{k=1}^{r_j} M_{T_{j,k}}(f)\,\mathrm{vol}(T_{j,k}),$$

where

$$m_{S_j}(f) = \inf\{f(\mathbf{x}) : \mathbf{x} \in S_j\}, \qquad M_{S_j}(f) = \sup\{f(\mathbf{x}) : \mathbf{x} \in S_j\},$$

$$m_{T_{j,k}}(f) = \inf\{f(\mathbf{x}) : \mathbf{x} \in T_{j,k}\}, \qquad M_{T_{j,k}}(f) = \sup\{f(\mathbf{x}) : \mathbf{x} \in T_{j,k}\}.$$

But $m_{T_{j,k}}(f) \geq m_{S_j}(f)$ and $M_{T_{j,k}}(f) \leq M_{S_j}(f)$, hence

$$\begin{aligned} L(R,f) &\geq \sum_{j=1}^{p} m_{S_j}(f) \sum_{k=1}^{r_j} \mathrm{vol}(T_{j,k}) = \sum_{j=1}^{p} m_{S_j}(f)\,\mathrm{vol}(S_j) = L(P,f), \\ U(R,f) &\leq \sum_{j=1}^{p} M_{S_j}(f) \sum_{k=1}^{r_j} \mathrm{vol}(T_{j,k}) = \sum_{j=1}^{p} M_{S_j}(f)\,\mathrm{vol}(S_j) = L(P,f), \end{aligned}$$

as required. ∎

Definition Let f be a bounded real-valued function defined over an n-rectangle C in $\mathbb{R}^n$. We define the *lower Riemann integral*

$$\mathcal{L}\int_C f(\mathbf{x})\,dx_1\,dx_2\ldots dx_n$$

of f on C to be the supremum (or least upper bound) of the quantities $L(P,f)$ as P ranges over all possible partitions of the n-rectangle C. Similarly we define the *upper Riemann integral*

$$\mathcal{U}\int_C f(\mathbf{x})\,dx_1\,dx_2\ldots dx_n$$

of f on C to be the infimum (or greatest lower bound) of the quantities $U(Q,f)$ as Q ranges over all possible partitions of the n-rectangle C.

Now if P and Q are partitions of the n-rectangle C then $L(P, f) \leq L(Q, f)$ for all bounded real-valued functions f on C. Indeed if R is a common refinement of the partitions P and Q then

$$L(P, f) \leq L(R, f) \leq U(R, f) \leq U(Q, f)$$

by Lemma 11.1. Therefore

$$\mathcal{L} \int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n \leq \mathcal{U} \int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n$$

for all bounded real-valued functions f on C.

Definition Let $f : C \to \mathbb{R}$ be a real-valued function defined on an n-rectangle C in $\mathbb{R}^n$. Suppose that f is bounded above and below on C. The function f is said to be *Riemann integrable* on C if and only if

$$\mathcal{L} \int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n = \mathcal{U} \int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n.$$

If f is Riemann-integrable on C then we define the Riemann integral

$$\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n$$

of f on C to be the common value of the lower and upper Riemann integrals of f on C.

The following result is the analogue for multiple integrals of Theorem 4.3.

Theorem 11.2 *Let $f: C \to \mathbb{R}$ be a bounded real-valued function defined on an n-rectangle C. Then f is Riemann-integrable on C if and only if, for every $\varepsilon > 0$ there exists a partition P of C for which*

$$U(P, f) - L(P, f) < \varepsilon.$$

Proof Suppose that f is Riemann-integrable on C. Let $\varepsilon > 0$ be any positive real number. We must show that there exists a partition P of C such that $U(P, f) - L(P, f) < \varepsilon$. Now if f is Riemann-integrable then there exists a partition Q of C such that

$$\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n - L(Q, f) < \tfrac{1}{2}\varepsilon.$$

Similarly there exists a partition R of C such that

$$U(R, f) - \int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n < \tfrac{1}{2}\varepsilon.$$

Then $U(R, f) - L(Q, f) < \varepsilon$. Let the partition P of C be a common refinement of the partitions Q and R. Using Lemma 11.1 we see that

$$L(Q, f) \leq L(P, f) \leq U(P, f) \leq U(R, f),$$

and hence $U(P, f) - L(P, f) < \varepsilon$. This shows that if f is Riemann-integrable then, given any $\varepsilon > 0$, there exists a partition P of C such that $U(P, f) - L(P, f) < \varepsilon$.

Conversely, let $f\colon C \to \mathbb{R}$ be a bounded real-valued function on C with the property that, given any $\varepsilon > 0$, there exists a partition P of C such that $U(P, f) - L(P, f) < \varepsilon$. We must show that f is Riemann-integrable on C. Now

$$L(P, f) \leq \mathcal{L}\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n \leq \mathcal{U}\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n \leq U(P, f)$$

for all partitions P of C. Therefore we conclude that

$$\mathcal{U}\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n - \mathcal{L}\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n < \varepsilon$$

for all $\varepsilon > 0$. But

$$\mathcal{U}\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n \geq \mathcal{L}\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n.$$

We conclude therefore that

$$\mathcal{U}\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n = \mathcal{L}\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n.$$

Thus f is Riemann-integrable on C. ∎

The following theorem is the analogue for multiple integrals of Theorem 4.5.

Theorem 11.3 *Let C be an n-rectangle in $\mathbb{R}^n$ and let f and g be be bounded Riemann-integrable functions on C. Let α be a real number. Then the functions $f + g$ and αf are Riemann-integrable on C, and*

$$\begin{aligned} \int_C (f(\mathbf{x}) + g(\mathbf{x}))\, dx_1\, dx_2 \ldots dx_n &= \int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n \\ &\quad + \int_C g(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n, \\ \int_C \alpha f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n &= \alpha \int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n. \end{aligned}$$

Proof Note that

$$m_S(f) + m_S(g) \le m_S(f+g) \le M_S(f+g) \le M_S(f) + M_S(g)$$

for all n-rectangles S in $\mathbb{R}^n$, where $m_S(f)$, $m_S(g)$ and $m_S(f+g)$ are the infima of the functions f, g and $f+g$ on S, and where $M_S(f)$, $M_S(g)$ and $M_S(f+g)$ are the suprema of the functions f, g and $f+g$ on S. It follows from this that

$$L(P,f) + L(P,g) \le L(P,f+g) \le U(P,f+g) \le U(P,f) + U(P,g)$$

for every partition P of the n-rectangle C.

Given $\varepsilon > 0$ choose partitions Q and R of C such that

$$\begin{aligned} L(Q,f) &> \int_C f(\mathbf{x})\,dx_1\,dx_2\ldots dx_n - \frac{1}{4}\varepsilon, \\ L(R,g) &> \int_C g(\mathbf{x})\,dx_1\,dx_2\ldots dx_n - \frac{1}{4}\varepsilon, \\ U(Q,f) &< \int_C f(\mathbf{x})\,dx_1\,dx_2\ldots dx_n + \frac{1}{4}\varepsilon, \\ U(R,g) &< \int_C g(\mathbf{x})\,dx_1\,dx_2\ldots dx_n + \frac{1}{4}\varepsilon. \end{aligned}$$

Let P be a common refinement of Q and R. Then

$$\begin{aligned} L(P,f+g) &> \int_C f(\mathbf{x})\,dx_1\,dx_2\ldots dx_n + \int_C g(\mathbf{x})\,dx_1\,dx_2\ldots dx_n - \frac{1}{2}\varepsilon, \\ U(P,f+g) &< \int_C f(\mathbf{x})\,dx_1\,dx_2\ldots dx_n + \int_C g(\mathbf{x})\,dx_1\,dx_2\ldots dx_n + \frac{1}{2}\varepsilon. \end{aligned}$$

Hence $U(P,f+g) - L(P,f+g) < \varepsilon$. This shows that $f+g$ is Riemann-integrable and that

$$\begin{aligned} \int_C (f(\mathbf{x}) + g(\mathbf{x}))\,dx_1\,dx_2\ldots dx_n &= \int_C f(\mathbf{x})\,dx_1\,dx_2\ldots dx_n \\ &\quad + \int_C g(\mathbf{x})\,dx_1\,dx_2\ldots dx_n. \end{aligned}$$

It is easily verified that αf is Riemann-integrable and that

$$\int_C \alpha f(\mathbf{x})\,dx_1\,dx_2\ldots dx_n = \alpha \int_C f(\mathbf{x})\,dx_1\,dx_2\ldots dx_n. \quad \blacksquare$$

Let S be an n-rectangle in $\mathbb{R}^n$, given by

$$S = \{\mathbf{x} \in \mathbb{R}^n : u_i \leq x_i \leq v_i\},$$

where $u_1, u_2, \ldots, u_n$ and $v_1, v_2, \ldots, v_n$ are real numbers which satisfy $u_i \leq v_i$ for each i. Given any continuous real-valued function f on S, let us denote by $I_S(f)$ the repeated integral of f over the n-rectangle S which is given by

$$I_S(f) = \int_{x_n=u_n}^{v_n} \left(\cdots \int_{x_2=u_2}^{v_2} \left(\int_{x_1=u_1}^{v_1} f(x_1, x_2, \ldots, x_n)\, dx_1 \right) dx_2 \ldots \right) dx_n$$

(i.e., $I_S(f)$ is obtained by integrating the function f first over the coordinate x_1, then over the coordinate x_2, and so on). Observe that if $\alpha \leq f(\mathbf{x}) \leq \beta$ on S for some constants α and β then

$$\alpha \operatorname{vol}(S) \leq I_S(f) \leq \beta \operatorname{vol}(S).$$

We shall use this fact to show that if f is a continuous function on some n-rectangle C in $\mathbb{R}^n$ then

$$I_C(f) = \int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n$$

(i.e., $I_C(f)$ is equal to the Riemann integral of f over C).

Theorem 11.4 *Let f be a continuous real-valued function defined on some n-rectangle C in $\mathbb{R}^n$, where*

$$C = \{\mathbf{x} \in \mathbb{R}^n : a_i \leq x_i \leq b_i\}.$$

Then f is Riemann-integrable on C, and moreover the Riemann integral

$$\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n$$

of f over C is equal to the repeated integral

$$\int_{x_n=a_n}^{b_n} \left(\cdots \int_{x_2=a_2}^{b_2} \left(\int_{x_1=a_1}^{b_1} f(x_1, x_2, \ldots, x_n)\, dx_1 \right) dx_2 \ldots \right) dx_n.$$

Proof Given a partition P of the n-rectangle C, we denote by $L(P, f)$ and $U(P, f)$ the quantities defined by

$$L(P, f) = \sum_{S \in \mathcal{R}(P)} m_S(f) \operatorname{vol}(S), \qquad U(P, f) = \sum_{S \in \mathcal{R}(P)} M_S(f) \operatorname{vol}(S)$$

where $\mathcal{R}(P)$ denotes the collection consisting of the subrectangles of C arising from the partition P of C, and where

$$m_S(f) = \inf\{f(\mathbf{x}) : \mathbf{x} \in S\}, \qquad M_S(f) = \sup\{f(\mathbf{x}) : \mathbf{x} \in S\}.$$

First we show that f is Riemann-integrable on C. We do this by showing that, for any $\varepsilon > 0$, there exists a partition P of C for which $U(P, f) - L(P, f) < \varepsilon$.

Now the n-rectangle C is a closed bounded set. Thus the continuous function f is bounded on C (by Theorem 3.1) and is uniformly continuous on C (by Theorem 3.3). Let $\varepsilon > 0$ be given. Then there exists some $\delta > 0$ such that

$$|f(\mathbf{x}) - f(\mathbf{y})| < \frac{\varepsilon}{2\operatorname{vol}(C)}$$

whenever $\mathbf{x}, \mathbf{y} \in C$ satisfy $|\mathbf{x} - \mathbf{y}| < \delta$. Choose a partition P of C which has the property that if S is any one of the subrectangles arising from the partition P of C then $|\mathbf{x} - \mathbf{y}| < \delta$ for all points $\mathbf{x}$ and $\mathbf{y}$ of S. (This can be done by ensuring that the length of the sides of the subrectangles arising from the partition P do not exceed $\delta/\sqrt{n}$.) Note that

$$M_S(f) - m_S(f) \le \frac{\varepsilon}{2\operatorname{vol}(C)}.$$

for every $S \in \mathcal{R}(P)$. But

$$U(P, f) - L(P, f) = \sum_{S \in \mathcal{R}(P)} (M_S(f) - m_S(f)) \operatorname{vol}(S),$$

hence

$$U(P, f) - L(P, f) \le \frac{\varepsilon}{2\operatorname{vol}(C)} \sum_{S \in \mathcal{R}(P)} \operatorname{vol}(S) = \tfrac{1}{2}\varepsilon.$$

Thus, given any $\varepsilon > 0$, there exists a partition P of the n-rectangle C such that $U(P, f) - L(P, f) < \varepsilon$. It follows from Theorem 11.1 that f is Riemann-integrable on C.

We now show that the Riemann integral of f over the n-rectangle C is equal to the repeated integral given above. Let P be a partition of C. Let S be any of the subrectangles of C arising from this partition. Then

$$m_S(f) \operatorname{vol}(S) \le I_S(f) \le M_S(f) \operatorname{vol}(S),$$

where

$$m_S(f) = \inf\{f(\mathbf{x}) : \mathbf{x} \in S\}, \qquad M_S(f) = \sup\{f(\mathbf{x}) : \mathbf{x} \in S\}.$$

Now

$$I_C(f) = \sum_{S \in \mathcal{R}(P)} I_S(f),$$

where $\mathcal{R}(P)$ is the collection consisting of all the subrectangles arising from the partition P of C. Thus if $L(P, f)$ and $U(P, f)$ are defined by

$$L(P, f) = \sum_{S \in \mathcal{R}(P)} m_S(f) \operatorname{vol}(S), \qquad U(P, f) = \sum_{S \in \mathcal{R}(P)} M_S(f) \operatorname{vol}(S)$$

then

$$L(P, f) \leq I_C(f) \leq U(P, f).$$

The Riemann integral of f is equal to the supremum of the quantities $L(P, f)$ as P ranges over all partitions of the n-rectangle C, hence

$$\int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n \leq I_C(f).$$

Similarly the Riemann integral of f is equal to the infimum of the quantities $U(P, f)$ as P ranges over all partitions of the n-rectangle C, hence

$$I_C(f) \leq \int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n.$$

Hence

$$I_C(f) = \int_C f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n,$$

as required. ∎

Note that the order in which the integrations are performed in the repeated integral plays no role in the above proof. We may therefore deduce the following important corollary.

Corollary 11.5 *Let f be a continuous real-valued function defined over some closed rectangle C in $\mathbb{R}^2$, where*

$$C = \{(x, y) \in \mathbb{R}^2 : a \leq x \leq b, \quad c \leq y \leq d\}.$$

Then

$$\int_a^b \left(\int_c^d f(x, y)\, dy \right) dx = \int_c^d \left(\int_a^b f(x, y)\, dx \right) dy.$$

Proof It follows directly from Theorem 11.4 that the repeated integrals

$$\int_a^b \left(\int_c^d f(x,y)\, dy \right) dx \text{ and } \int_c^d \left(\int_a^b f(x,y)\, dx \right) dy$$

are both equal to the Riemann integral of the function f over the rectangle C. Therefore these repeated integrals must be equal. ∎

The material in the remainder of this section of the course is ***NON-EXAMINABLE****. It is included for the sake of completeness. However proofs of the results described below are not presented here.*

In this section we have discussed integrals defined over n-rectangles in n-dimensional Euclidean space $\mathbb{R}^n$. However it is often necessary to integrate functions over regions (such as balls, tetrahedra, ellipsoids etc.) which are not n-rectangles. However if one is to make sense of the concept of the Riemann integral of a bounded function over such a set, it is neccessary to impose certain conditions on the set. We consider sets V which are closed and bounded. However the boundary of the set V must be 'sufficiently regular' in order to allow us to define the Riemann integral of a real-valued function over V. We give some definitions which will be applied when we discuss the nature of the conditions to be imposed on the set V.

Definition Let D be a subset of $\mathbb{R}^n$. The *closure* $\overline{D}$ of D is defined to be the smallest closed set containing D. Thus $\overline{D}$ is closed, and $\overline{D} \subset A$ for every closed set A containing D. Similarly the *interior* $\operatorname{int}(D)$ of D is defined to be the largest open set contained in D. Thus $\operatorname{int}(D)$ is an open set, and $U \subset \operatorname{int}(D)$ for every open set U contained in D. The *frontier* (or *boundary*) of D is defined to be the set $\overline{D} \setminus \operatorname{int}(D)$ (i.e., the frontier of D consists of all points of the closure $\overline{D}$ of D which do not belong to the interior $\operatorname{int}(D)$ of D.

Let F be the frontier of the set D. A point $\mathbf{x}$ of $\mathbb{R}^n$ is contained in the complement $\mathbb{R}^n \setminus F$ of F if and only if either there exists an open neighbourhood of $\mathbf{x}$ which is a subset of D or else there exists an open neighbourhood of $\mathbf{x}$ which is a subset of the complement $\mathbb{R}^n \setminus D$ of D.

Definition Let A be a subset of $\mathbb{R}^n$. Then the set A is said to have (n-dimensional) *content* 0 if and only if, for every $\varepsilon > 0$, there exists a finite collection $S_1, S_2, \ldots, S_k$ of n-rectangles in $\mathbb{R}^n$ such that

(i) each point of the set A is contained in the interior of one of these n-rectangles,

(ii) $\sum_{j=1}^{k} \operatorname{vol}(S_j) < \varepsilon$.

Definition A closed bounded subset V of $\mathbb{R}^n$ is said to be *Jordan-measurable* if and only if the frontier (or boundary) of the set V has content 0.

Let V be a subset of $\mathbb{R}^n$. The *characteristic function* χ_V of the set V is defined by

$$\chi_V(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in V; \\ 0 & \text{if } \mathbf{x} \notin V. \end{cases}$$

Theorem 11.6 *Let V be a closed bounded subset of $\mathbb{R}^n$ and let C be an n-rectangle in $\mathbb{R}^n$ which contains the set V. Then the characteristic function χ_V of V is Riemann-integrable over C if and only if V is Jordan-measurable.*

A proof of this result is to be found in *Calculus on manifolds* by M. Spivak (Theorem 3-9 on page 55).

In view of Theorem 11.6 we shall restrict our attention to defining the integral of real-valued functions over Jordan-measurable closed bounded sets. It is an 'obvious fact' that sets such as spheres, wedges etc. that one meets frequently in applied mathematics when evaluating multiple integrals are all Jordan-measurable.

Definition Let V be a closed bounded Jordan-measurable set in $\mathbb{R}^n$. Let $f: V \to \mathbb{R}$ be a bounded real-valued function defined on V and let $\tilde{f}: \mathbb{R}^n \to \mathbb{R}$ be the extension of f to the whole of $\mathbb{R}^n$ defined by

$$\tilde{f}(\mathbf{x}) = \begin{cases} f(\mathbf{x}) & \text{if } \mathbf{x} \in V; \\ 0 & \text{if } \mathbf{x} \notin V. \end{cases}$$

The function f is said to be *Riemann-integrable* over the set V if and only if $\tilde{f}$ is Riemann integrable over any n-rectangle which contains the set V. If f is Riemann-integrable over V then we define the Riemann integral

$$\int_V f(\mathbf{x})\, dx_1\, dx_2 \ldots dx_n$$

of f over V to be equal to the Riemann integral of the function $\tilde{f}$ over any n-rectangle which contains V.

It can be shown that if V is a closed bounded Jordan-measurable set and if $f: V \to \mathbb{R}$ is a continuous real-valued function on V then f is Riemann-integrable on V.

Finally we state a version of the change-of variables formula for multiple integrals. Let D and E be subsets of $\mathbb{R}^n$.

Theorem 11.7 (Change of Variables Formula) *Let D be an open subset of $\mathbb{R}^n$ and let $\varphi: D \to \mathbb{R}^n$ be a smooth map. Let V be a closed bounded Jordan-measurable subset of D and let f be a real-valued function defined over $\varphi(D)$ which is Riemann-integrable on $\varphi(D)$. Then*

$$\int_{\varphi(V)} f(\mathbf{y})\, dy_1\, dy_2 \ldots dy_n = \int_V f(\varphi(\mathbf{x}))|\det \varphi'(\mathbf{x})|\, dx_1\, dx_2 \ldots dx_n.$$

Here $\det \varphi'(\mathbf{x})$ denotes the determinant of the Jacobian matrix $\varphi'(\mathbf{x})$ of the smooth map φ at the point $\mathbf{x}$ of V.

The proof of this result is somewhat involved. A rigorous proof requires a number of results (such as the Inverse Function Theorem) which go well beyond the scope of this course. A proof of the change of variables formula can be found in *Calculus on manifolds* by M. Spivak (pages 66–72).

12 Curvilinear Coordinate Systems

In this section we shall discuss curvilinear coordinate systems defined over open sets in $\mathbb{R}^n$. We shall then discuss the representation of differential forms with respect to curvilinear coordinate systems.

Let D and E be subsets of $\mathbb{R}^n$ and let $\varphi: D \to E$ be a map from D into E. The map $\varphi: D \to E$ is said to be a *diffeomorphism* from D to E if and only if

(i) the function φ is a bijection from D to E (so that the inverse map $\varphi^{-1}: E \to D$ is well-defined),

(ii) $\varphi: D \to E$ and $\varphi^{-1}: E \to D$ are both smooth.

(Recall that a map $\varphi: D \to E$ is said to be *smooth* if and only if the partial derivatives

$$\frac{\partial^{k_1+k_2+\cdots+k_n}\varphi_j}{\partial x_1^{k_1}\, \partial x_2^{k_2} \ldots \partial x_n^{k_n}}$$

of the components $\varphi_1, \varphi_2, \ldots, \varphi_m$ of the map φ exist and are continuous for all choices of the non-negative integers $k_1, k_2, \ldots, k_n$.)

Lemma 12.1 *Let D and E be open sets in $\mathbb{R}^n$, and let $\varphi: D \to E$ be a diffeomorphism from D to E. Then the determinant $\det \varphi'(\mathbf{x})$ of the Jacobian matrix $\varphi'(\mathbf{x})$ of φ is non-zero at each point $\mathbf{x}$ of D.*

Proof Let $\psi: E \to D$ be the inverse φ^{-1} of the map φ. Then $\psi \circ \varphi$ is the identity mapping of D, so that the Jacobian matrix $(\psi \circ \varphi)'(\mathbf{x})$ of $\psi \circ \varphi$ at any point $\mathbf{x}$ of D is the identity matrix. But we see from the Chain Rule (Theorem 7.10) that $(\psi \circ \varphi)'(\mathbf{x}) = \psi'(\varphi(\mathbf{x}))\varphi'(\mathbf{x})$ for all $\mathbf{x} \in D$. Therefore

$$\det \psi'(\varphi(\mathbf{x})) \, \det \varphi'(\mathbf{x}) = \det \left(\psi'(\varphi(\mathbf{x}))\varphi'(\mathbf{x})\right) = 1$$

for all $\mathbf{x} \in D$, and hence $\det \varphi'(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in D$, as required. ∎

Definition Let D and E be open sets in $\mathbb{R}^n$ and let $\varphi: D \to E$ be a diffeomorphism from D to E. If $\det \varphi'(\mathbf{x}) > 0$ for all $\mathbf{x} \in D$ (where $\varphi'(\mathbf{x})$ denotes the Jacobian matrix of φ at $\mathbf{x}$) then the diffeomorphism φ is said to be *orientation-preserving.* If $\det \varphi'(\mathbf{x}) < 0$ for all $\mathbf{x} \in D$ then φ is said to be *orientation-reversing.*

We now discuss the notion of *path-connectedness.*

Definition A subset S of $\mathbb{R}^n$ is said to be *path-connected* if and only if, given any two points $\mathbf{u}$ and $\mathbf{v}$ of S there exists a continuous map $\gamma: [0, 1] \to S$ from the closed interval $[0, 1]$ into S such that $\gamma(0) = \mathbf{u}$ and $\gamma(1) = \mathbf{v}$ (i.e., the set S is path-connected if and only any two points of S can be joined by a continuous curve whose image is contained wholly within S).

Lemma 12.2 *Let S be a path-connected subset of $\mathbb{R}^n$ and let $f: S \to \mathbb{R}$ be a continuous real-valued function defined on S. Suppose that $f(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in S$. Then either $f(\mathbf{x}) > 0$ for all $\mathbf{x} \in S$ or else $f(\mathbf{x}) < 0$ for all $\mathbf{x} \in S$ (i.e., the function f has the same sign throughout S).*

Proof Let $\mathbf{u}$ be a point of S. Suppose that $f(\mathbf{u}) > 0$. Let $\mathbf{x}$ be a point of S. Then there exists a continuous map $\gamma: [0, 1] \to S$ such that $\gamma(0) = \mathbf{u}$ and $\gamma(1) = \mathbf{x}$. But then $f \circ \gamma$ is a continuous real-valued function defined on the closed interval $[0, 1]$ which is non-zero everywhere in $[0, 1]$. We deduce from this that $f(\mathbf{x}) > 0$ (since if it were the case that $f(\mathbf{x}) < 0$ then it would follow from the Intermediate Value Theorem that $f(\gamma(t)) = 0$ for some $t \in [0, 1]$, contradicting the fact that the function f is non-zero everywhere on S). Similarly if $f(\mathbf{u}) < 0$ then $f(\mathbf{x}) < 0$ for all $\mathbf{x} \in S$. ∎

We conclude immediately from Lemma 12.1 and Lemma 12.2 that if D and E are open sets in $\mathbb{R}^n$, if $\varphi: D \to E$ is a diffeomorphism, and if D is path-connected then either φ is orientation-preserving or else φ is orientation-reversing.

Definition Let U be an open set in $\mathbb{R}^n$ and let $\varphi_1, \varphi_2, \ldots, \varphi_n$ be smooth functions on U. We say that these functions define a *smooth curvilinear coordinate system* on U if and only if the map sending $\mathbf{x} \in U$ to

$$(\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots, \varphi_n(\mathbf{x}))$$

defines a diffeomorphism from U onto some open set in $\mathbb{R}^n$.

Example Let H be the half plane in $\mathbb{R}^3$ defined by

$$H = \{(x, y, z) : x \leq 0, \quad y = 0\}$$

and let U be the complement $\mathbb{R}^n \setminus H$ of H in $\mathbb{R}^n$. Then the spherical polar coordinates (r, θ, φ) form a smooth curvilinear coordinate system on U, where the values of θ and φ are chosen such that $0 < \theta < \pi$ and $-\pi < \varphi < \pi$. A point (x, y, z) of U is given in spherical polar coordinates by

$$x = r \sin\theta \, \cos\varphi, \qquad y = r \sin\theta \, \sin\varphi, \qquad z = r \cos\theta.$$

(We have excluded the half plane H from the domain U of our spherical polar coordinate system in order to ensure that the angles θ and φ are unambiguously defined and vary smoothly over the open set U.)

Similarly the cylindrical polar coordinates (ρ, φ, z) form a smooth curvilinear coordinate system on U, where $-\pi < \varphi < \pi$. A point (x, y, z) is given in cylindrical polar coordinates (ρ, φ, z) by

$$x = \rho \cos\varphi, \qquad y = \rho \sin\varphi.$$

Example Let u and v be the smooth real-valued functions on $\mathbb{R}^2$ defined by

$$u = x, \qquad v = y - x^2.$$

Then (u, v) is a smooth curvilinear coordinate system on $\mathbb{R}^2$. Note that x and y are given in terms of the curvilinear coordinates u and v by

$$x = y, \qquad y = v + u^2.$$

Let $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ be a smooth curvilinear coordinate system defined on an open subset U of $\mathbb{R}^n$. Let $\varphi\colon U \to \mathbb{R}^n$ be the smooth map defined by

$$\varphi(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots, \varphi_n(\mathbf{x})) .$$

Then φ maps U diffeomorphically onto some open subset $\varphi(U)$ of $\mathbb{R}^n$. It follows immediately from Lemma 12.1 that $\det \varphi'(\mathbf{x}) \neq 0$ for all $\mathbf{x} \in U$, where $\varphi'(\mathbf{x})$ is represented by the Jacobian matrix of φ at $\mathbf{x}$, given by

$$\varphi'(\mathbf{x}) = \begin{pmatrix} \dfrac{\partial \varphi_1}{\partial x_1} & \dfrac{\partial \varphi_1}{\partial x_2} & \cdots & \dfrac{\partial \varphi_1}{\partial x_n} \\ \dfrac{\partial \varphi_2}{\partial x_1} & \dfrac{\partial \varphi_2}{\partial x_2} & \cdots & \dfrac{\partial \varphi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial \varphi_n}{\partial x_1} & \dfrac{\partial \varphi_n}{\partial x_2} & \cdots & \dfrac{\partial \varphi_n}{\partial x_n} \end{pmatrix},$$

where the partial derivatives occurring in the matrix are evaluated at the point $\mathbf{x}$ of U.

Definition Let U be an open set in $\mathbb{R}^n$ and let $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ be a smooth curvilinear coordinate system defined on U. We say that this curvilinear coordinate system is *positively oriented* if and only if $\det \varphi'(\mathbf{x}) > 0$ for all $\mathbf{x} \in U$ (where $\varphi(\mathbf{x}) \equiv (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots, \varphi_n(\mathbf{x}))$ for all $\mathbf{x} \in U$). Similarly we say that this curvilinear coordinate system is *negatively oriented* if and only if $\det \varphi'(\mathbf{x}) < 0$ for all $\mathbf{x} \in U$.

Observe that the smooth curvilinear coordinate system $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ is positively oriented if and only if the map φ with components $\varphi_1, \varphi_2, \ldots, \varphi_n$ is an orientation-preserving diffeomorphism mapping U onto an open subset $\varphi(U)$ of $\mathbb{R}^n$. Similarly this curvilinear coordinate system is negatively oriented if and only if the map φ is an orientation-reversing diffeomorphism mapping U onto an open subset $\varphi(U)$ of $\mathbb{R}^n$.

If U is a path-connected open set in $\mathbb{R}^n$ then every curvilinear coordinate system defined over U is either positively-oriented or negatively-oriented, since any diffeomorphism from D to some open set in $\mathbb{R}^n$ is either orientation-preserving or else is orientation-reversing.

The orientation of a curvilinear coordinate system depends on the order in which the curvilinear coordinates are specified, as the following lemma shows.

Lemma 12.3 *Let $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ be a positively oriented smooth curvilinear coordinate system defined on an open set U in $\mathbb{R}^n$. Let π be a permutation of $\{1, 2, \ldots, n\}$ and let $j_1, j_2, \ldots, j_n$ be defined by $j_i = \pi(i)$ for each i. If π is an even permutation then the curvilinear coordinate system $(\varphi_{j_1}, \varphi_{j_2}, \ldots, \varphi_{j_n})$ is positively oriented. If π is an odd permutation then the curvilinear coordinate system $(\varphi_{j_1}, \varphi_{j_2}, \ldots, \varphi_{j_n})$ is negatively oriented.*

Proof The permutation π determines a corresponding permutation of the rows of the Jacobian matrix φ' of the transformation $\varphi\colon U \to \mathbb{R}^n$ represented by $\varphi_1, \varphi_2, \ldots, \varphi_n$. If the permutation is even then it leaves the determinant of the Jacobian matrix unchanged. If the permutation is odd then it multiplies the determinant of the Jacobian matrix by -1.The required result follows immediately. ∎

Example Let H be the half plane in $\mathbb{R}^3$ defined by

$$H = \{(x, y, z) : x \leq 0, \quad y = 0\}$$

and let U be the complement $\mathbb{R}^n \setminus H$ of H in $\mathbb{R}^n$. Then the spherical polar coordinates (r, θ, φ) form a positively oriented curvilinear coordinate system on U. To prove this we must show that the determinant of the Jacobian matrix J

$$\begin{pmatrix} \dfrac{\partial r}{\partial x} & \dfrac{\partial r}{\partial y} & \dfrac{\partial r}{\partial z} \\ \dfrac{\partial \theta}{\partial x} & \dfrac{\partial \theta}{\partial y} & \dfrac{\partial \theta}{\partial z} \\ \dfrac{\partial \varphi}{\partial x} & \dfrac{\partial \varphi}{\partial y} & \dfrac{\partial \varphi}{\partial z} \end{pmatrix},$$

is everywhere positive on U. We could do this by expressing the Cartesian coordinates (x, y, z) in terms of (r, θ, φ) and then calculating the components of the Jacobian matrix above directly. However it is more convenient to proceed by observing that the inverse J^{-1} of J is the Jacobian matrix of the transformation specifying the Cartesian coordinates (x, y, z) in terms of the spherical polar coordinates (r, θ, φ), so that

$$J^{-1} = \begin{pmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} & \dfrac{\partial x}{\partial \varphi} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} & \dfrac{\partial y}{\partial \varphi} \\ \dfrac{\partial z}{\partial r} & \dfrac{\partial z}{\partial \theta} & \dfrac{\partial z}{\partial \varphi} \end{pmatrix}.$$

Now

$$x = r \sin\theta \cos\varphi, \qquad y = r \sin\theta \sin\varphi, \qquad z = r \cos\theta.$$

A straightforward calculation shows that $\det J^{-1} = r^2 \sin\theta$. Hence

$$\det J = \frac{1}{r^2 \sin\theta},$$

where $0 < \theta < \pi$. We conclude that $\det J > 0$ everywhere on U. Thus the spherical polar coordinates (r, θ, φ) (when taken in this order) form a positively oriented curvilinear coordinate system on U.

12.1 Representation of Differential Forms in Curvilinear Coordinate Systems

Lemma 12.4 *Let $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ be a smooth curvilinear coordinate system defined over an open set D in $\mathbb{R}^n$. Define $E = \varphi(D)$, where $\varphi\colon D \to \mathbb{R}^n$ is defined by*

$$\varphi(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots, \varphi_n(\mathbf{x}))\,.$$

Then every smooth p-form on D can be expressed as a sum of terms of the form

$$F(\varphi_1, \varphi_2, \ldots, \varphi_n)\, d\varphi_{i_1} \wedge d\varphi_{i_2} \wedge \cdots \wedge d\varphi_{i_p},$$

where $F\colon E \to \mathbb{R}$ is a smooth function on E and where $i_1, i_2, \ldots, i_p$ are integers between 1 and n.

Proof It follows from the definition of the differential of a function that

$$d\varphi_i = \sum_{j=1}^{n} \frac{\partial \varphi_i}{\partial x_j}\, dx_j$$

for $i = 1, 2, \ldots, n$. But the Jacobian matrix

$$\begin{pmatrix} \dfrac{\partial \varphi_1}{\partial x_1} & \dfrac{\partial \varphi_1}{\partial x_2} & \cdots & \dfrac{\partial \varphi_1}{\partial x_n} \\ \dfrac{\partial \varphi_2}{\partial x_1} & \dfrac{\partial \varphi_2}{\partial x_2} & \cdots & \dfrac{\partial \varphi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial \varphi_n}{\partial x_1} & \dfrac{\partial \varphi_n}{\partial x_2} & \cdots & \dfrac{\partial \varphi_n}{\partial x_n} \end{pmatrix}$$

of φ is invertible at every point of D (since φ maps D diffeomorphically onto E). Let its inverse matrix be given by

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{pmatrix}$$

where $a_{ji}\colon D \to \mathbb{R}$ is a smooth function on D for each pair (j, i) of integers between 1 and n. Then

$$\sum_{i=1}^{n} a_{ji} \frac{\partial \varphi_i}{\partial x_k} = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{if } j \neq k, \end{cases}$$

hence

$$\sum_{i=1}^{n} a_{ji} d\varphi_i = \sum_{i=1}^{n} \sum_{k=1}^{n} a_{ji} \frac{\partial \varphi_i}{\partial x_k}\, dx_k = dx_j$$

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

Now each smooth p-form on D can be expressed as a sum of terms of the form

$$f(x_1, x_2, \ldots, x_n)\, dx_{j_1} \wedge dx_{j_2} \wedge \cdots \wedge dx_{j_p},$$

where f is a smooth real-valued function on D and where $j_1, j_2, \ldots, j_p$ are integers between 1 and n. However we have shown that

$$dx_j = \sum_{i=1}^{n} a_{ji} d\varphi_i$$

for each j. If we express each dx_j in terms of $d\varphi_1, d\varphi_2, \ldots, d\varphi_n$ in this fashion we conclude that each smooth p-from on D can be expressed as a sum of terms of the form

$$g(x_1, x_2, \ldots, x_n)\, d\varphi_{j_1} \wedge d\varphi_{j_2} \wedge \cdots \wedge d\varphi_{j_p},$$

where g is a smooth real-valued function on E. However

$$g(x_1, x_2, \ldots, x_n) = F(\varphi_1, \varphi_2, \ldots, \varphi_n),$$

where $F \colon E \to \mathbb{R}$ is defined by $F = g \circ \varphi^{-1}$. Thus every smooth p-form on D can be expressed as a sum of terms of the form

$$F(\varphi_1, \varphi_2, \ldots, \varphi_n)\, d\varphi_{i_1} \wedge d\varphi_{i_2} \wedge \cdots \wedge d\varphi_{i_p},$$

as required. ∎

Let $f \colon D \to \mathbb{R}$ be a smooth real-valued function defined over some open set D in $\mathbb{R}^n$ and let $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ be a smooth curvilinear coordinate system on D. Define $E = \varphi(D)$, where $\varphi \colon D \to \mathbb{R}^n$ is defined by

$$\varphi(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots, \varphi_n(\mathbf{x}))\,.$$

Suppose that

$$f = F(\varphi_1, \varphi_2, \ldots, \varphi_n),$$

where F is a smooth function on E. It follows from the Chain Rule that

$$\frac{\partial f}{\partial x_i} = \sum_{j=1}^{n} \frac{\partial F(\varphi_1, \varphi_2, \ldots, \varphi_n)}{\partial \varphi_j} \frac{\partial \varphi_j}{\partial x_i}.$$

But

$$df = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}\, dx_i, \qquad d\varphi_j = \sum_{i=1}^{n} \frac{\partial \varphi_j}{\partial x_i}\, dx_i.$$

It follows that

$$df = \sum_{j=1}^{n} \frac{\partial F(\varphi_1, \varphi_2, \ldots, \varphi_n)}{\partial \varphi_j}\, d\varphi_j.$$

Lemma 12.5 *Let $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ be a smooth curvilinear coordinate system defined over an open set D in $\mathbb{R}^n$. Let ω be the smooth p-form on D given by*

$$\omega = F(\varphi_1, \varphi_2, \ldots, \varphi_n)\, d\varphi_{i_1} \wedge d\varphi_{i_2} \wedge \cdots \wedge d\varphi_{i_p},$$

(where F is a smooth function of its arguments). Then

$$d\omega = \sum_{j=1}^{n} \frac{\partial F(\varphi_1, \varphi_2, \ldots, \varphi_n)}{\partial \varphi_j}\, d\varphi_j\, d\varphi_{i_1} \wedge d\varphi_{i_2} \wedge \cdots \wedge d\varphi_{i_p}.$$

Proof Recall that if η_1 and η_2 are differential forms on D then

$$d(\eta_1 \wedge \eta_2) = d\eta_1 \wedge \eta_2 + (-1)^t \eta_1 \wedge d\eta_2,$$

where t is the degree of the differential form η_1 (see Lemma 9.4). Now $d(d\varphi_i) = 0$ for all i (see Lemma 9.5). Using these facts one can show (e.g., by induction on p) that

$$d\left(d\varphi_{i_1} \wedge d\varphi_{i_2} \wedge \cdots \wedge d\varphi_{i_p}\right) = 0.$$

Thus

$$\begin{aligned} d\omega &= d\left(F(\varphi_1, \varphi_2, \ldots, \varphi_n)\right) \wedge d\varphi_{i_1} \wedge d\varphi_{i_2} \wedge \cdots \wedge d\varphi_{i_p} \\ &= \sum_{j=1}^{n} \frac{\partial F(\varphi_1, \varphi_2, \ldots, \varphi_n)}{\partial \varphi_j}\, d\varphi_j \wedge d\varphi_{i_1} \wedge d\varphi_{i_2} \wedge \cdots \wedge d\varphi_{i_p}, \end{aligned}$$

as required. ∎

Example Let H be the half plane in $\mathbb{R}^3$ defined by

$$H = \{(x, y, z) : x \leq 0, \quad y = 0\}$$

and let U be the complement $\mathbb{R}^n \setminus H$ of H in $\mathbb{R}^n$. Let (r, θ, φ) be spherical polar coordinates on U, so that

$$x = r \sin\theta\, \cos\varphi, \qquad y = r \sin\theta\, \sin\varphi, \qquad z = r \cos\theta,$$

where (x, y, z) denote the standard Cartesian coordinates of the point with polar coordinates (r, θ, φ). Using Lemma 12.5 we see that

$$\begin{aligned} dx &= \sin\theta\,\cos\varphi\,dr + r\cos\theta\,\cos\varphi\,d\theta - r\sin\theta\,\sin\varphi\,d\varphi, \\ dy &= \sin\theta\,\sin\varphi\,dr + r\cos\theta\,\sin\varphi\,d\theta + r\sin\theta\,\cos\varphi\,d\varphi, \\ dz &= \cos\theta\,dr - r\sin\theta\,d\theta. \end{aligned}$$

Suppose that we define 1-forms ω_1, ω_2 and ω_3 by

$$\omega_1 = dr, \quad \omega_2 = r\,d\theta, \quad \omega_3 = r\sin\theta\,d\varphi.$$

Then

$$\omega_1 \wedge \omega_2 = r\,dr \wedge d\theta, \quad \omega_2 \wedge \omega_3 = r^2 \sin\theta\,d\theta \wedge d\varphi, \quad \omega_3 \wedge \omega_1 = r\sin\theta\,d\varphi \wedge dr,$$

$$\omega_1 \wedge \omega_2 \wedge \omega_3 = r^2 \sin\theta\,dr \wedge d\theta \wedge d\varphi,$$

and

$$\begin{aligned} d\omega_1 &= 0, \\ d\omega_2 &= dr \wedge d\theta = \frac{1}{r}\,\omega_1 \wedge \omega_2, \\ d\omega_3 &= \sin\theta\,dr \wedge d\varphi + r\cos\theta\,d\theta \wedge d\varphi = \frac{\cot\theta}{r}\omega_2 \wedge \omega_3 - \frac{1}{r}\omega_3 \wedge \omega_1. \end{aligned}$$

Lemma 12.6 *Let D and E be open sets in $\mathbb{R}^n$ and let $\varphi\colon D \to E$ be a diffeomorphism from D to E. Let $\varphi_1, \varphi_2, \ldots, \varphi_n$ be the components of the map φ. Then*

$$d\varphi_1 \wedge d\varphi_2 \wedge \cdots \wedge d\varphi_n = (\det \varphi')\,dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n,$$

where $\varphi'(\mathbf{x})$ denotes the Jacobian matrix of the diffeomorphism φ at the point $\mathbf{x}$.

We do not present the proof of this result here in full generality. Instead we shall show that this formula is valid in the cases when $n = 1$, $n = 2$ and $n = 3$.

Proof of Lemma 12.6 in the case when $n = 1$. In this case D and E are open sets in the set $\mathbb{R}$ of real numbers, and

$$\det \varphi'(x) = \frac{d\varphi(x)}{dx}$$

for all $x \in D$. But it follows from the definition of the differential of a function that

$$d\varphi = \frac{d\varphi}{dx} dx.$$

Thus Lemma 12.6 is valid when $n = 1$.

Proof of Lemma 12.6 in the case when $n = 2$. The differentials $d\varphi_1$ and $d\varphi_2$ of φ_1 and φ_2 are given in this case by

$$d\varphi_1 = \frac{\partial \varphi_1}{\partial x_1}\, dx_1 + \frac{\partial \varphi_1}{\partial x_2}\, dx_2, \qquad d\varphi_2 = \frac{\partial \varphi_2}{\partial x_1}\, dx_1 + \frac{\partial \varphi_2}{\partial x_2}\, dx_2.$$

Now $dx_1 \wedge dx_1 = 0$, $dx_2 \wedge dx_2 = 0$ and $dx_2 \wedge dx_1 = -dx_1 \wedge dx_2$. Therefore

$$\begin{aligned} d\varphi_1 \wedge d\varphi_2 &= \left(\frac{\partial \varphi_1}{\partial x_1}\frac{\partial \varphi_2}{\partial x_2} - \frac{\partial \varphi_1}{\partial x_2}\frac{\partial \varphi_2}{\partial x_1} \right) dx_1 \wedge dx_2 \\ &= \begin{vmatrix} \dfrac{\partial \varphi_1}{\partial x_1} & \dfrac{\partial \varphi_1}{\partial x_2} \\ \dfrac{\partial \varphi_2}{\partial x_1} & \dfrac{\partial \varphi_2}{\partial x_2} \end{vmatrix} dx_1 \wedge dx_2, \end{aligned}$$

as required.

Proof of Lemma 12.6 in the case when $n = 3$. In this case

$$\begin{aligned} d\varphi_1 &= \frac{\partial \varphi_1}{\partial x_1}\, dx_1 + \frac{\partial \varphi_1}{\partial x_2}\, dx_2 + \frac{\partial \varphi_1}{\partial x_3}\, dx_3, \\ d\varphi_2 &= \frac{\partial \varphi_2}{\partial x_1}\, dx_1 + \frac{\partial \varphi_2}{\partial x_2}\, dx_2 + \frac{\partial \varphi_2}{\partial x_3}\, dx_3, \\ d\varphi_3 &= \frac{\partial \varphi_3}{\partial x_1}\, dx_1 + \frac{\partial \varphi_3}{\partial x_2}\, dx_2 + \frac{\partial \varphi_3}{\partial x_3}\, dx_3. \end{aligned}$$

Thus

$$\begin{aligned} d\varphi_1 \wedge d\varphi_2 \wedge d\varphi_3 &= \quad \frac{\partial \varphi_1}{\partial x_1}\frac{\partial \varphi_2}{\partial x_2}\frac{\partial \varphi_3}{\partial x_3}\, dx_1 \wedge dx_2 \wedge dx_3 \\ &\quad + \frac{\partial \varphi_1}{\partial x_2}\frac{\partial \varphi_2}{\partial x_3}\frac{\partial \varphi_3}{\partial x_1}\, dx_2 \wedge dx_3 \wedge dx_1 \\ &\quad + \frac{\partial \varphi_1}{\partial x_3}\frac{\partial \varphi_2}{\partial x_1}\frac{\partial \varphi_3}{\partial x_2}\, dx_3 \wedge dx_1 \wedge dx_2 \\ &\quad + \frac{\partial \varphi_1}{\partial x_3}\frac{\partial \varphi_2}{\partial x_2}\frac{\partial \varphi_3}{\partial x_1}\, dx_3 \wedge dx_2 \wedge dx_1 \\ &\quad + \frac{\partial \varphi_1}{\partial x_1}\frac{\partial \varphi_2}{\partial x_3}\frac{\partial \varphi_3}{\partial x_2}\, dx_1 \wedge dx_3 \wedge dx_2 \\ &\quad + \frac{\partial \varphi_1}{\partial x_2}\frac{\partial \varphi_2}{\partial x_1}\frac{\partial \varphi_3}{\partial x_3}\, dx_2 \wedge dx_1 \wedge dx_3 \\ &= \Bigl(\frac{\partial \varphi_1}{\partial x_1}\frac{\partial \varphi_2}{\partial x_2}\frac{\partial \varphi_3}{\partial x_3} + \frac{\partial \varphi_1}{\partial x_2}\frac{\partial \varphi_2}{\partial x_3}\frac{\partial \varphi_3}{\partial x_1} \\ &\quad + \frac{\partial \varphi_1}{\partial x_3}\frac{\partial \varphi_2}{\partial x_1}\frac{\partial \varphi_3}{\partial x_2} - \frac{\partial \varphi_1}{\partial x_3}\frac{\partial \varphi_2}{\partial x_2}\frac{\partial \varphi_3}{\partial x_1} \end{aligned}$$

$$
\left. - \frac{\partial\varphi_1}{\partial x_1}\frac{\partial\varphi_2}{\partial x_3}\frac{\partial\varphi_3}{\partial x_2} - \frac{\partial\varphi_1}{\partial x_2}\frac{\partial\varphi_2}{\partial x_1}\frac{\partial\varphi_3}{\partial x_3}\right) dx_1 \wedge dx_2 \wedge dx_3
$$

$$
= \begin{vmatrix} \dfrac{\partial\varphi_1}{\partial x_1} & \dfrac{\partial\varphi_1}{\partial x_2} & \dfrac{\partial\varphi_1}{\partial x_3} \\ \dfrac{\partial\varphi_2}{\partial x_1} & \dfrac{\partial\varphi_2}{\partial x_2} & \dfrac{\partial\varphi_2}{\partial x_3} \\ \dfrac{\partial\varphi_3}{\partial x_1} & \dfrac{\partial\varphi_3}{\partial x_2} & \dfrac{\partial\varphi_3}{\partial x_3} \end{vmatrix} dx_1 \wedge dx_2 \wedge dx_3,
$$

as required.

Remark One can derive Lemma 12.6 immediately from a theorem proved in Course 211. This theorem states that if $u_1, u_2, \ldots, u_r$ and $v_1, v_2, \ldots v_r$ are elements of some vector space V and if

$$
u_j = \sum_{i=1}^{r} a_j^i v_j
$$

for $i = 1, 2, \ldots, r$ then

$$
u_1 \wedge u_2 \wedge \cdots \wedge u_r = (\det A)\, v_1 \wedge v_2 \wedge \cdots \wedge v_r,
$$

where A is the $r \times r$ matrix of scalar coefficients given by

$$
A = \begin{pmatrix} a_1^1 & a_2^1 & \ldots & a_r^1 \\ a_1^2 & a_2^2 & \ldots & a_r^2 \\ \vdots & \vdots & \ddots & \ldots \\ a_1^3 & a_2^3 & \ldots & a_r^3 \end{pmatrix}.
$$

Remark We conclude immediately from Lemma 12.6 that if $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ is a smooth curvilinear coordinate system defined over some open set D in $\mathbb{R}^n$ then

$$
d\varphi_1 \wedge d\varphi_2 \wedge \cdots \wedge d\varphi_n = (\det \varphi')\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n,
$$

where φ' is the Jacobian matrix defined by

$$
\varphi' = \begin{pmatrix} \dfrac{\partial\varphi_1}{\partial x_1} & \dfrac{\partial\varphi_1}{\partial x_2} & \cdots & \dfrac{\partial\varphi_1}{\partial x_n} \\ \dfrac{\partial\varphi_2}{\partial x_1} & \dfrac{\partial\varphi_2}{\partial x_2} & \cdots & \dfrac{\partial\varphi_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial\varphi_n}{\partial x_1} & \dfrac{\partial\varphi_n}{\partial x_2} & \cdots & \dfrac{\partial\varphi_n}{\partial x_n} \end{pmatrix}.
$$

13 Integration of Differential Forms

Let ω be a continuous n-form defined over an open set D in $\mathbb{R}^n$, and let V be a closed bounded Jordan-measurable subset of D. We can write

$$\omega = f(x_1, x_2, \ldots, x_n)\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n,$$

where $f\colon D \to \mathbb{R}$ is a continuous function on D. We recall that the Riemann integral of continuous function over a closed bounded Jordan-measurable set in $\mathbb{R}^n$ is well-defined. Thus let us define

$$\int_V \omega \equiv \int_V f(x_1, x_2, \ldots, x_n)\, dx_1 dx_2 \ldots dx_n.$$

Lemma 13.1 *Let ω be a continuous n-form on some open set D in $\mathbb{R}^n$. Let $(x_1, x_2, \ldots, x_n)$ be the standard Cartesian coordinates on $\mathbb{R}^n$. Let π be a permutation of the set $\{1, 2, \ldots, n\}$, and let $j_i = \pi(i)$ for $i = 1, 2, \ldots, n$. Suppose that*

$$\omega = f(\mathbf{x})\, dx_{j_1} \wedge dx_{j_2} \wedge \cdots \wedge dx_{j_n}.$$

Then

$$\int_V \omega = \epsilon_\pi \int_V f(\mathbf{x})\, dx_1 dx_2 \ldots dx_n$$

for all closed bounded Jordan-measurable subsets V of $\mathbb{R}^n$, where

$$\epsilon_\pi = \begin{cases} +1 & \text{if the permutation } \pi \text{ is even;} \\ -1 & \text{if the permutation } \pi \text{ is odd.} \end{cases}$$

Proof This follows directly from the definition of the integral of ω, using the fact that

$$dx_{j_1} \wedge dx_{j_2} \wedge \cdots \wedge dx_{j_n} = \epsilon_\pi dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n. \quad \blacksquare$$

Example Let ω be a 3-form on $\mathbb{R}^3$ which has the form

$$\omega = f(x, y, z)\, dz \wedge dy \wedge dx.$$

Then

$$\int_V \omega = -\int_V f(x, y, z)\, dx\, dy\, dz$$

for all closed bounded Jordan-measurable subsets V of $\mathbb{R}^3$.

Using the Change of Variables Formula for multiple integrals (Theorem 11.7 we can deduce the following result.

Theorem 13.2 *Let D and E be open sets in $\mathbb{R}^n$ and let $\varphi\colon D \to E$ be a diffeomorphism from D to E. Let ω be a continuous n-form on E, and let V be a closed bounded Jordan-measurable subset of D. If $\varphi\colon D \to E$ is orientation-preserving then*

$$\int_{\varphi(V)} \omega = \int_V \varphi^*\omega,$$

(where $\varphi^*\omega$ is the pullback of ω under the map φ. Similarly if $\varphi\colon D \to E$ is orientation-reversing then

$$\int_{\varphi(V)} \omega = -\int_V \varphi^*\omega.$$

Proof We can write

$$\omega = f(x_1, x_2, \ldots, x_n)\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n,$$

where $f\colon E \to \mathbb{R}$ is continuous. If we apply the change of variables formula (Theorem 11.7) we see that

$$\begin{aligned} \int_{\varphi(V)} \omega &= \int_{\varphi(V)} f(\mathbf{x})\, dx_1 dx_2 \ldots dx_n. \\ &= \int_V f(\varphi(\mathbf{x}))|\det \varphi'(\mathbf{x})|\, dx_1 dx_2 \ldots dx_n. \end{aligned}$$

On the other hand

$$\begin{aligned} \varphi^*\omega &= (f \circ \varphi)\, d\varphi_1 \wedge d\varphi_2 \wedge \cdots \wedge d\varphi_n \\ &= (f \circ \varphi)(\det \varphi')\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n. \end{aligned}$$

by Lemma 12.6. If $\varphi\colon D \to E$ is orientation-preserving then $\det \varphi'(\mathbf{x}) > 0$ for all $\mathbf{x} \in D$, and hence

$$\int_{\varphi(V)} \omega = \int_V \varphi^*\omega.$$

Similarly if $\varphi\colon D \to E$ is orientation-reversing then $\det \varphi'(\mathbf{x}) < 0$ for all $\mathbf{x} \in D$, and hence

$$\int_{\varphi(V)} \omega = -\int_V \varphi^*\omega. \quad \blacksquare$$

Corollary 13.3 *Let ω be a continuous n-form defined over some open subset D of $\mathbb{R}^n$. Let $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ be a smooth positively-oriented curvilinear coordinate system on D. Suppose that*

$$\omega = F(\varphi_1, \varphi_2, \ldots, \varphi_n)\, d\varphi_1 \wedge d\varphi_2 \wedge \cdots \wedge d\varphi_n,$$

Let V be a closed bounded Jordan-measurable subset of D. Then

$$\int_V \omega = \int_{\varphi(V)} F(u_1, u_2, \ldots, u_n)\, du_1\, du_2 \ldots du_n.$$

Proof Let $E = \varphi(D)$, where

$$\varphi(\mathbf{x}) = (\varphi_1(\mathbf{x}), \varphi_2(\mathbf{x}), \ldots, \varphi_n(\mathbf{x})).$$

The diffeomorphism $\varphi D \to E$ is orientation-preserving (since the coordinate system $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ is positively oriented. Moreover $\omega = \varphi^*\eta$, where η is the n-form on E defined by

$$\eta = F\, dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n.$$

It follows that

$$\begin{aligned} \int_V \omega &= \int_V \varphi^*\eta = \int_{\varphi(V)} \eta \\ &= \int_{\varphi(V)} F(u_1, u_2, \ldots, u_n)\, du_1\, du_2 \ldots du_n, \end{aligned}$$

as required. ∎

We may deduce from this the following result.

Corollary 13.4 *Let ω be a continuous n-form defined over some open subset D of $\mathbb{R}^n$. Let $(\varphi_1, \varphi_2, \ldots, \varphi_n)$ be a smooth positively-oriented curvilinear coordinate system on D. Let π be a permutation of the set $\{1, 2, \ldots, n\}$. Suppose that*

$$\omega = F(\varphi_1, \varphi_2, \ldots, \varphi_n)\, d\varphi_{j_1} \wedge d\varphi_{j_2} \wedge \cdots \wedge d\varphi_{j_n},$$

where $j_i = \pi(i)$ for each i. Let V be a closed bounded Jordan-measurable subset of D. Then

$$\int_V \omega = \epsilon_\pi \int_{\varphi(V)} F(u_1, u_2, \ldots, u_n)\, du_1\, du_2 \ldots du_n.$$

where

$$\epsilon_\pi = \begin{cases} 1 & \text{if } \pi \text{ is even;} \\ -1 & \text{if } \pi \text{ is odd.} \end{cases}$$

Example Let (u, v) be the smooth curvilinear coordinate system on $\mathbb{R}^2$ defined by

$$u = x, \qquad v = y - x^2.$$

Then

$$\begin{vmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \\ \dfrac{\partial v}{\partial x} & \dfrac{\partial v}{\partial y} \end{vmatrix} = 1,$$

hence the coordinate system (u, v) is positively oriented. Let Q be the region in $\mathbb{R}^2$ defined by

$$Q = \{(x, y) \in \mathbb{R}^2 : -1 \leq x \leq 1, \quad x^2 \leq y \leq x^2 + 1\},$$

and let ω be the 2-form defined by $\omega = (y^2 - 2yx^2 + x^4)dx \wedge dy$. Now

$$du = dx, \qquad dv = dy - 2x\, dx,$$

hence $du \wedge dv = dx \wedge dy$. It follows that $\omega = v^2 du \wedge dv$ and hence

$$\int_Q \omega = \int_{v=0}^{1} \int_{u=-1}^{1} v^2\, du\, dv = \frac{2}{3}.$$

Example Let (r, θ, φ) be spherical polar coordinates on $\mathbb{R}^3$. Let ω be the 3-form on $\mathbb{R}^3$ defined by $\omega = r^2 \sin\theta\, dr \wedge d\theta \wedge d\varphi$. Let V be the closed set in $\mathbb{R}^3$ defined by

$$B = \{(x, y, z) \in \mathbb{R}^3 : x \geq 0, \quad 3z^2 \leq x^2 + y^2, \quad 1 \leq x^2 + y^2 + z^2 \leq 4\}.$$

Note that the set B is the set of all points in $\mathbb{R}$ for which $1 \leq r \leq 2$, $-\pi/2 \leq \varphi \leq \pi/2$ and $\pi/3 \leq \theta \leq 2\pi/3$. Therefore

$$\begin{aligned} \int_B \omega &= \int_{\varphi=-\frac{\pi}{2}}^{\frac{\pi}{2}} \int_{\theta=\frac{\pi}{3}}^{\frac{2\pi}{3}} \int_{r=1}^{2} r^2 \sin\theta\, dr\, d\theta\, d\varphi = \frac{7\pi}{3} \int_{\frac{pi}{3}}^{\frac{2\pi}{3}} \sin\theta\, d\theta \\ &= \frac{14\pi}{3} \cos\frac{\pi}{3} = \frac{7\pi}{3}. \end{aligned}$$

Example Let (r, θ) be polar coordinates on $\mathbb{R}^2$. The coordinate system (r, θ) is positively oriented. Consider the 2-form η defined by $\eta = r^3 d\theta \wedge dr$. Let D be the closed unit disk in $\mathbb{R}^2$ defined by

$$D\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

Then

$$\int_D \eta = -\int_{\theta=0}^{2\pi}\int_{r=0}^{1} r^3\, dr\, d\theta = -\frac{\pi}{2}.$$

(Note the minus sign in this formula: this occurs because the coordinate system (r,θ) is positively oriented, whereas the coordinate system (θ,r) is negatively oriented. Thus if f is any function on D then

$$\int_D f\, d\theta \wedge dr = -\int_D f\, dr\, d\theta,$$

by Corollary 13.4.)

13.1 Line Integrals

We define a *parameterized smooth curve* in $\mathbb{R}^n$ to be a smooth map $\gamma\colon [a,b] \to \mathbb{R}^n$ which maps some interval $[a,b]$ into $\mathbb{R}^n$.

Definition Let η be a continuous 1-form on some open set D in $\mathbb{R}^n$, given by

$$\eta = f_1\, dx_1 + f_2\, dx_2 + \cdots + f_n\, dx_n$$

for some continuous real-valued functions $f_1, f_2, \ldots f_n$ on D. Let $\gamma\colon [a,b] \to \mathbb{R}^n$ be a parameterized smooth curve in D. let $\gamma_1, \gamma_2, \ldots, \gamma_n$ be the components of the map γ (so that

$$\gamma(t) = (\gamma_1(t), \gamma_2(t), \ldots \gamma_n(t))$$

for all $t \in [a,b]$). We define the *line integral* of η along the curve γ by

$$\int_\gamma \eta = \sum_{i=1}^{n} \int_a^b f_i(\gamma(t)) \frac{d\gamma_i(t)}{dt}\, dt.$$

Observe that

$$\int_\gamma \eta = \int_{[a,b]} \gamma^* \eta$$

for all 1-forms η defined over some open set containing the image of the curve γ.

An important property of line integrals is their invariance under smooth reparameterizations of the curves along which the line integral is taken. Suppose that $\alpha\colon [a,b] \to \mathbb{R}^n$ and $\beta\colon [c,d] \to \mathbb{R}^n$ are smooth curves in $\mathbb{R}$. We say that the curve β is a *smooth reparameterization* of the curve α if and only if $\beta = \alpha \circ \rho$ for some smooth increasing function $\rho\colon [c,d] \to [a,b]$, where $\rho(c) = a$, $\rho(d) = b$.

Lemma 13.5 *Let η a continuous 1-form defined on some open set D in $\mathbb{R}^n$ and let $\alpha\colon [a,b] \to D$ and $\beta\colon [c,d] \to D$ be smooth curves in D. Suppose that β is a smooth reparameterization of the curve α. Then*

$$\int_\beta \eta = \int_\alpha \eta.$$

Proof Let η be given by

$$\eta = f_1\, dx_1 + f_2\, dx_2 + \cdots + f_n\, dx_n$$

for some continuous real-valued functions $f_1, f_2, \ldots f_n$ on D. Now $\beta = \alpha \circ \rho$ for some smooth increasing function $\rho\colon [c,d] \to [a,b]$, where $\rho(c) = a$, $\rho(d) = b$. Thus

$$\begin{aligned} \int_\beta \eta &= \sum_{i=1}^n \int_c^d f_i(\beta(t))\beta_i'(t)\, dt \\ &= \sum_{i=1}^n \int_c^d f_i(\alpha(\rho(t))\alpha_i'(\rho(t))\rho'(t)\, dt \\ &= \sum_{i=1}^n \int_a^b f_i(\alpha(u)\alpha_i'(u)\, du \\ &= \int_\alpha \eta \end{aligned}$$

(where we have used the Chain Rule for functions of one variable, together with the rule for integration by substitution) ∎.

The sign of the line integral of a 1-form η along a curve γ depends on the direction in which the curve γ is traversed. Indeed suppose that $\alpha\colon [a,b] \to \mathbb{R}^n$ and $\beta\colon [c,d] \to \mathbb{R}^n$ are parameterized smooth curves in $\mathbb{R}^n$, where $\beta = \alpha \circ \rho$ for some smooth *decreasing* function $\rho\colon [c,d] \to [a,b]$ with $\rho(c) = b$ and $\rho(d) = a$. Then

$$\int_\beta \eta = -\int_\alpha \eta.$$

Thus if we change the direction in which the curve is traversed then this changes the sign of the line integral of any 1-form along that curve.

Example Let η be the smooth 1-form on $\mathbb{R}^2 \setminus \{(0,0)\}$ defined by

$$\eta = \frac{x}{x^2+y^2}\, dy - \frac{y}{x^2+y^2}\, dx.$$

We integrate η around the unit circle in $\mathbb{R}^2$ traversed in the anticlockwise direction. Thus let the parameterization of the unit circle be given by $\gamma\colon [0, 2\pi] \to \mathbb{R}^2$, where

$$\gamma(t) = (\cos t, \sin t).$$

Then

$$\begin{aligned} \int_\gamma \eta &= \int_0^{2\pi} \cos t \frac{d(\sin t)}{dt}\, dt - \int_0^{2\pi} \sin t \frac{d(\cos t)}{dt}\, dt \\ &= \int_0^{2\pi} (\cos^2 t + \sin^2 t)\, dt = 2\pi. \end{aligned}$$

Example Let η be the 1-form on $\mathbb{R}^3$ defined by

$$\eta = x^2\, dy + xy\, dx - x\, dz.$$

Let $\gamma\colon [0, 1] \to \mathbb{R}^3$ be the parameterized smooth curve defined by $\gamma(t) = (t, t^2, t^3)$. Then

$$\int_\gamma \eta = \int_0^1 \left(t^2 \frac{d(t^2)}{dt} + t^3 \frac{dt}{dt} - t \frac{d(t^3)}{dt} \right) dt = 0.$$

Lemma 13.6 *Let D be an open set in $\mathbb{R}^n$ and let $\gamma\colon [a, b] \to D$ be a smooth curve in D. Let $f\colon D \to \mathbb{R}$ be a continuously differentiable real-valued function on D. Then*

$$\int_\gamma df = f(\mathbf{v}) - f(\mathbf{u}),$$

where $\mathbf{u}$ and $\mathbf{v}$ are the endpoints of the curve γ, given by $\mathbf{u} = \gamma(a)$ and $\mathbf{v} = \gamma(b)$.

Proof The 1-form df if defined by

$$df = \sum_{i=1}^n \frac{\partial f}{\partial x_i} dx_i.$$

Thus if we write $\gamma(t) = (\gamma_1(t), \gamma_2(t), \ldots, \gamma_n(t))$ for all $t \in [a, b]$ then

$$\begin{aligned} \int_\gamma df &= \sum_{i=1}^n \int_a^b \left.\frac{\partial f(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\gamma(t)} \frac{d\gamma_i(t)}{dt}\, dt \\ &= \int_a^b \frac{df(\gamma(t))}{dt}\, dt \\ &= f(\gamma(b) - f(\gamma(a)), \end{aligned}$$

by the Chain Rule, as required. ∎

Let D be an open set in $\mathbb{R}^n$. A curve $\gamma\colon [a,b] \to D$ is said to be *closed* if and only if $\gamma(b) = \gamma(a)$. We conclude immediately from Lemma 13.6 that if $f\colon D \to \mathbb{R}$ is continuously differentiable then

$$\int_\gamma df = 0$$

for all smooth closed curves γ in D.

Example Let D be the open set in $\mathbb{R}^2$ defined by $D = \mathbb{R}^2 \setminus \{0,0\}$ (i.e., D is obtained from $\mathbb{R}^2$ by removing the origin). Let η be the 1-form on D defined by

$$\eta = \frac{x}{x^2+y^2}\,dy - \frac{y}{x^2+y^2}\,dx.$$

Then

$$\begin{aligned} d\eta &= \frac{\partial}{\partial x}\left(\frac{x}{x^2+y^2}\right) dx \wedge dy + \frac{\partial}{\partial y}\left(\frac{y}{x^2+y^2}\right) dx \wedge dy \\ &= \left(\frac{y^2-x^2}{x^2+y^2} + \frac{x^2-y^2}{x^2+y^2}\right) dx \wedge dy \\ &= 0. \end{aligned}$$

However if $\gamma\colon [0,2\pi] \to D$ is the closed curve in D given by $\gamma(t) = (\cos t, \sin t)$ (so that the curve γ winds once around to origin in the anticlockwise direction) then

$$\int_\gamma \eta = 2\pi.$$

We conclude that there cannot exist a smooth function $f\colon D \to \mathbb{R}$ on D with the property that $df = \eta$, since the integral of η around the smooth closed curve γ is non-zero. Thus the conclusions of the Poincaré Lemma do not apply on the domain D defined above (i.e., there exists a differential form η on D which satisfies $d\eta = 0$ but which is not of the form df for any smooth function f defined over the *whole* of the domain D). Note that the domain D is not star-shaped.

We can also integrate 1-forms along piecewise-smooth curves in $\mathbb{R}^n$. A continuous curve $\gamma\colon [a,b] \to \mathbb{R}^n$ is said to be *piecewise-smooth* if and only if there exists a partition $\{t_0, t_1, \ldots, t_r\}$ of the interval $[a,b]$, where

$$a = t_0 < t_1 < \cdots < t_{r-1} < t_r = b,$$

such that γ is smooth on $[t_{j-1}, t_j]$ for $j = 1, 2, \ldots, r$. If η is a continuous 1-form defined along the piecewise-smooth curve γ then we define

$$\int_\gamma \eta = \sum_{j=1}^{r} \int_{\gamma|[t_{j-1}, t_j]} \eta$$

(where $\gamma|[t_{j-1}, t_j]\colon [t_{j-1}, t_j] \to \mathbb{R}^n$ is the restriction of the curve γ to the closed interval $[t_{j-1}, t_j]$).

13.2 Surface Integrals

We have seen how to integrate 1-forms along smooth curves. Similarly we can integrate 2-forms over oriented smooth surface. First we define the concept of a smooth surface.

Definition Let S be a subset of $\mathbb{R}^n$. Then S is said to be a smooth surface (without boundary) in $\mathbb{R}^n$ if and only if, for each point $\mathbf{p}$ of S there exists an open set U in $\mathbb{R}^n$ containing $\mathbf{p}$ and a smooth curvilinear coordinate system $(\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_n)$ defined on U such that

$$S \cap U = \{\mathbf{x} \in U : \tilde{\varphi}_i(\mathbf{x}) = 0 \text{ for } i = 3, \ldots, n\}.$$

Definition Let S be a smooth surface in $\mathbb{R}^n$, and let $(\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_n)$ be a smooth curvilinear coordinate system defined over some open set U in $\mathbb{R}^n$. We say that this coordinate system is *adapted to the surface* S if and only if

$$S \cap U = \{\mathbf{x} \in U : \tilde{\varphi}_i(\mathbf{x}) = 0 \text{ for } i = 3, \ldots, n\}.$$

Example Let r, θ and φ be spherical polar coordinates defined over some suitable open set U in $\mathbb{R}^3$. Thus if (x, y, z) are the standard Cartesian coordinates on $\mathbb{R}^3$ then

$$x = r \sin\theta \cos\varphi, \qquad y = r \sin\theta \sin\varphi, \qquad z = r \cos\theta$$

at each point of the open set U. We see that, for each positive number R, the curvilinear coordinate system (θ, φ, r) is adapted to the sphere of radius R about the origin, where this sphere is defined to be the set

$$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = R^2.$$

Example Let ρ, φ and z be cylindrical polar coordinates defined over some suitable open set U in $\mathbb{R}^3$. Thus if (x, y, z) are the standard Cartesian coordinates on $\mathbb{R}^3$ then

$$x = \rho \cos\varphi, \qquad y = \rho \sin\varphi.$$

at each point of the open set U. We see that, for each positive number R the curvilinear coordinate system (φ, z, r) is adapted to the cylinder of radius R about the z-axis, where this cylinder is defined to be the set

$$\{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = R^2\}.$$

Definition Let S be a smooth surface in $\mathbb{R}^n$, and let V be a subset of S. Let U be an open set in $\mathbb{R}^n$ which contains V and let $(\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_n)$, be a smooth curvilinear coordinate system for $\mathbb{R}^n$ which is defined over U and which is adapted to the surface S. Let $\varphi_1: V \to \mathbb{R}$ and $\varphi_2: V \to \mathbb{R}$ be the restrictions of $\tilde{\varphi}_1$ and $\tilde{\varphi}_2$ respectively to V. Then we say that (φ_1, φ_2) is a *smooth coordinate system for the surface* S defined over the set V.

Example Let S^2 denote the unit sphere in $\mathbb{R}^3$ defined by

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}.$$

Let r, θ and φ be spherical polar coordinates defined over some suitable open set U in $\mathbb{R}^3$. Then (θ, φ) is a smooth coordinate system for the sphere S^2 defined over $S^2 \cap U$.

Example Let Z be the cylinder in $\mathbb{R}^3$ defined by

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1\}.$$

Let ρ, φ and z be cylindrical polar coordinates defined over some suitable open set U in $\mathbb{R}^3$. Then (φ, z) is a smooth coordinate system for the cylinder Z defined over $Z \cap U$.

We now discuss the concept of *orientation* for a smooth surface S in $\mathbb{R}^n$.

Definition Let S be a smooth surface in $\mathbb{R}^n$, let V be a subset of S and let (φ_1, φ_2) and (ψ_1, ψ_2) be smooth coordinate systems for the surface S which are defined over V. We say that these two coordinate systems induce the same *orientation* on V if and only if

$$\begin{vmatrix} \dfrac{\partial \psi_1}{\partial \varphi_1} & \dfrac{\partial \psi_1}{\partial \varphi_2} \\ \dfrac{\partial \psi_2}{\partial \varphi_1} & \dfrac{\partial \psi_2}{\partial \varphi_2} \end{vmatrix} > 0$$

at all points of V.

Definition Let S be a smooth surface in $\mathbb{R}^n$ and let $\mathcal{A}$ be a collection of smooth coordinate systems, each defined over some suitable subset of S known as the *domain* of that coordinate system. The collection $\mathcal{A}$ of coordinate systems is said to be an *oriented atlas* if and only if the following conditions are satisfied:

(i) every point of S belongs to the domain of at least one of the smooth coordinate systems in the collection $\mathcal{A}$,

(ii) if (φ_1, φ_2) and (ψ_1, ψ_2) are smooth coordinate systems for the surface S which belong to the collection $\mathcal{A}$ and if U and V are the domains of these coordinate systems, then the coordinate systems (φ_1, φ_2) and (ψ_1, ψ_2) induce the same orientation on $U \cap V$.

The surface S is said to be *orientable* if and only if there exists an oriented atlas for S. Such an oriented atlas is said to define an *orientation* on the surface S. If such an orientation has been chosen for the surface S then we say that S is *oriented*.

Definition Let S be an oriented surface and let $\mathcal{A}$ be an oriented atlas for S which induces the chosen orientation on S. Let (φ_1, φ_2) be a smooth coordinate system for the surface S which is defined over some subset U of S. The coordinate system (φ_1, φ_2) is said to be *positively oriented* if and only if the following condition is satisfied:

if (ψ_1, ψ_2) is a smooth coordinate system with domain D which belongs to the atlas $\mathcal{A}$, then

$$\begin{vmatrix} \dfrac{\partial \psi_1}{\partial \varphi_1} & \dfrac{\partial \psi_1}{\partial \varphi_2} \\ \dfrac{\partial \psi_2}{\partial \varphi_1} & \dfrac{\partial \psi_2}{\partial \varphi_2} \end{vmatrix} > 0$$

at all points of $U \cap D$.

Similarly the coordinate system (φ_1, φ_2) is said to be *negatively oriented* if and only if the following condition is satisfied:

if (ψ_1, ψ_2) is a smooth coordinate system with domain D which belongs to the atlas $\mathcal{A}$, then

$$\begin{vmatrix} \dfrac{\partial \psi_1}{\partial \varphi_1} & \dfrac{\partial \psi_1}{\partial \varphi_2} \\ \dfrac{\partial \psi_2}{\partial \varphi_1} & \dfrac{\partial \psi_2}{\partial \varphi_2} \end{vmatrix} < 0$$

at all points of $U \cap D$.

We observe that if (φ_1, φ_2) and (ψ_1, ψ_2) are positively oriented coordinate systems for some oriented surface S defined over some subset V of S then

$$\begin{vmatrix} \dfrac{\partial \psi_1}{\partial \varphi_1} & \dfrac{\partial \psi_1}{\partial \varphi_2} \\ \dfrac{\partial \psi_2}{\partial \varphi_1} & \dfrac{\partial \psi_2}{\partial \varphi_2} \end{vmatrix} > 0$$

at all points of V.

Remark Let S be a smooth surface in $\mathbb{R}^n$. Suppose that it is possible to find a smooth curvilinear coordinate system (φ_1, φ_2) for the surface S which is defined over the whole of the surface S. Then one can specify an orientation on the surface S simply by stating whether this coordinate system on S is positively oriented or negatively oriented with respect to the chosen orientation on S.

Example Let S^2 be the 2-sphere in $\mathbb{R}^3$ defined by

$$S^2 = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}.$$

Then we can specify an orientation on S^2 by requiring that the coordinate system (θ, φ) on S^2 be positively oriented, where

$$x = \sin\theta \cos\varphi, \qquad y = \sin\theta \sin\varphi, \qquad z = \cos\theta.$$

Let S be an oriented surface in $\mathbb{R}^n$ and let V be a subset of S. Suppose that there exists a positively oriented coordinate system (φ_1, φ_2) for the surface S defined over V. Let us define the subset W of $\mathbb{R}^2$ by

$$W = \{(\xi, \eta) \in \mathbb{R}^2 : \xi = \varphi_1(\mathbf{p}) \text{ and } \eta = \varphi_2(\mathbf{p}) \text{ for some } \mathbf{p} \in V\}.$$

We let $\alpha: W \to V$ be the smooth map which expresses the standard Cartesian coordinates $(x_1, x_2, \ldots, x_n)$ at a point of V in terms of φ_1 and φ_2. Thus

$$x_i = \alpha_i(\varphi_1, \varphi_2) \qquad (i = 1, 2, \ldots, n),$$

where $\alpha_i: W \to \mathbb{R}$ is the ith component of the map α. One can readily verify that the map $\alpha: W \to V$ has the following properties:

(i) the subset W of the plane $\mathbb{R}^2$ is mapped homeomorphically onto V by the map α (i.e., $\alpha: W \to V$ is a bijection and both $\alpha: W \to V$ and $\alpha^{-1}: V \to W$ are continuous).

(ii) the Jacobian matrix of the map $\alpha\colon W \to \mathbb{R}^n$ has rank 2 at each point of W.

We refer to a map $\alpha\colon W \to V$ which arises from some coordinate system (φ_1, φ_2) for the surface S in the manner described above as a *local parameterization* of the surface.

Now let us suppose that the domain W of the local parameterization $\alpha\colon W \to V$ is Jordan-measurable. Then we define

$$\int_V \omega = \int_W \alpha^* \omega.$$

for each continuous 2-form defined over some open set which contains V. We now show that the value of this integral is well-defined independently of the choice of the positively oriented smooth coordinate system (φ_1, φ_2) chosen on V.

Suppose that (φ_1, φ_2) and (ψ_1, ψ_2) are smooth positively oriented coordinate systems for the oriented surface S defined over the subset V of S. We define

$$\begin{aligned} W &= \{(\xi, \eta) \in \mathbb{R}^2 : \xi = \varphi_1(\mathbf{p}) \text{ and } \eta = \varphi_2(\mathbf{p}) \text{ for some } \mathbf{p} \in V\}, \\ \tilde{W} &= \{(\xi, \eta) \in \mathbb{R}^2 : \xi = \psi_1(\mathbf{p}) \text{ and } \eta = \psi_2(\mathbf{p}) \text{ for some } \mathbf{p} \in V\}. \end{aligned}$$

We let $\alpha\colon W \to V$ and $\beta\colon \tilde{W} \to V$ be the local parameterizations of the surface defined such that

$$x_i = \alpha_i(\varphi_1, \varphi_2) = \beta_i(\psi_1, \psi_2) \qquad (i = 1, 2, \ldots, n),$$

where α_i and β_i denote the ith components of the maps α and β respectively. There exists a diffeomorphism $\rho\colon W \to \tilde{W}$ characterized by the property that

$$(\psi_1(\mathbf{p}), \psi_2(\mathbf{p})) = \rho(\varphi_1(\mathbf{p}, \varphi_2(\mathbf{p})$$

for all $\mathbf{p} \in V$. This diffeomorphism $\rho\colon W \to \tilde{W}$ is orientation preserving (since the positively oriented coordinate systems (φ_1, φ_2) and (ψ_1, ψ_2) determine the same orientation on V). It follows from the definitions of α, β and ρ that $\alpha = \beta \circ \rho$. Therefore

$$\int_W \alpha^* \omega = \int_W \rho^*(\beta^* \omega) = \int_{\tilde{W}} \beta^* \omega$$

by Theorem 13.2. This shows that the value of the integral $\int_V \omega$ is well-defined independently of the choice of the positively oriented smooth coordinate system (φ_1, φ_2) chosen on V, as required.

The discussion above explains how to calculate the integral of a 2-form over a portion V of some oriented smooth surface S in $\mathbb{R}^n$, provided that the region V is contained within the domain of some smooth coordinate system for the surface S and provided that V corresponds by means of this coordinate system to a closed bounded Jordan-measurable subset of $\mathbb{R}^2$.

Now suppose that V is some portion of an oriented smooth surface S, where the region V is a closed bounded set contained in S which is bounded by piecewise-smooth curves lying in the surface S. It may not be possible to find a smooth coordinate system defined over the whole of V. However if V is a closed bounded set then one can decompose V into subregions $V_1, V_2, \ldots, V_r$, where each of these subregions of the surface S is bounded by a piecewise smooth curve in S and is contained in the domain of some positively oriented smooth coordinate system S. The intersection of any pair of these subregions $V_1, V_2, \ldots, V_r$ should be contained within the boundary of these subregions. Let ω be a continuous 2-form defined on V. The integral of the 2-form ω over each of the regions $V_1, V_2, \ldots, V_r$ can be defined as described above, since each of these regions is contained within the domain of some smooth coordinate system on the surface S. Therefore we define

$$\int_V \omega = \sum_{j=1}^r \int_{V_j} \omega.$$

It is not difficult to show that the value of this integral is independent of the fashion in which we divide up the region V into the subregions $V_1, V_2, \ldots, V_r$.

Example Let S be a smooth surface in $\mathbb{R}^3$ of the form

$$S = \{(x, y, z) \in \mathbb{R}^3 : z = x^2 + y^2\}.$$

We let (u, v) be the coordinate system on S defined by $u = x$ and $v = y$. We choose the orientation on S such that (u, v) is a positively oriented coordinate system on S. The coordinate system (u, v) defines a local parameterization $\alpha \colon \mathbb{R}^2 \to \mathbb{R}^3$, where

$$x = \alpha_1(u, v), \qquad y = \alpha_2(u, v), \qquad z = \alpha_3(u, v).$$

Thus

$$\alpha_1(u, v) = u, \qquad \alpha_2(u, v) = v, \qquad \alpha_3(u, v) = u^2 + v^2.$$

Let ω be the 2-form on $\mathbb{R}^3$ defined by

$$\omega = x\, dy \wedge dz + y\, dz \wedge dx + z\, dx \wedge dy,$$

and let V be the region of the surface S defined by

$$V = \{(x, y, z) \in S : x^2 + y^2 \leq 1\}.$$

Then

$$\int_V \omega = \int_W \alpha^* \omega,$$

where $W = \{(u, v) \in \mathbb{R}^2 : u^2 + v^2 \leq 1\}$. Now

$$\begin{aligned} \alpha^* \omega &= u\, dv \wedge d(u^2 + v^2) + v\, d(u^2 + v^2) \wedge du + (u^2 + v^2)\, du \wedge dv \\ &= (-2u^2 - 2v^2 + u^2 + v^2)\, du \wedge dv = -(u^2 + v^2)\, du \wedge dv \end{aligned}$$

Therefore

$$\int_V \omega = -\int_W (u^2 + v^2)\, du\, dv = -2\pi \int_0^1 \rho^3\, d\rho = -\frac{\pi}{2}.$$

Example Let Z be the cylinder in $\mathbb{R}^3$ defined by

$$Z = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 = 1, \quad -1 \leq z \leq 1\}.$$

We choose the orientation on Z such that (φ, z) is a positively oriented coordinate system on Z, where $(x, y, z) = (\cos\varphi, \sin\varphi, z)$. Let W be the subset of $\mathbb{R}^2$ defined by

$$W = \{(\xi, \eta) \in \mathbb{R}^2 : 0 \leq \eta \leq 2\pi, \quad -1 \leq \eta \leq 1\}.$$

and let $\alpha\colon W \to Z$ be the parameterization of Z defined by

$$\alpha(\xi, \eta) = (\cos\xi, \sin\xi, \eta).$$

Let ω be the 2-form on $\mathbb{R}^2$ defined by

$$\omega = z^2 x\, dy \wedge dz = z^2 y\, dx \wedge dz.$$

Then

$$\begin{aligned} \alpha^* \omega &= \eta^2 \cos\xi\, d(\sin\xi) \wedge d\eta - \eta^2 \sin\xi\, d(\cos\xi) \wedge d\eta \\ &= \eta^2 (\cos^2\xi d\xi \wedge d\eta + \sin^2\xi d\xi \wedge d\eta) \\ &= \eta^2\, d\xi \wedge d\eta, \end{aligned}$$

and hence

$$\int_Z \omega = \int_W \alpha^* \omega = \int_{\eta=-1}^1 \int_{\xi=0}^{2\pi} \eta^2\, d\xi\, d\eta = \frac{4\pi}{3}.$$

13.3 Smooth Surfaces in $\mathbb{R}^3$

Let Z be a smooth surface in $\mathbb{R}^3$. Given any smooth coordinate system (ξ, η) for the surface Z let us define

$$\frac{\partial \mathbf{r}}{\partial \xi} = \begin{pmatrix} \dfrac{\partial x}{\partial \xi} \\ \dfrac{\partial y}{\partial \xi} \\ \dfrac{\partial z}{\partial \xi} \end{pmatrix}, \qquad \frac{\partial \mathbf{r}}{\partial \eta} = \begin{pmatrix} \dfrac{\partial x}{\partial \eta} \\ \dfrac{\partial y}{\partial \eta} \\ \dfrac{\partial z}{\partial \eta} \end{pmatrix}.$$

The vector product

$$\frac{\partial \mathbf{r}}{\partial \xi} \times \frac{\partial \mathbf{r}}{\partial \eta}$$

is normal to the surface Z.

Lemma 13.7 *Let Z be a smooth surface in $\mathbb{R}^3$, and let (ξ, η) and (σ, τ) be smooth coordinate systems for the surface Z defined over some subset V of Z. Then*

$$\frac{\partial \mathbf{r}}{\partial \xi} \times \frac{\partial \mathbf{r}}{\partial \eta} = \begin{vmatrix} \dfrac{\partial \sigma}{\partial \xi} & \dfrac{\partial \sigma}{\partial \eta} \\ \dfrac{\partial \tau}{\partial \xi} & \dfrac{\partial \tau}{\partial \eta} \end{vmatrix} \frac{\partial \mathbf{r}}{\partial \sigma} \times \frac{\partial \mathbf{r}}{\partial \tau}$$

at each point of V.

Proof Using the fact that

$$\begin{aligned} \frac{\partial \mathbf{r}}{\partial \xi} &= \frac{\partial \mathbf{r}}{\partial \sigma}\frac{\partial \sigma}{\partial \xi} + \frac{\partial \mathbf{r}}{\partial \tau}\frac{\partial \tau}{\partial \xi}, \\ \frac{\partial \mathbf{r}}{\partial \eta} &= \frac{\partial \mathbf{r}}{\partial \sigma}\frac{\partial \sigma}{\partial \eta} + \frac{\partial \mathbf{r}}{\partial \tau}\frac{\partial \tau}{\partial \eta}, \end{aligned}$$

we see that

$$\begin{aligned} \frac{\partial \mathbf{r}}{\partial \xi} \times \frac{\partial \mathbf{r}}{\partial \eta} &= \left(\frac{\partial \mathbf{r}}{\partial \sigma}\frac{\partial \sigma}{\partial \xi} + \frac{\partial \mathbf{r}}{\partial \tau}\frac{\partial \tau}{\partial \xi}\right) \times \left(\frac{\partial \mathbf{r}}{\partial \sigma}\frac{\partial \sigma}{\partial \eta} + \frac{\partial \mathbf{r}}{\partial \tau}\frac{\partial \tau}{\partial \eta}\right) \\ &= \left(\frac{\partial \sigma}{\partial \xi}\frac{\partial \tau}{\partial \eta} - \frac{\partial \tau}{\partial \xi}\frac{\partial \sigma}{\partial \eta}\right) \frac{\partial \mathbf{r}}{\partial \sigma} \times \frac{\partial \mathbf{r}}{\partial \tau} \\ &= \begin{vmatrix} \dfrac{\partial \sigma}{\partial \xi} & \dfrac{\partial \sigma}{\partial \eta} \\ \dfrac{\partial \tau}{\partial \xi} & \dfrac{\partial \tau}{\partial \eta} \end{vmatrix} \frac{\partial \mathbf{r}}{\partial \sigma} \times \frac{\partial \mathbf{r}}{\partial \tau}, \end{aligned}$$

as required. ∎

Suppose that the surface Z is orientable. Let us choose an orientation on Z. Then this orientation determines a smooth normal vector field $\mathbf{n}$ on Z. Indeed suppose that (ξ, η) is a positively oriented smooth coordinate system for the surface Z. Then the vector product

$$\frac{\partial \mathbf{r}}{\partial \xi} \times \frac{\partial \mathbf{r}}{\partial \eta}$$

is normal to the surface Z. We define the unit normal vector field $\mathbf{n}$ to be the unit vector field on Z characterized by the property that

$$\frac{\partial \mathbf{r}}{\partial \xi} \times \frac{\partial \mathbf{r}}{\partial \eta} = \left| \frac{\partial \mathbf{r}}{\partial \xi} \times \frac{\partial \mathbf{r}}{\partial \eta} \right| \mathbf{n}$$

on the domain of the coordinate system (ξ, η). It follows from Lemma 13.7 that if (ξ, η) and (σ, τ) are positively oriented coordinate systems for the surface Z then the directions of the unit normal vector field $\mathbf{n}$ determined by these coordinate systems will be consistent, since

$$\begin{vmatrix} \dfrac{\partial \sigma}{\partial \xi} & \dfrac{\partial \sigma}{\partial \eta} \\ \dfrac{\partial \tau}{\partial \xi} & \dfrac{\partial \tau}{\partial \eta} \end{vmatrix} > 0,$$

at each point of the surface Z that belongs to the domain of both of these coordinate systems. Thus we conclude that if the surface Z is oriented then the orientation determines a smooth unit normal vector field $\mathbf{n}\colon Z \to \mathbb{R}^3$ over the whole of Z. Conversely if such a smooth normal vector field $\mathbf{n}$ can be defined over the whole of some smooth surface Z then Z is orientable, and the unit normal vector field $\mathbf{n}$ determines an orientation of the surface Z characterized by the property that a smooth coordinate system (ξ, η) for the surface Z is positively oriented if and only if

$$\frac{\partial \mathbf{r}}{\partial \xi} \times \frac{\partial \mathbf{r}}{\partial \eta} = f\, \mathbf{n},$$

where $f(\mathbf{p}) > 0$ at each point $\mathbf{p}$ of the domain of the coordinate system (ξ, η).

Let Z be an oriented smooth surface in $\mathbb{R}^3$, where the set Z is a closed bounded subset of $\mathbb{R}^3$. and let $\mathbf{n}$ be the unit normal vector field on Z determined by the orientation. Let $\mathbf{B}$ be a continuous vector field on Z. The surface integral

$$\int_Z \mathbf{B}.\mathbf{n}\, dS$$

is evaluated into the following manner:

(i) divide up the surface Z into regions $V_1, V_2, \ldots, V_r$ such that each of these regions is in the domain of some positively oriented smooth coordinate system for the surface Z,

(ii) choose a positively oriented coordinate system (ξ, η) on the region V_j and define

$$\int_{V_j} \mathbf{B}.\mathbf{n}\, dS = \int_W \mathbf{B}(x(\xi,\eta), y(\xi,\eta), z(\xi,\eta)).\left(\frac{\partial \mathbf{r}}{\partial \xi} \times \frac{\partial \mathbf{r}}{\partial \eta}\right) d\xi d\eta,$$

where

$$W = \{(u,v) \in \mathbb{R}^2 : u = \xi(\mathbf{p}) \text{ and } v = \eta(\mathbf{p}) \text{ for some } \mathbf{p} \in V_j\}.$$

(iii) Define

$$\int_Z \mathbf{B}.\mathbf{n}\, dS = \sum_{p=1}^{r} \int_{V_j} \mathbf{B}.\mathbf{n}\, dS,$$

where the integral over the region V_j is defined as described above.

One can verify (using the change of variables formula) that the value of the surface integral is independent of the manner in which the surface Z is partitioned into the regions $V_1, V_2, \ldots, V_r$ and is independent of the choice of coordinate system chosen on each of these regions.

Remark One can interpret the surface integral

$$\int_Z \mathbf{B}.\mathbf{n}\, dS$$

as the integral of the scalar product $\mathbf{B}.\mathbf{n}$ with respect to *surface area* on the surface Z. The reason for this is that if (ξ, η) is a smooth coordinate system on the surface Z then

$$\left(\frac{\partial \mathbf{r}}{\partial \xi}\,\delta\xi\right) \times \left(\frac{\partial \mathbf{r}}{\partial \eta}\,\delta\eta\right) = (\delta A)\mathbf{n}$$

where δA is the area of the paralellogram whose sides are determined by the vectors

$$\frac{\partial \mathbf{r}}{\partial \xi}\,\delta\xi \text{ and } \frac{\partial \mathbf{r}}{\partial \eta}\,\delta\eta.$$

Lemma 13.8 *Let Z be an oriented smooth surface in $\mathbb{R}^3$, and let $\mathbf{n}$ be the unit normal vector field on Z determined by the orientation on Z. Let $\mathbf{B}$ be a continuous vector field defined over some open set D in $\mathbb{R}^3$ which contains the surface Z. Let B_1, B_2 and B_3 be the Cartesian components of the vector field $\mathbf{B}$ and let ω be the 2-form on D defined by*

$$\omega = B_1\, dy \wedge dz + B_2\, dz \wedge dx + B_3\, dx \wedge dy.$$

Then

$$\int_Z \mathbf{B}.\mathbf{n}\, dS = \int_Z \omega.$$

Proof We may suppose (without loss of generality) that there exists some coordinate system (ξ, η) for the surface Z which is defined over the whole of Z (since otherwise we could subdivide the surface Z into portions possessing this property and then prove the result for the surface Z by proving it for each of these portions). Let W be the subset of $\mathbb{R}^2$ defined by

$$W = \{(u, v) \in \mathbb{R}^2 : u = \xi(\mathbf{p}) \text{ and } v = \eta(\mathbf{p}) \text{ for some } \mathbf{p} \in Z\}.$$

and let $\alpha\colon W \to Z$ be the parameterization of the surface Z defined by the coordinate system (ξ, η), so that

$$(x, y, z) = \alpha(\xi, \eta)$$

at each point of the surface Z. Then

$$\begin{aligned}
\alpha^*\omega &= B_1(\alpha(\xi,\eta)) \left(\frac{\partial y}{\partial \xi}\frac{\partial z}{\partial \eta} - \frac{\partial z}{\partial \xi}\frac{\partial y}{\partial \eta} \right) d\xi \wedge d\eta \\
&\quad + B_2(\alpha(\xi,\eta)) \left(\frac{\partial z}{\partial \xi}\frac{\partial x}{\partial \eta} - \frac{\partial x}{\partial \xi}\frac{\partial z}{\partial \eta} \right) d\xi \wedge d\eta \\
&\quad + B_3(\alpha(\xi,\eta)) \left(\frac{\partial x}{\partial \xi}\frac{\partial y}{\partial \eta} - \frac{\partial y}{\partial \xi}\frac{\partial x}{\partial \eta} \right) d\xi \wedge d\eta \\
&= \mathbf{B}(\alpha(\xi,\eta)). \left(\frac{\partial \mathbf{r}}{\partial \xi} \times \frac{\partial \mathbf{r}}{\partial \eta} \right) d\xi \wedge d\eta.
\end{aligned}$$

But

$$\frac{\partial \mathbf{r}}{\partial \xi} \times \frac{\partial \mathbf{r}}{\partial \eta} = \begin{pmatrix} \dfrac{\partial y}{\partial \xi}\dfrac{\partial z}{\partial \eta} - \dfrac{\partial z}{\partial \xi}\dfrac{\partial y}{\partial \eta} \\ \dfrac{\partial z}{\partial \xi}\dfrac{\partial x}{\partial \eta} - \dfrac{\partial x}{\partial \xi}\dfrac{\partial z}{\partial \eta} \\ \dfrac{\partial x}{\partial \xi}\dfrac{\partial y}{\partial \eta} - \dfrac{\partial y}{\partial \xi}\dfrac{\partial x}{\partial \eta} \end{pmatrix},$$

and hence

$$\begin{aligned}\int_Z \mathbf{B}.\mathbf{n}\,dS &= \int_W \mathbf{B}(\alpha(\xi,\eta)).\left(\frac{\partial \mathbf{r}}{\partial \xi}\times\frac{\partial \mathbf{r}}{\partial \eta}\right)d\xi d\eta \\ &= \int_W \alpha^*\omega = \int_Z \omega,\end{aligned}$$

as required. ∎

14 Stokes' Theorem for Differential Forms

We shall prove a generalization of Stokes' Theorem which applies to the integrals of smooth differential forms over submanifolds of $\mathbb{R}^n$. We first define the concept of a k-dimensional *smooth submanifold* of $\mathbb{R}^n$, and explain how to defined the integral of a continuous k-form over an oriented k-dimensional submanifold of $\mathbb{R}^n$. We shall then state the Generalized Stokes' Theorem (Theorem 14.1) and deduce a number of classical theorems of vector calculus that follow from this theorem. We conclude this section with a discussion of the proof of the Generalized Stokes' Theorem.

14.1 Submanifolds of $\mathbb{R}^n$

Definition Let M be a subset of $\mathbb{R}^n$, and let k be an integer between 1 and n. We say that M is a *smooth k-dimensional submanifold of* $\mathbb{R}^n$ *without boundary* if and only if the following condition is satisfied:

> if $\mathbf{p}$ is a point of M then there exists a smooth curvilinear coordinate system $(\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_n)$ defined over an open neighbourhood U of $\mathbf{p}$ in $\mathbb{R}^n$ such that
>
> $$M \cap U = \{\mathbf{x} \in U : \tilde{\varphi}_j(\mathbf{x}) = 0 \text{ for } j = k+1, \ldots, n\}.$$

Observe that a smooth 1-dimensional submanifold of $\mathbb{R}^n$ is a smooth curve in $\mathbb{R}^n$, and a smooth 2-dimensional submanifold is a smooth surface in $\mathbb{R}^n$. Thus, for any integer k between 1 and n, a k-dimensional smooth submanifold of $\mathbb{R}^n$ is the k-dimensional analogue of a smooth surface in $\mathbb{R}^n$.

Definition Let M be a smooth k-dimensional submanifold of $\mathbb{R}^n$ without boundary, and let $(\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_n)$ be a smooth curvilinear coordinate system defined over an open neighbourhood U of $\mathbf{p}$ in $\mathbb{R}^n$. We say that this coordinate system is *adapted to the submanifold S* if and only if

$$M \cap U = \{\mathbf{x} \in U : \tilde{\varphi}_j(\mathbf{x}) = 0 \text{ for } j = k+1, \ldots, n\}.$$

We now define the concept of a *submanifold of* $\mathbb{R}^n$ *with boundary.*

Definition Let M be a subset of $\mathbb{R}^n$ and let ∂M be a subset of M. Let k be an integer between 1 and n. We say that M is a *smooth k-dimensional submanifold of* $\mathbb{R}^n$ *with boundary* ∂M if and only if the following conditions are satisfied:

(i) the boundary ∂M of M is a smooth $(k-1)$-dimensional submanifold of $\mathbb{R}^n$ without boundary,

(ii) if $\mathbf{p}$ is a point of M then there exists a smooth curvilinear coordinate system $(\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_n)$ defined over an open neighbourhood U of $\mathbf{p}$ in $\mathbb{R}^n$ such that

$$M \cap U = \{\mathbf{x} \in U : \tilde{\varphi}_1(\mathbf{x}) \leq 0 \text{ and } \tilde{\varphi}_j(\mathbf{x}) = 0 \text{ for } j = k+1, \ldots, n\}$$

and

$$\partial M \cap U = \{\mathbf{x} \in U : \tilde{\varphi}_1(\mathbf{x}) = 0 \text{ and } \tilde{\varphi}_j(\mathbf{x}) = 0 \text{ for } j = k+1, \ldots, n\}$$

Definition Let M be a smooth submanifold of $\mathbb{R}^n$ of dimension k with boundary ∂M, and let $(\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_n)$ be a smooth curvilinear coordinate system defined over an open neighbourhood U of $\mathbf{p}$ in $\mathbb{R}^n$. We say that this coordinate system is *adapted to the submanifold S* if and only if

$$M \cap U = \{\mathbf{x} \in U : \tilde{\varphi}_1(\mathbf{x}) \leq 0 \text{ and } \tilde{\varphi}_j(\mathbf{x}) = 0 \text{ for } j = k+1, \ldots, n\}$$

and

$$\partial M \cap U = \{\mathbf{x} \in U : \tilde{\varphi}_1(\mathbf{x}) = 0 \text{ and } \tilde{\varphi}_j(\mathbf{x}) = 0 \text{ for } j = k+1, \ldots, n\}$$

Definition Let M be a smooth k-dimensional submanifold in $\mathbb{R}^n$ (with or without boundary), and let V be a subset of M. Let U be an open set in $\mathbb{R}^n$ which contains V and let $(\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_n)$, be a smooth curvilinear coordinate system for $\mathbb{R}^n$ which is defined over U which is adapted to the submanifold M. Let $\varphi_1, \varphi_2, \ldots, \varphi_k$ be the restrictions of $\tilde{\varphi}_1, \tilde{\varphi}_2, \ldots, \tilde{\varphi}_k$ to V. Then we say that $(\varphi_1, \varphi_2, \ldots, \varphi_k)$ is a *smooth coordinate system for the submanifold M* defined over the set V.

A subset S of $\mathbb{R}^n$ is said to be *compact* if and only if every open cover of the set S has a finite subcover. A well-known theorem states that a subset S of $\mathbb{R}^n$ is compact if and only if the set S is closed and bounded. If M is a compact submanifold of $\mathbb{R}^n$ then one can find a *finite* collection of coordinate systems for the submanifold M such that every point of M belongs to the domain of at least one of these coordinate systems.

We now discuss the concept of *orientation* for a smooth submanifold M in $\mathbb{R}^n$.

Definition Let M be a smooth submanifold in $\mathbb{R}^n$ of dimension k (with or without boundary), let V be a subset of M and let $(\varphi_1, \ldots, \varphi_k)$ and $(\psi_1, \ldots, \psi_k)$ be smooth coordinate systems for the submanifold M which are defined over V. We say that these two coordinate systems induce the same *orientation* on V if and only if

$$\begin{vmatrix} \dfrac{\partial \psi_1}{\partial \varphi_1} & \dfrac{\partial \psi_1}{\partial \varphi_2} & \cdots & \dfrac{\partial \psi_1}{\partial \varphi_k} \\ \dfrac{\partial \psi_2}{\partial \varphi_1} & \dfrac{\partial \psi_2}{\partial \varphi_2} & \cdots & \dfrac{\partial \psi_2}{\partial \varphi_k} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial \psi_k}{\partial \varphi_1} & \dfrac{\partial \psi_k}{\partial \varphi_2} & \cdots & \dfrac{\partial \psi_k}{\partial \varphi_k} \end{vmatrix} > 0$$

at all points of V.

Definition Let M be a smooth submanifold in $\mathbb{R}^n$ of dimension k (with or without boundary) and let $\mathcal{A}$ be a collection of smooth coordinate systems, each defined over some suitable subset of M known as the *domain* of that coordinate system. The collection $\mathcal{A}$ of coordinate systems is said to be an *oriented atlas* if and only if the following conditions are satisfied:

(i) every point of M belongs to the domain of at least one of the smooth coordinate systems in the collection $\mathcal{A}$,

(ii) if $(\varphi_1, \ldots, \varphi_k)$ and $(\psi_1, \ldots, \psi_k)$ are smooth coordinate systems for the submanifold M which belong to the collection $\mathcal{A}$ and if U and V are the domains of these coordinate systems, then the coordinate systems $(\varphi_1, \ldots, \varphi_k)$ and $(\psi_1, \ldots, \psi_k)$ induce the same orientation on $U \cap V$.

The submanifold M is said to be *orientable* if and only if there exists an oriented atlas for M. Such an oriented atlas is said to define an *orientation* on the submanifold M. If such an orientation has been chosen for the submanifold M then we say that M is *oriented*.

Definition Let M be an oriented submanifold of $\mathbb{R}^n$ of dimension k (with or without boundary) and let $\mathcal{A}$ be an oriented atlas for M which induces the chosen orientation on M. Let $(\varphi_1, \ldots, \varphi_k)$ be a smooth coordinate system for the submanifold M which is defined over some subset U of M. The coordinate system $(\varphi_1, \ldots, \varphi_k)$ is said to be *positively oriented* if and only if the following condition is satisfied:

if $(\psi_1, \ldots, \psi_k)$ is a smooth coordinate system with domain D which belongs to the atlas $\mathcal{A}$, then

$$\begin{vmatrix} \dfrac{\partial\psi_1}{\partial\varphi_1} & \dfrac{\partial\psi_1}{\partial\varphi_2} & \cdots & \dfrac{\partial\psi_1}{\partial\varphi_k} \\ \dfrac{\partial\psi_2}{\partial\varphi_1} & \dfrac{\partial\psi_2}{\partial\varphi_2} & \cdots & \dfrac{\partial\psi_2}{\partial\varphi_k} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial\psi_k}{\partial\varphi_1} & \dfrac{\partial\psi_k}{\partial\varphi_2} & \cdots & \dfrac{\partial\psi_k}{\partial\varphi_k} \end{vmatrix} > 0$$

at all points of $U \cap D$.

Similarly the coordinate system $(\varphi_1, \ldots, \varphi_k)$ is said to be *negatively oriented* if and only if the following condition is satisfied:

if $(\psi_1, \ldots, \psi_k)$ is a smooth coordinate system with domain D which belongs to the atlas $\mathcal{A}$, then

$$\begin{vmatrix} \dfrac{\partial\psi_1}{\partial\varphi_1} & \dfrac{\partial\psi_1}{\partial\varphi_2} & \cdots & \dfrac{\partial\psi_1}{\partial\varphi_k} \\ \dfrac{\partial\psi_2}{\partial\varphi_1} & \dfrac{\partial\psi_2}{\partial\varphi_2} & \cdots & \dfrac{\partial\psi_2}{\partial\varphi_k} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial\psi_k}{\partial\varphi_1} & \dfrac{\partial\psi_k}{\partial\varphi_2} & \cdots & \dfrac{\partial\psi_k}{\partial\varphi_k} \end{vmatrix} < 0$$

at all points of $U \cap D$.

We observe that if $(\varphi_1, \ldots, \varphi_k)$ and $(\theta_1, \ldots, \theta_k)$ are positively oriented coordinate systems for some oriented submanifold M defined over some subset V of M then

$$\begin{vmatrix} \dfrac{\partial\theta_1}{\partial\varphi_1} & \dfrac{\partial\theta_1}{\partial\varphi_2} & \cdots & \dfrac{\partial\theta_1}{\partial\varphi_k} \\ \dfrac{\partial\theta_2}{\partial\varphi_1} & \dfrac{\partial\theta_2}{\partial\varphi_2} & \cdots & \dfrac{\partial\theta_2}{\partial\varphi_k} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial\theta_k}{\partial\varphi_1} & \dfrac{\partial\theta_k}{\partial\varphi_2} & \cdots & \dfrac{\partial\theta_k}{\partial\varphi_k} \end{vmatrix} > 0$$

at all points of V.

Let M be a smooth orientable submanifold of $\mathbb{R}^n$ of dimension k with boundary ∂M. If we choose an orientation of M then this induces an orientation of the boundary ∂M of M. Suppose that $\mathbf{p}$ is a point of ∂M.

Let $(\varphi_1, \varphi_2, \ldots, \varphi_k)$ be a smooth coordinate system for the submanifold M, defined over some neighbourhood V of $\mathbf{p}$ in M which has the following properties:

(i) the coordinate system $(\varphi_1, \varphi_2, \ldots, \varphi_k)$ is positively oriented (with respect to the chosen orientation on the submanifold M).

(ii) $\varphi_1 \leq 0$ on V,

(iii) $\partial M \cap V = \{\mathbf{p} \in V : \varphi_1(\mathbf{p}) = 0\}$,

(It follows from the definition of a submanifold M of $\mathbb{R}^n$ with boundary ∂M that there exists such a coordinate system defined around any point of the boundary ∂M of M). Then the restriction of $(\varphi_2, \ldots, \varphi_k)$ to $\partial M \cap V$ defines a coordinate system for ∂M defined over $\partial M \cap V$. The *induced orientation* on ∂M is characterized by the property that the coordinate system $(\varphi_2, \ldots, \varphi_k)$ for ∂M is positively oriented for each coordinate system $(\varphi_1, \varphi_2, \ldots, \varphi_k)$ for M satisfying the three properties listed above.

We now describe how to integrate a continuous k-form over an oriented k-dimensional submanifold M of $\mathbb{R}^n$ (with or without boundary). The definition of the integral is completely analogous to the definition of the integral of a continuous 2-form over a smooth surface.

Let M be an oriented submanifold in $\mathbb{R}^n$ and let V be a subset of M. Suppose that there exists a positively oriented coordinate system $(\varphi_1, \ldots, \varphi_k)$ for the submanifold M defined over V. Let us define the subset W of $\mathbb{R}^k$ by

$$W = \{\mathbf{u} \in \mathbb{R}^k : \mathbf{u} = (\varphi_1(\mathbf{p}), \ldots, \varphi_k(\mathbf{p})) \text{ for some } \mathbf{p} \in V\}.$$

We let $\alpha\colon W \to V$ be the smooth map which expresses the standard Cartesian coordinates $(x_1, x_2, \ldots, x_n)$ at a point of V in terms of $\varphi_1, \ldots, \varphi_k$. Thus

$$x_i = \alpha_i(\varphi_1, \ldots, \varphi_k) \qquad (i = 1, 2, \ldots, n),$$

where $\alpha_i\colon W \to \mathbb{R}$ is the ith component of the map α. One can readily verify that the map $\alpha\colon W \to V$ has the following properties:

(i) the subset W of $\mathbb{R}^k$ is mapped homeomorphically onto V by the map α (i.e., $\alpha\colon W \to V$ is a bijection and both $\alpha\colon W \to V$ and $\alpha^{-1}\colon V \to W$ are continuous).

(ii) the Jacobian matrix of the map $\alpha\colon W \to \mathbb{R}^n$ has rank k at each point of W.

We refer to a map $\alpha: W \to V$ which arises from some coordinate system $(\varphi_1, \ldots, \varphi_k)$ for the submanifold M in the manner described above as a *local parameterization* of the submanifold.

Now let us suppose that the domain W of the local parameterization $\alpha: W \to V$ is Jordan-measurable. Then we define

$$\int_V \omega = \int_W \alpha^* \omega.$$

for each continuous k-form defined over some open set which contains V. We now show that the value of this integral is well-defined independently of the choice of the positively oriented smooth coordinate system $(\varphi_1, \ldots, \varphi_k)$ chosen on V.

Suppose that $(\varphi_1, \ldots, \varphi_k)$ and $(\psi_1, \ldots, \psi_k)$ are smooth positively oriented coordinate systems for the oriented submanifold M defined over the subset V of M. We define

$$\begin{aligned} W &= \{\mathbf{u} \in \mathbb{R}^k : \mathbf{u} = (\varphi_1(\mathbf{p}), \ldots, \varphi_k(\mathbf{p})) \text{ for some } \mathbf{p} \in V\}, \\ \tilde{W} &= \{\mathbf{u} \in \mathbb{R}^k : \mathbf{u} = (\psi_1(\mathbf{p}), \ldots, \psi_k(\mathbf{p})) \text{ for some } \mathbf{p} \in V\}. \end{aligned}$$

We let $\alpha: W \to V$ and $\beta: \tilde{W} \to V$ be the local parameterizations of the submanifold defined such that

$$x_i = \alpha_i(\varphi_1, \ldots, \varphi_k) = \beta_i(\psi_1, \ldots, \psi_k) \qquad (i = 1, 2, \ldots, n),$$

where α_i and β_i denote the ith components of the maps α and β respectively. There exists a diffeomorphism $\rho: W \to \tilde{W}$ characterized by the property that

$$(\psi_1(\mathbf{p}), \ldots, \psi_k(\mathbf{p})) = \rho(\varphi_1(\mathbf{p}, \ldots, \varphi_k(\mathbf{p})$$

for all $\mathbf{p} \in V$. This diffeomorphism $\rho: W \to \tilde{W}$ is orientation preserving (since the positively oriented coordinate systems $(\varphi_1, \ldots, \varphi_k)$ and $(\psi_1, \ldots, \psi_k)$ determine the same orientation on V). It follows from the definitions of α, β and ρ that $\alpha = \beta \circ \rho$. Therefore

$$\int_W \alpha^* \omega = \int_W \rho^*(\beta^* \omega) = \int_{\tilde{W}} \beta^* \omega$$

by Theorem 13.2. This shows that the value of the integral $\int_V \omega$ is well-defined independently of the choice of the positively oriented smooth coordinate system $(\varphi_1, \ldots, \varphi_k)$ chosen on V, as required.

The discussion above explains how to calculate the integral of a k-form over a portion V of some oriented smooth submanifold M in $\mathbb{R}^n$, provided that the region V is contained within the domain of some smooth coordinate system for the submanifold M and provided that V corresponds by means of this coordinate system to a closed bounded Jordan-measurable subset of $\mathbb{R}^k$.

If M is a compact oriented smooth submanifold of $\mathbb{R}^k$ (with or without boundary) then it is possible to partition M into regions $V_1, V_2, \ldots, V_k$, where each region V_i is a closed bounded set which belongs to the domain of some smooth coordinate system for the submanifold M and which corresponds by means of that coordinate system to a closed bounded Jordan-measurable set in $\mathbb{R}^k$. Let ω be a continuous k-form defined on M. The integral of the k-form ω over each of the regions $V_1, V_2, \ldots, V_r$ can be defined as described above. We define

$$\int_V \omega = \sum_{j=1}^{r} \int_{V_j} \omega.$$

It is not difficult to show that the value of this integral is independent of the fashion in which we divide up the region V into the subregions $V_1, V_2, \ldots, V_r$.

14.2 The Generalized Stokes' Theorem

We now state a generalization of Stokes' Theorem which applies to integrals of smooth differential forms over smooth submanifolds of $\mathbb{R}^n$.

We recall that if M as a smooth submanifold of $\mathbb{R}^n$ with boundary ∂M then any orientation of M induces a corresponding orientation of the boundary ∂M of M. This orientation on ∂M is known as the *induced orientation* on ∂M.

Theorem 14.1 (The Generalized Stokes' Theorem) *Let M be a compact smooth oriented k-dimensional submanifold of $\mathbb{R}^n$ with boundary ∂M. Then*

$$\int_M d\omega = \int_{\partial M} \omega$$

for every smooth $(k-1)$-form ω on M, where the integral over ∂M is taken with respect to the induced orientation on ∂M.

We shall present a proof of this theorem below. First however we show that a number of well-known theorems of vector calculus can be deduced as corollaries of this result.

The first of these Corollaries is *Green's Theorem in the Plane*. Let D be a region on the plane $\mathbb{R}^2$ which is bounded by smooth closed curves $\gamma_1, \gamma_2, \ldots, \gamma_r$. Then there is an induced orientation (i.e., direction of travel) on these boundary curves characterized by the property that the region D lies to ones left as one traverses each of these boundary curves.

Theorem 14.2 (Green's Theorem in the Plane) *Let D be a closed bounded region on the plane $\mathbb{R}^2$ which is bounded by smooth curves $\gamma_1, \gamma_2, \ldots, \gamma_r$. Let P and Q be smooth functions on D. Then*

$$\sum_{j=1}^{r} \int_{\gamma_j} (P\,dx + Q\,dy) = \int_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx\,dy,$$

where the orientation of the closed curves $\gamma_1, \gamma_2, \ldots, \gamma_r$ is the induced orientation described above.

Proof This result follows from Theorem 14.1 in the particular case when $M = D$, $\partial M = \gamma_1 \cup \cdots \cup \gamma_r$, and $\omega = P\,dx + Q\,dy$, so that

$$d\omega = \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx \wedge dy. \quad \blacksquare$$

The next corollary is *Gauss' Theorem* (also known as the *Divergence Theorem*).

Theorem 14.3 (Gauss' Theorem) *Let V be a closed bounded region in $\mathbb{R}^3$ which is bounded by some smooth surface Σ. Let $\mathbf{n}$ be the unit normal vector field on Σ directed outwards (i.e., directed away from the region V). Let B be a smooth vector field on V. Then*

$$\int_V \operatorname{div} \mathbf{B}\,dx\,dy\,dz = \int_\Sigma \mathbf{B}.\mathbf{n}\,dS.$$

Proof Let B_1, B_2 and B_3 be the Cartesian components of the vector field $\mathbf{B}$. Let ω be the smooth vector field on V defined by

$$\omega = B_1\,dy \wedge dz + B_2\,dz \wedge dx + B_3\,dx \wedge dy$$

(where (x, y, z) are the standard Cartesian coordinates on $\mathbb{R}^3$. Then

$$\int_\Sigma \mathbf{B}.\mathbf{n}\,dS = \int_\Sigma \omega.$$

by Lemma 13.8, where the orientation on Σ is the orientation induced on the boundary Σ of the region V by the standard orientation of V. Also

$$d\omega = \operatorname{div} \mathbf{B}\,dx \wedge dy \wedge dz,$$

hence

$$\int_V d\omega = \int_V \operatorname{div} \mathbf{B}\,dx\,dy\,dz.$$

We conclude from Theorem 14.1 that

$$\int_V \operatorname{div} \mathbf{B}\,dx\,dy\,dz = \int_\Sigma \mathbf{B}.\mathbf{n}\,dS. \quad \blacksquare$$

We now show how the classical form of Stokes' Theorem can be deduced from Theorem 14.1. Let Z be an oriented smooth surface in $\mathbb{R}^3$ bounded by smooth closed curves $\gamma_1, \gamma_2, \ldots, \gamma_r$. The orientation on the surface Z determines a unit normal vector field $\mathbf{n}$ on Z (see Section 13). The orientation on Z induces an orientation on the boundary ∂Z of Z. Moreover

$$\int_{\partial Z} \eta = \sum_{j=1}^{r} \int_{\gamma_j} \eta,$$

for all continuous 1-forms η on the boundary of Z, where each of the curves $\gamma_1, \gamma_2 \ldots, \gamma_r$ is traversed in the direction determined by the induced orientation on the boundary of Z. This direction may be described as follows: if $\mathbf{p}$ is a point which lies on the curve γ_j, if $\mathbf{T}$ is a vector tangential to the curve γ_j at $\mathbf{p}$ in the direction in which this curve is traversed, if $\mathbf{U}$ is a vector directed into the surface Z at $\mathbf{p}$, and if $\mathbf{n}$ is the unit normal vector to the surface Z determined by the orientation on Z, then $(\mathbf{T}, \mathbf{U}, \mathbf{n})$ is a positively oriented basis of $\mathbb{R}^3$. Let B be a vector field on Z whose Cartesian components are (B_1, B_2, B_3). We denote the line integral

$$\int_{\gamma_j} B_1\, dx + B_2\, dy + B_3\, dz$$

by $\oint_{\gamma_j} \mathbf{B}.d\mathbf{r}$.

Theorem 14.4 (Stokes' Theorem) *Let Z be a smooth oriented surface in $\mathbb{R}^3$ bounded by smooth curves $\gamma_1, \gamma_2, \ldots, \gamma_r$, and let $\mathbf{n}$ be the normal vector field on Z determined by the orientation of Z. Then*

$$\int_Z (\operatorname{curl} \mathbf{B}).\mathbf{n}\, dS = \sum_{j=1}^{r} \oint_{\gamma_j} \mathbf{B}.d\mathbf{r}$$

for all smooth vector fields $\mathbf{B}$ on the surface Z, where the boundary curves $\gamma_1, \gamma_2 \ldots, \gamma_r$ are traversed in the direction determined by the orientation induced on the boundary of Z by the orientation of Z.

Proof Consider the smooth 1-form η on Z defined by

$$\eta = B_1\, dx + B_2\, dy + B_3\, dz,$$

where (B_1, B_2, B_3) are the Cartesian components of the vector field $\mathbf{B}$. Then

$$\begin{aligned} d\eta &= \left(\frac{\partial B_2}{\partial z} - \frac{\partial B_3}{\partial y} \right) dy \wedge dz + \left(\frac{\partial B_3}{\partial x} - \frac{\partial B_1}{\partial z} \right) dz \wedge dx \\ &\quad + \left(\frac{\partial B_1}{\partial y} - \frac{\partial B_2}{\partial x} \right) dx \wedge dy. \end{aligned}$$

It follows from Lemma 13.8 that

$$\int_Z d\eta = \int_Z (\mathrm{curl}\,\mathbf{B}).\mathbf{n}\,dS,$$

where

$$\mathrm{curl}\,\mathbf{B} = \left(\frac{\partial B_2}{\partial z} - \frac{\partial B_3}{\partial y}, \frac{\partial B_3}{\partial x} - \frac{\partial B_1}{\partial z}, \frac{\partial B_1}{\partial y} - \frac{\partial B_2}{\partial x} \right).$$

Thus

$$\begin{aligned} \int_Z (\mathrm{curl}\,\mathbf{B}).\mathbf{n}\,dS &= \int_Z d\eta = \int_{\partial Z} \eta \\ &= \sum_{j=1}^{r} \oint_{\gamma_j} \mathbf{B}.d\mathbf{r}, \end{aligned}$$

by Theorem 14.1. ∎

14.3 The Proof of the Generalized Stokes' Theorem

*The material in the remainder of this section of the course is **NON-EXAMINABLE**. We give a proof of the Generalized Stokes' Theorem below, after first describing a number of concepts and results that we use in the proof of this theorem.*

Definition Let $f\colon D \to \mathbb{R}$ be a continuous function defined over some subset D of $\mathbb{R}^n$. The *support* of the function f is defined to be the closure (in $\mathbb{R}^n$) of the set

$$\{\mathbf{x} \in D : f(\mathbf{x}) \neq 0\}.$$

We see from this definition that the support $\mathrm{supp}\, f$ of the function f is the smallest closed set with the property that the function f vanishes throughout the complement of this set. We can generalize this definition to differential forms: if ω is a continuous differential form defined over some subset D of $\mathbb{R}^n$ then the *support* $\mathrm{supp}\,\omega$ of the differential form ω is the smallest closed set in $\mathbb{R}^n$ with the property that ω vanishes throughout the complement of this set.

Definition Let ω be a continuous differential form on $\mathbb{R}^n$. Then ω is said to have *bounded support* if and only if there exists some $R \geq 0$ with the property that ω vanishes throughout the set

$$\{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x}| > R\}.$$

We see from these definitions that a differential form on $\mathbb{R}^n$ has bounded support if and only if the support of the differential form is a bounded set in $\mathbb{R}^n$. If η is a continuous n-form on $\mathbb{R}^n$ then the integral $\int_{\mathbb{R}^n} \eta$ is well-defined, since η vanishes outside some bounded set. Similarly the integral $\int_H \eta$ is well-defined, where H is any half-space in $\mathbb{R}^n$.

Lemma 14.5 *Let H be the half-space in $\mathbb{R}^n$ defined by*

$$H = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1 \leq 0\},$$

and let ∂H be the boundary of H, given by

$$\partial H = \{(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n : x_1 = 0\},$$

Let ∂H be oriented so that $(x_2, x_3, \ldots, x_n)$ is a positively oriented coordinate system on ∂H. Let ω be a continuously differentiable $(n-1)$-form with bounded support on H. Then

$$\int_H d\omega = \int_{\partial H} \omega.$$

Proof Let $\sigma_1, \sigma_2, \ldots, \sigma_n$ be the $n-1$-forms on H defined by

$$\begin{aligned}
\sigma_1 &= dx_2 \wedge dx_3 \wedge \cdots \wedge dx_n \\
\sigma_2 &= -dx_1 \wedge dx_3 \wedge \cdots \wedge dx_n \\
&\vdots \\
\sigma_n &= (-1)^{n-1} dx_1 \wedge dx_3 \wedge \cdots \wedge dx_n.
\end{aligned}$$

(Thus $\sigma_i = (-1)^{i-1} dx_1 \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots \wedge dx_n$ for $i = 1, 2, \ldots, n$.) Note that $dx_j \wedge \sigma_i = 0$ if $j \neq i$ and that

$$dx_i \wedge \sigma_i = dx_1 \wedge dx_2 \wedge \cdots \wedge dx_n$$

for $i = 1, 2, \ldots, n$. If ω is a continuously differentiable $(n-1)$-form on H then we can write

$$\omega = \sum_{i=1}^{n} f_i \sigma_i,$$

where $f_1, f_2, \ldots, f_n$ are continuously differentiable functions on H. But then

$$d\omega = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial f_i}{dx_j} dx_j \wedge \sigma_i = \sum_{i=1}^{n} \frac{\partial f_i}{dx_i} dx_i \wedge dx_2 \wedge \cdots \wedge dx_n.$$

Therefore

$$\int_H d\omega = \int_H \left(\sum_{i=1}^n \frac{\partial f_i}{dx_i} \right) dx_1\, dx_2 \ldots, dx_n.$$

However there exists some $R > 0$ with the property that $f(\mathbf{x}) = 0$ whenever $\mathbf{x} \in H$ satisfies $|\mathbf{x}| > R$. It follows from this that

$$\int_H \frac{\partial f_i}{dx_i}\, dx_1\, dx_2 \ldots, dx_n = 0$$

if $i \neq 1$ and that

$$\begin{aligned} \int_H \frac{\partial f_1}{dx_1}\, dx_1\, dx_2 \ldots, dx_n &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\int_{-\infty}^{0} \frac{\partial f_i}{dx_1}\, dx_1 \right) dx_2\, dx_3 \ldots,\ dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_1(0, x_2, \ldots, x_n)\, dx_2\, dx_3 \ldots,\ dx_n \\ &= \int_{\partial H} f_1 \sigma_1. \end{aligned}$$

Moreover

$$\int_{\partial H} \omega = \int_{\partial H} f_1 \sigma_1,$$

since $i^* \sigma_i = 0$ on ∂H whenever $i \neq 1$, where $i\colon \partial H \hookrightarrow H$ is the inclusion map. If we combine these results we see that

$$\int_H d\omega = \int_{\partial H} \omega,$$

as required. ∎

In the proof of Theorem 14.1 we shall make use of the following result, which we state without proof.

Theorem 14.6 *Let M be a compact smooth submanifold of $\mathbb{R}^n$ (with or without boundary). Then there exist smooth real-valued functions $f_1, f_2, \ldots, f_r$ defined over $\mathbb{R}^n$ such that*

$$f_1 + f_2 + \cdots + f_r = 1$$

and such that the support of each function f_j is contained within the domain of some smooth curvilinear coordinate system adapted to the submanifold M.

We can now prove the Generalized Stokes' Theorem for differential forms.

Proof of Theorem 14.1. In order to prove Stokes' Theorem we consider first the case when the support of the $(k-1)$-form ω is contained within the domain of some smooth curvilinear coordinate system for $\mathbb{R}^n$ which is adapted to the submanifold M. Thus let us suppose that there exists some open set U in $\mathbb{R}^n$ and a smooth curvilinear coordinate system

$$(\psi_1, \psi_2, \ldots \psi_n)$$

on U such that the following conditions are satisfied:

(i) $M \cap U = \{\mathbf{x} \in U : \psi_1(\mathbf{x}) \leq 0 \text{ and } \psi_i(\mathbf{x}) = 0 \text{ for } i = k+1, \ldots, n\}$ and $\partial M \cap U = \{\mathbf{x} \in U : \psi_1(\mathbf{x}) = 0 \text{ and } \psi_i(\mathbf{x}) = 0 \text{ for } i = k+1, \ldots, n\}$.

(ii) $(\psi_1, \psi_2, \ldots, \psi_k)$ defines a positively-oriented coordinate system on $M \cap U$,

(iii) the support $\operatorname{supp} \omega$ of the $(k-1)$-form ω is contained in U.

It follows from condition (ii), together with the definition of the induced orientation on ∂M, that the restriction of $(\psi_2, \ldots, \psi_k)$ to $\partial M \cap U$ defines a positively oriented coordinate system on $\partial M \cap U$.

Let $\alpha \colon W \to M$ be the local parameterization of the submanifold M determined by the coordinate system $(\psi_1, \psi_2, \ldots, \psi_k)$, where

$$W = \left\{\mathbf{u} \in \mathbb{R}^k : \mathbf{u} = (\psi_1(\mathbf{p}), \psi_2(\mathbf{p}), \ldots, \psi_k(\mathbf{p})) \text{ for some } \mathbf{p} \in M \cap U\right\}.$$

The local parameterization α expresses the standard Cartesian coordinates $(x_1, x_2, \ldots, x_n)$ in terms of $(\psi_1, \psi_2, \ldots, \psi_k)$. Thus

$$x_i = \alpha_i(\psi_1, \psi_2, \ldots, \psi_k) \qquad (i = 1, 2, \ldots, n)$$

on $M \cap U$ (where α_i denotes the ith component of the map α). Let H be the half-space defined by $H = \{\mathbf{u} \in \mathbb{R}^k : u_1 \leq 0\}$. Let $(u_1, u_2, \ldots, u_k)$ denote the standard Cartesian coordinate system on H. Let σ be the $(k-1)$-form on H defined such that $\sigma = \alpha^*\omega$ on W and $\sigma = 0$ on $H \setminus W$. Then σ is a smooth $(k-1)$-form on H. (This follows from the fact that ω is a smooth $(k-1)$-form whose support is contained in U.) Therefore

$$\int_H d\sigma = \int_{\partial H} \sigma$$

by Lemma 14.5, where ∂H is the $k-1$ dimensional subspace of $\mathbb{R}^k$ defined by $\partial H = \{\mathbf{u} \in \mathbb{R}^k : u_1 = 0\}$, and where ∂H is oriented so that $(u_2, u_3, \ldots, u_k)$

is a positively oriented coordinate system on ∂H. But then

$$\begin{aligned}\int_M d\omega &= \int_W \alpha^*(d\omega) = \int_W d(\alpha^*\omega) = \int_W d\sigma \\ &= \int_H d\sigma = \int_{\partial H} \sigma = \int_{\partial H \cap W} \alpha^*\omega \\ &= \int_{\partial M} \omega,\end{aligned}$$

where we have used the definition of $\int_M d\omega$ and $\int_{\partial M} \omega$ together with the fact that $\alpha^*(d\omega) = d(\alpha^*\omega)$ (see Lemma 9.7). This proves the theorem in the special case when the support of the $(k-1)$-form ω is contained within the domain of some coordinate system on $\mathbb{R}^n$ which is adapted to the submanifold M.

We now prove the result in the general case. It follows from Theorem 14.6 that there exist smooth functions $f_1, f_2, \ldots, f_r$ defined over $\mathbb{R}^n$ such that

$$f_1 + f_2 + \cdots + f_r = 1$$

and such that the support of each function f_j is contained within the domain of some smooth curvilinear coordinate system adapted to the submanifold M. It then follows from the special case proved above that

$$\int_M d(f_j\omega) = \int_{\partial M} f_j\omega \qquad (j = 1, 2, \ldots, r).$$

Therefore

$$\int_M d\omega = \sum_{j=1}^r \int_M d(f_j\omega) = \sum_{j=1}^r \int_{\partial M} f_j\omega = \int_{\partial M} \omega,$$

as required. ∎