MATHEMATICS 218: NOTES

Robert C. Gunning

January 27, 2010
Introduction

These are notes of honors courses on calculus of several variables given at Princeton University during the academic years 2007-2010. I would particularly like to thank Lillian Pierce, who taught the course with me during the fall of 2008 and suggested and largely was responsible for supplementing the theoretical material with problem sets emphasizing the explicit calculations that are essential to understanding and using this material. The basis for these notes is the set of notes of the fall 2008 course taken by Robert Haraaway, Adam Hesterberg, Jay Holt and Alex Schiller, whom I should also like to thank for their careful and thorough notes and their many suggestions.

Robert C. Gunning
Fine Hall
Princeton University
January 2010
# Contents

Introduction iii

1 Background 1
   1.1 Norms ........................................ 1
   1.2 Topological Preliminaries ...................... 6
   1.3 Continuous Mappings .......................... 10

2 Differentiable Mappings 13
   2.1 The Derivative .................................. 13
   2.2 The Chain Rule .................................. 16
   2.3 Higher Derivatives ............................. 19
   2.4 Functions ....................................... 22

3 The Rank Theorem 29
   3.1 The Inverse Mapping Theorem ..................... 29
   3.2 The Implicit Function Theorem ................... 36
   3.3 The Rank Theorem ................................ 43

4 Integration 53
   4.1 Riemann Integral ............................... 53
   4.2 Fubini’s Theorem ................................ 63
   4.3 Limits and Improper Integrals ................... 67
   4.4 Change of Variables ............................. 72

5 Differential Forms 77
   5.1 Line Integrals ................................... 77
   5.2 Differential Forms ............................... 83
   5.3 Stokes’s Theorem ................................. 93

A Exterior Algebra 107
Chapter 1

Background

1.1 Norms

Some fairly standard set-theoretic notation and terminology will be used consistently. In particular, $a \in A$ indicates that $a$ is an element of, or a point in, the set $A$, while $A \subset B$ indicates that $A$ is a subset of $B$, possibly equal to $B$. The union $A \cup B$ of sets $A$ and $B$ consists of all elements that belong to either $A$ or $B$ or both $A$ and $B$, while the intersection $A \cap B$ consists of all elements that belong to both $A$ and $B$. The complement $A \sim B$ consists of all elements of $A$ that do not belong to $B$, whether or not $B$ is a subset of $A$. If the set $A$ is understood from context it may be omitted; so $\sim B$ consists of all elements in a set $A$ that are not in $B$, where the set $A$ is understood, and if $A$ and $B$ are both understood to be subsets of a set $E$ then $A \sim B = A \cap (\sim B)$. A mapping $f : A \rightarrow B$ associates to each element $a \in A$ an element $f(a) \in B$, the image of the point $a$. This mapping $f$ is injective if distinct elements of $A$ have distinct images in $B$, is surjective if every element of $B$ is the image of some element of $A$, and is bijective if it is both injective and surjective; thus $f$ is bijective if and only if is a one-to-one mapping between the sets $A$ and $B$, and consequently has an inverse mapping $f^{-1} : B \rightarrow A$ that is also bijective.

The $n$-dimensional real vector space will be denoted by $\mathbb{R}^n$, following the Bourbaki convention. It will be viewed as the space of real column vectors of length $n$, elements of which are

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix};$$

but for notational convenient a vector $x \in \mathbb{R}^n$ sometimes will be indicated by listing its coordinates in the form $x = \{x_j\} = \{x_1, \ldots, x_n\}$. The origin in $\mathbb{R}^n$ is the vector $0 = \{0\} = \{0, 0, \ldots, 0\}$. When considering a function $f$, the notation $f(x_1, \ldots, x_n)$ also will be used when viewing coordinates of the vector $x$ as individual variables; but the vector $x$ still will be viewed as a column vector. Addition is the usual addition of vectors by adding their coordinates; and scalar multiplication amounts to multiplying
all the entries of the vector by a scalar, so that

\[ a \mathbf{x} = \begin{pmatrix} ax_1 \\ \vdots \\ ax_n \end{pmatrix} \quad \text{for any } a \in \mathbb{R}. \]

Linear mappings between vector spaces are described by matrix multiplication; for example, a 2 × 3 matrix \( A = \{a_{ij}\} \) describes the mapping \( A : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \) that takes a vector \( \mathbf{x} = \{x_j\} \in \mathbb{R}^3 \) to the vector

\[ A\mathbf{x} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{pmatrix} \in \mathbb{R}^2. \]

There are a various norms measuring the sizes or lengths of vectors in \( \mathbb{R}^n \), only two of which will be considered here:

- the Euclidean norm or Cartesian norm of a vector \( \mathbf{x} = \{x_j\} \) is defined by
  \[ \|\mathbf{x}\|_2 = \sqrt{\sum_{j=1}^{n} x_j^2} = \sqrt{x_1^2 + \cdots + x_n^2} \quad \text{(with the non-negative square root)}; \]

- the supremum norm or sup norm of a vector \( \mathbf{x} = \{x_j\} \) is defined by
  \[ \|\mathbf{x}\|_\infty = \max_{1 \leq j \leq n} |x_j|. \]

In general, a norm on the vector space \( \mathbb{R}^n \) is a mapping \( \mathbb{R}^n \rightarrow \mathbb{R} \) that associates to any \( \mathbf{x} \in \mathbb{R}^n \) a real number \( \|\mathbf{x}\| \in \mathbb{R} \) with the following properties:

1. positivity: \( \|\mathbf{x}\| \geq 0 \) and \( \|\mathbf{x}\| = 0 \) if and only if \( \mathbf{x} = 0 \);
2. homogeneity: \( \|c\mathbf{x}\| = |c| \|\mathbf{x}\| \) for any \( c \in \mathbb{R} \);
3. the triangle inequality: \( \|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \).

That the supremum norm satisfies these three properties is obvious, except perhaps for the triangle inequality; to verify that, if \( \mathbf{x}, \mathbf{y} \in \mathbb{R}^n \) then since \( |x_j| \leq \|\mathbf{x}\|_\infty \) and \( |y_j| \leq \|\mathbf{y}\|_\infty \) for \( 1 \leq j \leq n \) it follows that \( |x_j + y_j| \leq |x_j| + |y_j| \leq \|\mathbf{x}\|_\infty + \|\mathbf{y}\|_\infty \) for \( 1 \leq j \leq n \) and consequently that \( \|\mathbf{x} + \mathbf{y}\| = \max_{1 \leq j \leq n} |x_j + y_j| \leq \|\mathbf{x}\|_\infty + \|\mathbf{y}\|_\infty \).

That the Euclidean norm satisfies these three properties also is obvious, except for the triangle inequality; it is convenient to demonstrate that together with an inequality involving the inner product of two vectors, which is defined by

\[ (\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{n} x_j y_j \quad \text{for vectors } \mathbf{x} = \{x_j\} \text{ and } \mathbf{y} = \{y_j\} \in \mathbb{R}^n. \]

The inner product is quite commonly written \( (\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y} \) and called the dot product of the two vectors \( \mathbf{x} \) and \( \mathbf{y} \). It is characterized by the following properties:

1. linearity: \( (c_1 \mathbf{x}_1 + c_2 \mathbf{x}_2, \mathbf{y}) = c_1 (\mathbf{x}_1, \mathbf{y}) + c_2 (\mathbf{x}_2, \mathbf{y}) \) for any \( c \in \mathbb{R} \); 
2. symmetry: \( (\mathbf{x}, \mathbf{y}) = (\mathbf{y}, \mathbf{x}) \);
3. positivity: \( (\mathbf{x}, \mathbf{x}) \geq 0 \) and \( (\mathbf{x}, \mathbf{x}) = 0 \) if and only if \( \mathbf{x} = 0 \).

These three properties follow almost immediately from the defining equation (1.1). It is apparent from symmetry that the inner product \( (\mathbf{x}, \mathbf{y}) \) is also a linear function of the vector \( \mathbf{y} \). The Euclidean norm can be defined in terms of the inner product by

\[ \|\mathbf{x}\|_2 = \sqrt{(\mathbf{x}, \mathbf{x})} \quad \text{(with the non-negative square root)}, \]

(1.2)
1.1. NORMS

as is quite clear from the definitions. Conversely the inner product can be defined in terms of the norm by

\[(1.3) \quad (x, y) = \frac{1}{4} \|x + y\|_2^2 - \frac{1}{4} \|x - y\|_2^2 \]

since

\[\|x + y\|_2^2 - \|x - y\|_2^2 = (x + y, x + y) - (x - y, x - y)\]
\[= (x, x) + 2(x, y) + (y, y) - ((x, x) - 2(x, y) + (y, y))\]
\[= 4(x, y).\]

Equation (1.3) is called the polarization identity.

**Theorem 1.1**

(i) \(|(x, y)| \leq \|x\|_2 \|y\|_2\) for any \(x, y \in \mathbb{R}^n\); and this is an equality if and only if the two vectors are linearly dependent.

(ii) \(\|x + y\|_2 \leq \|x\|_2 + \|y\|_2\) for any \(x, y \in \mathbb{R}^n\); and this is an equality if and only if one of the two vectors is a non-negative multiple of the other.

**Proof:** If \(x, y \in \mathbb{R}^n\) where \(y \neq 0\) and \(t \in \mathbb{R}\) introduce the continuous function \(f(t)\) of the variable \(t\) defined by

\[(1.4) \quad f(t) = \|x + ty\|_2^2 = \sum_{j=1}^{n} (x_j + ty_j)^2\]
\[= \sum_{j=1}^{n} x_j^2 + 2t \sum_{j=1}^{n} x_jy_j + t^2 \sum_{j=1}^{n} y_j^2\]
\[= \|x\|_2^2 + 2t(x, y) + t^2 \|y\|_2^2.\]

It is clear that the function \(f(t)\) becomes large for \(|t|\) large, since \(\|y\|_2 > 0\) by assumption, so it must take a minimum value at some point. Since

\[f'(t) = 2(x, y) + 2t \|y\|_2^2 \]

there is a single point at which \(f'(t) = 0\), the point \(t_0 = -\frac{(x, y)}{\|y\|_2^2}\), so this must be the point at which the function \(f(t)\) takes its minimum value; and the minimum value is

\[(1.5) \quad f(t_0) = \|x\|_2^2 + 2 \left(\frac{(x, y)}{\|y\|_2^2}\right)(x, y) + \frac{(x, y)^2}{\|y\|_2^2} \|y\|_2^2 = \frac{\|x\|_2^2 \|y\|_2^2 - (x, y)^2}{\|y\|_2^2}.\]

It is clear from (1.4) that \(f(t) \geq 0\) at all points \(t \in \mathbb{R}\), so in particular at the point \(t_0\), and that yields the inequality (i). It is clear from (1.5) that this inequality is an equality if and only if \(f(t_0) = 0\), hence if and only if \(x + t_0y = 0\) since \(f(t_0) = \|x + t_0y\|_2^2\); and since \(y \neq 0\) that is just the condition that the vectors \(x\) and \(y\) are linearly dependent. Next from (1.4) with \(t = 1\) and from the inequality (i) it follows that

\[\|x + y\|_2^2 = \|x\|_2^2 + 2(x, y) + \|y\|_2^2 \leq \|x\|_2^2 + 2\|x\|_2 \|y\|_2 + \|y\|_2^2 = (\|x\|_2 + \|y\|_2)^2,\]

which yields the inequality (ii). This inequality is an equality if and only if \((x, y) = \|x\|_2 \|y\|_2\), or equivalently if and only if (i) is an equality and \((x, y) \geq 0\); and since
y ≠ 0 it follows from what has already been demonstrated that the inequality (i) is an equality if and only if x = cy for some real number c, and then (x, y) = c|y|^2 ≥ 0 if and only if c ≥ 0. That suffices for the proof.

The very useful inequality (i) in the preceding theorem, called the **Cauchy-Schwarz inequality**, can be written

\[
\frac{(x, y)}{\|x\| \|y\|} \leq 1 \text{ if } x \neq 0, y \neq 0;
\]

as a consequence there is an angle \( \theta \), called the **angle** between the nonzero vectors \( x \) and \( y \), that is determined uniquely up to a multiple of \( 2\pi \) by

\[
(1.6) \quad \cos \theta = \frac{(x, y)}{\|x\| \|y\|}.
\]

In particular if \( \theta = 0 \) or \( \pi \) so \( \cos \theta = \pm 1 \) then by Theorem 1.1 (i) the two vectors \( x \) and \( y \) are linearly dependent; they are parallel and in the same direction if \( \theta = 0 \) so that \( (x, y) > 0 \), or parallel and in the opposite direction if \( \theta = \pi \) so that \( (x, y) < 0 \). The two vectors are **orthogonal** or perpendicular to one another when the angle is either \( \pi/2 \) or \( 3\pi/2 \) so that \( (x, y) = 0 \). The geometrical interpretation of the norm \( \|x\|_2 \) in terms of the inner product makes that norm particularly useful in many applications.

In general, two norms \( \|x\|_a \) and \( \|x\|_b \) on a vector space \( \mathbb{R}^n \) are **equivalent** if there are nonzero constants \( c_a \) and \( c_b \) such that \( \|x\|_a \leq c_a \|x\|_b \) and \( \|x\|_b \leq c_b \|x\|_a \) for all \( x \in \mathbb{R}^n \). The Euclidean and supremum norms are equivalent, since they are related by the very useful inequalities

\[
(1.7) \quad \|x\|_\infty \leq \|x\|_2 \leq \sqrt{n} \|x\|_\infty \text{ for any } x \in \mathbb{R}^n.
\]

To verify these inequalities, if \( x = \{x_j\} \in \mathbb{R}^n \) then \( |x_j| \leq \|x\|_\infty \) for \( 1 \leq j \leq n \) so

\[
\|x\|_2^2 = \sum_{j=1}^{n} x_j^2 \leq \sum_{j=1}^{n} |x_j|^2 = n \|x\|_\infty^2 \text{ hence } \|x\|_2 \leq \sqrt{n} \|x\|_\infty;
\]

on the other hand

\[
\|x\|_2^2 = \sum_{j=1}^{n} x_j^2 \geq x_j^2 \text{ so } \|x\|_2 \geq |x_j| \text{ for } 1 \leq j \leq n \text{ hence } \|x\|_2 \geq \|x\|_\infty.
\]

It follows from (1.7) that whenever \( \|x\|_2 \) is small then \( \|x\|_\infty \) is also small, and conversely; so for many purposes the two norms are interchangeable. If it is not really necessary to specify which norm is meant the notation \( \|x\| \) will be used, meaning either \( \|x\|_2 \) or \( \|x\|_\infty \). Of course some care must be taken; in particular \( \|x\| \) should have the same meaning in any single equation, or usually throughout any single proof, for the quite obvious reasons.

For some purposes it is convenient to view an \( m \times n \) matrix

\[
A = \left\{ a_{ij} \mid 1 \leq i \leq m, 1 \leq j \leq n \right\} \in \mathbb{R}^{m \times n}
\]

as a vector in \( \mathbb{R}^{mn} \), and so to consider its norm when viewed as a vector; thus

\[
(1.8) \quad \|A\|_\infty = \max_{1 \leq i \leq m} \min_{1 \leq j \leq n} |a_{ij}| \text{ and } \|A\|_2 = \sqrt{\sum_{1 \leq i \leq m, 1 \leq j \leq n} a_{ij}^2} \quad \text{(the nonnegative square root)}.
\]
1.1. NORMS

If \( x \in \mathbb{R}^n \) then \( Ax \in \mathbb{R}^m \), and since

\[
\left| \sum_{j=1}^{n} a_{ij} x_j \right| \leq \sum_{j=1}^{n} |a_{ij}| x_j \leq \sum_{j=1}^{n} \|A\|_\infty \|x\|_\infty = n \|A\|_\infty \|x\|_\infty
\]

it follows that

\[
\|Ax\|_\infty = \max_{1 \leq i \leq m} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \leq n \|A\|_\infty \|x\|_\infty.
\]

An \( \epsilon \)-neighborhood \( \mathcal{N}_\epsilon(a) \) of a point \( a \in \mathbb{R}^n \), an open \( \epsilon \)-neighborhood to be more specific, is a subset of \( \mathbb{R}^n \) defined by

\[
\mathcal{N}_\epsilon(a) = \left\{ x \in \mathbb{R}^n \mid d(x,a) = \|x-a\| < \epsilon \right\};
\]

a closed \( \epsilon \)-neighborhood is defined correspondingly by

\[
\overline{\mathcal{N}}_\epsilon(a) = \left\{ x \in \mathbb{R}^n \mid d(x,a) = \|x-a\| \leq \epsilon \right\}.
\]

The boundary of an open or closed \( \epsilon \)-neighborhood is the closed set

\[
\partial \mathcal{N}_\epsilon(a) = \partial \overline{\mathcal{N}}_\epsilon(a) = \left\{ x \in \mathbb{R}^n \mid d(x,a) = \|x-a\| = \epsilon \right\}.
\]

If it is necessary or convenient to be specific, an open \( \epsilon \) neighborhood in the Euclidean norm is denoted by \( \mathcal{N}_{\epsilon,2}(a) \) while an open \( \epsilon \)-neighborhood in the supremum norm is denoted by \( \mathcal{N}_{\epsilon,\infty}(a) \). It is clear from the inequality (1.7) that

\[
\mathcal{N}_{\epsilon,2}(a) \subset \mathcal{N}_{\epsilon,\infty}(a) \subset \mathcal{N}_{\epsilon,\sqrt{n},2}(a) \quad \text{in} \quad \mathbb{R}^n.
\]

The geometrical shapes of these neighborhoods depend on the norms used in their definitions; for example in the plane \( \mathbb{R}^2 \) neighborhoods have the shapes as shown in Figure 1.1.

Figure 1.1: Neighborhoods \( \mathcal{N}_\epsilon(a) \subset \mathbb{R}^2 \).

Other useful subsets of \( \mathbb{R}^n \) are open cells, subsets of the form

\[
\Delta = \left\{ x = \{x_j\} \in \mathbb{R}^n \mid a_j < x_j < b_j \quad \text{for} \quad 1 \leq j \leq n \right\}
\]

for arbitrary real numbers \( a_j < b_j \), and closed cells, subsets of the form

\[
\overline{\Delta} = \left\{ x = \{x_j\} \in \mathbb{R}^n \mid a_j \leq x_j \leq b_j \quad \text{for} \quad 1 \leq j \leq n \right\}
\]
for arbitrary real numbers \(a_j \leq b_j\). Closed cells may have \(a_j = b_j\) for some indices \(j\), while of course open cells for which \(a_j = b_j\) for any index \(j\) must be the empty set. The boundary of an open or closed cell is the closed set

\[
\partial \Delta = \partial \Delta = \left\{ x \in \mathbb{R}^n \mid a_j \leq x_j \leq b_j \text{ for } 1 \leq j \leq n, \text{ and at least one inequality is an equality} \right\}.
\]

A customary notation for cells in \(\mathbb{R}^1\), which are also called intervals, is

\[
(a, b) = \{ x \in \mathbb{R}^1 \mid a < x < b \} \quad \text{and} \quad [a, b] = \{ x \in \mathbb{R}^1 \mid a \leq x \leq b \};
\]

these can be mixed, as for instance \((a, b] = \{ x \in \mathbb{R}^1 \mid a < x \leq b \}\).

### 1.2 Topological Preliminaries

Some familiarity with the basic topological notions for subsets of the real line is assumed; but a brief review is included here, focusing on the extension of these notions to subsets of \(\mathbb{R}^n\), and it may be a sufficient introduction even for those seeing these concepts for the first time. However if these notions are not entirely familiar it is advisable as an additional exercise to prove explicitly and in detail those statements that are labeled "clear" or "evident" in the discussion here.

The basic topological property of the real number system is its completeness, the property that any set \(S\) of real numbers that is bounded from above has a least upper bound or supremum, denoted by \(\inf S\), or equivalently that any set \(S\) of real numbers that is bounded from below has a greatest lower bound or infimum, denoted by \(\inf S\); an alternative statement of this property is that any Cauchy sequence of real numbers converges to a real limit.

There are a variety of topological notions and terms that are in common use. A subset \(U \subset \mathbb{R}^n\) is said to be open if for each \(a \in U\) there is an \(\epsilon > 0\) such that \(N_{\epsilon}(a) \subset U\). It is clear from the inequalities (1.7) that the condition that a set be open is independent of which of the two norms \(\|x\|_2\) or \(\|x\|_\infty\) is used to define the neighborhood \(N_{\epsilon}(a)\). Intuitively a set \(U \subset \mathbb{R}^n\) is open if for any point \(a \in U\) all points that are near enough to \(a\) are also in the set \(U\). For example an open \(\epsilon\)-neighborhood \(N_{\epsilon}(a)\) of a point \(a \in \mathbb{R}^n\) and an open cell \(\Delta \subset \mathbb{R}^n\) are open subsets. An open subset of \(\mathbb{R}^n\) containing a point \(a\) is often called an open neighborhood of the point \(a\); an open \(\epsilon\)-neighborhood \(N_{\epsilon}(a)\) is thus an open neighborhood of the point \(a\) in this sense as well. The collection \(T\) of open subsets of \(\mathbb{R}^n\) has the following characteristic properties:

1. If \(U_\alpha \in T\) then \(\bigcup_\alpha U_\alpha \in T\);

2. If \(U_i \in T\) for \(1 \leq i \leq N\) for some \(N\) then \(\bigcap_{i=1}^N U_i \in T\);

3. \(\emptyset \in T\) and \(\mathbb{R}^n \in T\).

These properties can be summarized in the statement that an arbitrary union of open sets is open, a finite intersection of open sets is open, and the empty set and the set \(\mathbb{R}^n\) itself are open. That the open subsets of \(\mathbb{R}^n\) satisfy these three conditions is quite clear. It is also clear that an infinite intersection of open sets is not necessarily open; for instance \(\bigcap_{i=1}^\infty N_{\frac{1}{i}}(a) = a\) and a single point is not open. A collection \(T\) of subsets
of an arbitrary abstract set \( S \) having these three characteristic properties is said to be a \textit{topology} on the set \( S \); and the sets \( U \in \mathcal{T} \) are said to be the \textit{open sets} in this topology. For example if \( S \subset \mathbb{R}^n \) is an arbitrary subset of \( \mathbb{R}^n \) the intersections \( U \cap S \) of \( S \) with open sets \( U \subset \mathbb{R}^n \) are a topology on \( S \), called the \textit{relative topology} on \( S \) induced by the topology on \( \mathbb{R}^n \); and the sets \( U \in \mathcal{T} \) are called the \textit{relatively open} sets in this topology. It is evident that relatively open sets can be defined in parallel with the definition of open sets in \( \mathbb{R}^n \) as those sets \( E \subset S \) such that for any point \( a \in E \) there is an \( \epsilon > 0 \) for which \( \mathcal{N}_\epsilon(a) \cap S \subset E \). Clearly a set \( E \subset S \) can be open in the relative topology of \( S \) but not open in the usual topology of \( \mathbb{R}^n \).

A point \( a \in \mathbb{R}^n \) is a \textit{limit point} of a subset \( E \subset \mathbb{R}^n \) if the intersection \( \mathcal{N}_\epsilon(a) \cap E \) contains points of \( E \) other than \( a \) for all \( \epsilon > 0 \); clearly that is equivalent to the condition that \( \mathcal{N}_\epsilon(a) \cap E \) is an infinite set of points for all \( \epsilon > 0 \). The set of limit points of \( E \) is denoted by \( E' \) and is called the \textit{derived set} of \( E \). The set \( E \) is said to be \textit{closed} if \( E' \subset E \). For example a closed neighborhood \( \mathcal{N}_s(a) \) of a point \( a \in \mathbb{R}^n \) and a closed cell \( \Delta \) in \( \mathbb{R}^n \) are closed subsets. The \textit{closure} of a set \( E \), denoted by \( \overline{E} \), is defined to be the union \( \overline{E} = E \cup E' \), and is readily seen to be a closed set. A closed neighborhood \( \mathcal{N}_s(a) \) of a point \( a \in \mathbb{R}^n \) is the closure of the open neighborhood \( \mathcal{N}_\epsilon(a) \) and a closed cell \( \Delta \) in \( \mathbb{R}^n \) is the closure of the open cell \( \Delta \) provided that \( \Delta \neq \emptyset \). A basic topological result is that a subset \( E \subset \mathbb{R}^n \) is closed if and only if its complement \( F = \mathbb{R}^n \setminus E \) is open. To verify that, if \( E \) is closed and \( a \notin E \) then \( a \notin E' \) so \( a \) is not a limit point of \( E \) hence there must be some \( \epsilon > 0 \) such that \( \mathcal{N}_\epsilon(a) \cap E = \emptyset \); consequently \( \mathcal{N}_\epsilon(a) \subset F \) so \( F \) is open. Conversely if \( E \) is open then no point of \( E \) can be a limit point of \( F \), since for any point \( a \in E \) there is an \( \epsilon > 0 \) such that \( \mathcal{N}_\epsilon(a) \subset E \) hence that \( \mathcal{N}_\epsilon(a) \cap F = \emptyset \); consequently all the limit points of \( F \) are contained in \( F \) so \( F \) is closed. It is clear that the closed sets have the properties that a finite union of closed sets is closed, that any intersection of closed sets is closed, and that the empty set and the full set \( \mathbb{R}^n \) are both closed; a topology can be defined alternatively by giving a collection of sets satisfying these three conditions and defining the open sets to be their complements, although it is customary to define topologies directly in terms of the open sets. The \textit{boundary} of a subset \( E \subset \mathbb{R}^n \) is the closed set \( \partial E = \overline{E} \setminus \overline{E'} \). The boundaries of \( \epsilon \) neighborhoods \( \mathcal{N}_\epsilon(a) \) and of cells \( \Delta \) as defined earlier are also their boundaries in this sense. For more general sets some caution is necessary, since the boundary may not always correspond to what naively might be viewed as the boundary of the set; for example If \( E \) is the set of all points in an open cell \( \Delta \) that have rational coordinates then \( \partial E = \Delta \).

Two subsets \( E, F \subset \mathbb{R}^n \) are \textit{separated} if \( \overline{E} \cap \overline{F} = \emptyset \). This is somewhat stronger than that the two sets are disjoint; for example \( E = [0, 1) \) and \( F = [1, 2] \) are disjoint subsets of \( \mathbb{R} \) but are not separated since \( \overline{E} \cap F = 1 \), but \( E = [0, 1) \) and \( G = (1, 2] \) are separated subsets of \( \mathbb{R} \). A subset \( S \subset \mathbb{R}^n \) is \textit{connected} if it cannot be written as the union of two nonempty separated subsets. In terms of open sets alone, a topological space \( S \), such as a subset of \( \mathbb{R}^n \) with the relative topology, is connected if it cannot be written as a disjoint union of two relatively open subsets of \( S \), or equivalently if there is no subset of \( S \) that is both relatively open and relatively closed, other than the empty set and all of \( S \). The equivalence of these conditions is quite obvious. It is easy to see that the entire space \( \mathbb{R}^n \) is connected. Indeed if \( U \subset \mathbb{R}^n \) is a subset other than the empty set or all of \( \mathbb{R}^n \) that is both open and closed, choose a point \( a \in U \) and a point \( b \in \mathbb{R}^n \sim U \). The set of real numbers \( s \) such that \( a + s(b - a) \in U \) for \( 0 \leq t < s \) is nonempty, since \( U \) is open, and it is bounded above since \( b \notin U \), so this set has a least upper bound \( s_0 \), one of the characteristic properties of the real number system. Since \( U \) is closed it must be the case that \( a + s_0 b \in U \),
for this is the limit of the points \( a + sb \in U \) as \( s \to s_0 \); but then since \( U \) is open it also must be the case that \( a + sb \in U \) for some real numbers \( s > s_0 \), a contradiction. The same argument shows that an open neighborhood \( \mathcal{N}_r(a) \) and an open cell in \( \mathbb{R}^n \) are connected sets.

An open covering of a subset \( E \subset \mathbb{R}^n \) is a collection of open sets \( U_\alpha \), not necessarily a countable collection, such that \( E \subset \bigcup U_\alpha \). If some of the sets \( U_\alpha \) are redundant they can be eliminated and the remaining sets are also an open covering of \( E \), called a subcovering of the set \( E \). A set \( E \) is said to be compact if every open covering of \( E \) has a finite subcovering, that is, if for any open covering \{\( U_\alpha \)\} of \( E \) finitely many of the sets \( U_\alpha \) actually cover all of \( E \). This is a rather subtle notion, but is very important and frequently used. An example of a non-compact subset of \( \mathbb{R}^n \) is a nonempty open neighborhood \( \mathcal{N}_1(0) \) of the origin in \( \mathbb{R}^n \); this set is covered by the open subsets \( U_\nu = \mathcal{N}_{1 - \frac{1}{\nu}}(0) \) for \( \nu = 1, 2, 3, \ldots \), but the union of any finite collection of these subsets will just be the set \( U_\nu \) for the largest \( \nu \) in the collection, and that is a proper subset of \( \mathcal{N}_1(0) \). An open cell is also noncompact, for essentially the same reason. A closed cell is an example of a compact subset; but the proof is a bit subtler and rests on the basic topological properties of the real number system, just as did the proof that \( \mathbb{R}^n \) is connected. For the proof it is convenient to define the edgelsize of an open or closed cell \( \Delta = \{ x = (x_j) \in \mathbb{R}^n \mid a_j \leq x_j \leq b_j \} \subset \mathbb{R}^n \) to be the nonnegative number \( d(\Delta) = \max_{1 \leq j \leq n} (b_j - a_j) \).

Lemma 1.1 If \( \overline{\Delta}_\nu \subset \mathbb{R}^n \) are closed cells in \( \mathbb{R}^n \) for \( \nu = 1, 2, \ldots \) such that \( \overline{\Delta}_{\nu+1} \subset \overline{\Delta}_\nu \) and \( \lim_{\nu \to \infty} d(\overline{\Delta}_\nu) = 0 \) then \( \bigcap_\nu \overline{\Delta}_\nu \) is a single point of \( \mathbb{R}^n \).

Proof: If \( \overline{\Delta}_\nu \) = \{ \( x = (x_j) \mid a_j^\nu \leq x_j \leq b_j^\nu \) \} then for each \( j \) clearly \( a_j^{\nu+1} \geq a_j^\nu \) and \( b_j^{\nu+1} \leq b_j^\nu \); and since \( \overline{\Delta}_\nu \subset \overline{\Delta}_1 \) it is also the case that \( a_j^\nu \leq b_j^\nu \) and \( b_j^\nu \geq a_j^\nu \). The basic completeness property of the real number system implies that any increasing sequence of real numbers bounded from above and any decreasing sequence of real numbers bounded from below have limiting values; therefore \( \lim_{\nu \to \infty} a_j^\nu = a_j \) and \( \lim_{\nu \to \infty} b_j^\nu = b_j \) for some uniquely determined real numbers \( a_j, b_j \), and it is clear that the cell \( \Delta = \{ x = (x_j) \mid a_j \leq x_j \leq b_j \} \) is contained in the intersection \( \bigcap_\nu \overline{\Delta}_\nu \). On the other hand since \( (b_j - a_j) \leq (b_j^\nu - a_j^\nu) \leq d(\overline{\Delta}_\nu) \) and \( \lim_{\nu \to \infty} d(\overline{\Delta}_\nu) = 0 \) it must be the case that \( b_j = a_j \) so the limiting cell \( \Delta \) is just a single point of \( \mathbb{R}^n \), which concludes the proof.

Theorem 1.2 A closed cell in \( \mathbb{R}^n \) is compact.

Proof: If a closed cell \( \overline{\Delta} = \{ x = (x_j) \mid a_j \leq x_j \leq b_j \} \) is not compact there is an open covering \{\( U_\alpha \)\} of \( \overline{\Delta} \) that does not admit any finite subcovering. The cell \( \overline{\Delta} \) can be written as the union of the closed cells arising from bisecting each of its sides. If finitely many of the sets \( \{ U_\alpha \} \) covered each of the subcells then finitely many would cover the entire set \( \overline{\Delta} \), which is not the case; hence at least one of the subcells cannot be covered by finitely many of the sets \( \{ U_\alpha \} \). Then bisect each of the sides of that subcell, and repeat the process. The result is that there is a collection of closed cells \( \overline{\Delta}_\nu \) which cannot be covered by finitely many of the open sets \( \{ U_\alpha \} \) and for which \( \overline{\Delta}_{\nu+1} \subset \overline{\Delta}_\nu \subset \overline{\Delta} \) and \( \lim_{\nu \to \infty} d(\overline{\Delta}_\nu) \to 0 \). It then follows from the preceding lemma that \( \bigcap_\nu \overline{\Delta}_\nu = a, \) a single point in \( \mathbb{R}^n \). This point must be contained within one of the sets \( \{ U_\alpha \} \), and if \( \nu \) is sufficiently large then \( \overline{\Delta}_\nu \subset U_{a_0} \) as well; but that is a contradiction, since the cells \( \overline{\Delta}_\nu \) were chosen so that none of them could be covered
by finitely many of the sets \( \{U_\alpha\} \). Therefore the cell \( \Delta \) is compact, which concludes the proof.

**Theorem 1.3** A closed subset of a compact set is compact.

**Proof:** Suppose that \( E \subset F \) where \( F \) is compact and \( E \) is closed, and that \( \{U_\alpha\} \) is an open covering of the set \( E \). The sets \( U_\alpha \) together with the open set \( \mathbb{R}^n \sim E \) form a covering of the compact set \( F \), so finitely many of these sets cover \( F \) hence also cover \( E \). Clearly the set \( \mathbb{R}^n \sim E \) covers none of the points of \( E \), so the remaining finitely many sets \( U_\alpha \) necessarily cover \( E \). Therefore \( E \) is compact, which suffices for the proof.

**Theorem 1.4 (Heine-Borel Theorem)** A subset \( E \subset \mathbb{R}^n \) is compact if and only if it is closed and bounded.

**Proof:** A bounded subset \( E \subset \mathbb{R}^n \) is contained in a sufficiently large closed cell \( \Delta \subset \mathbb{R}^n \); the set \( \Delta \) is compact by Theorem 1.2, so if \( E \) is also closed then by Theorem 1.3 it is compact. Conversely suppose that \( E \) is a compact set. If \( E \) is not bounded it can be covered by the collection of open neighborhoods \( \mathcal{N}_\nu(0) \) for \( \nu = 1, 2, \ldots \), but it is not covered by any finite set of these neighborhoods; that contradicts the compactness of \( E \), so a compact set is bounded. If \( E \) is not closed then there is a limit point \( a \in E' \) that is not contained in \( E \). Each closed neighborhood \( \overline{\mathcal{N}}_{1/\nu}(a) \) of the point \( a \) must contain a point of \( E \), since \( a \in E' \), but the intersection of all of these neighborhoods consists of the point \( a \) itself, which is not contained in \( E \). The complements \( U_\nu = \mathbb{R}^n \sim \overline{\mathcal{N}}_{1/\nu}(a) \) then form an open covering of \( E \), but no finite number of these sets suffice to cover \( E \) since the union of finitely many such sets must be one of the sets \( U_\nu \) and hence does not cover the points of \( E \) contained in \( \overline{\mathcal{N}}_{1/\nu}(a) \); and that again is a contradiction, which suffices to conclude the proof.

**Theorem 1.5 (Casorati-Weierstrass Theorem)** A subset \( E \subset \mathbb{R}^n \) is compact if and only if every sequence of distinct points in \( E \) has a limit point in \( E \).

**Proof:** Suppose that \( E \) is compact. If \( a_\nu \subset E \) is a collection of distinct points of \( E \) with no limit points in \( E \) then the points \( a_\nu \) can have no limit points in \( \mathbb{R}^n \), for since \( E \) is closed these limit points would necessarily lie in \( E \); in particular the set \( \bigcup a_\nu \) is a closed set. Each point \( a_\nu \) has an open neighborhood \( U_\nu \) that contains none of the other points, since otherwise \( a_\nu \) would itself be a limit point of this collection of points. These open sets \( U_\nu \) together with the open set \( \mathbb{R}^n \sim (\bigcup a_\nu) \) form an open covering of \( E \); and since \( E \) is compact finitely many of these sets already cover \( E \). That is a contradiction, since the set \( \mathbb{R}^n \sim (\bigcup a_\nu) \) does not cover any of the points \( a_\nu \) and no finite collection of the sets \( U_\nu \) cover all the points \( a_\nu \). On the other hand, suppose that \( E \subset \mathbb{R}^n \) is not compact. Then by the Heine-Borel Theorem either \( E \) is not bounded or \( E \) is not closed. If \( E \) is not bounded it must contain a sequence of distinct points \( a_\nu \) such that \( |a_\nu| \) is a strictly increasing sequence of real numbers with no finite limit, and this sequence can have no limit point in \( E \). On the other hand if \( E \) is not closed it contains a sequence of distinct points \( a_\nu \) with a limit point not contained in \( E \). That suffices to conclude the proof.

An incidental observation of interest is that the property that a set is compact is more intrinsic than the property of it being open or closed. It is clear that if \( S \subset \mathbb{R}^n \)
has the relative topology, a subset \( E \subset S \) that is open in the relative topology of \( S \) need not be open in the relative topology of \( \mathbb{R}^n \), and a subset \( E \subset S \) that is closed in the relative topology of \( S \) need not be closed in the relative topology of \( \mathbb{R}^n \); indeed an open interval in the line \( S = \mathbb{R}^1 \) is not open in \( \mathbb{R}^2 \) when the line is imbedded in the plane, and an open cell \( S = \Delta \subset \mathbb{R}^n \) is a closed subset of itself in the topology inherited from \( \mathbb{R}^n \), since it is the whole set \( S \), but is not closed in \( \mathbb{R}^n \). However a subset \( E \subset S \) is compact as a subset of \( S \) in the topology of \( S \) if and only if it compact as a subset of \( \mathbb{R}^n \) in the topology of \( \mathbb{R}^n \), as a clear consequence of the definition of the relatively open subsets of \( S \) as the intersections with \( S \) of open subsets in \( \mathbb{R}^n \).

1.3 Continuous Mappings

A mapping \( f \) from a subset \( U \subset \mathbb{R}^m \) to a subset \( V \subset \mathbb{R}^n \) associates to each point \( x = \{x_j\} \in U \) a point \( f(x) = y = \{y_j\} \in V \); the coordinates \( y_j \) depend on the point \( x \) so can be viewed as given by functions \( y_j = f_j(x) \), which are the coordinate functions of the mapping \( f \). The mapping \( f \) is continuous at a point \( a \in U \) if for every \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that \( ||f(a) - f(x)|| \leq \varepsilon \) whenever \( ||x - a|| \leq \delta \), or equivalently, for every \( \varepsilon > 0 \) there is a \( \delta > 0 \) such that \( f(x) \in N_\varepsilon(a) \) whenever \( x \in N_\delta(a) \). It is clear from the inequalities (1.7) that in the definition of continuity the norm \( ||x|| \) can be either the Euclidean norm \( ||x||_2 \) or the supremum norm \( ||x||_\infty \). It is also clear that the mapping \( f \) is continuous at a point \( a = \{a_j\} \) if and only if each of its coordinate functions \( f_j \) is continuous at the point \( a = \{a_j\} \). The mapping is said to be continuous on the subset \( U \) if it is continuous at each point of \( U \).

**Theorem 1.6** A mapping \( f : S \rightarrow \mathbb{R}^n \) from a subset \( S \subset \mathbb{R}^m \) into \( \mathbb{R}^n \) is continuous in \( S \) if and only if \( f^{-1}(U) \) is an open subset of \( S \) in the relative topology of \( S \) for any open subset \( U \subset \mathbb{R}^n \).

**Proof:** If \( f : S \rightarrow V \) is continuous, \( U \subset \mathbb{R}^n \) is an open subset and \( a \in f^{-1}(U) \) then \( f(a) = b \in U \), and since \( U \) is open there is an open neighborhood \( N_\varepsilon(b) \subset U \). Since \( f \) is continuous there is a \( \delta \) such that \( f(x) \in N_\varepsilon(f(a)) \) whenever \( x \in N_\delta(a) \); and consequently \( N_\delta(a) \subset f^{-1}(U) \), so \( f^{-1}(U) \) is an open subset. On the other hand if \( f^{-1}(U) \) is an open subset of \( S \) for any open subset \( U \subset \mathbb{R}^n \) then in particular for any \( \varepsilon > 0 \) the set \( f^{-1}(N_\varepsilon(b)) \) is an open subset of \( U \), so \( N_\delta(a) \subset f^{-1}(N_\varepsilon(b)) \) for some \( \delta \), and consequently the mapping \( f \) is continuous at any point \( a \in U \). That concludes the proof.

**Corollary 1.1** A mapping \( f : S \rightarrow \mathbb{R}^n \) from a subset \( S \subset \mathbb{R}^m \) into \( \mathbb{R}^n \) is continuous in \( S \) if and only if \( f^{-1}(E) \) is a closed subset of \( S \) in the relative topology of \( S \) for any closed subset \( E \subset \mathbb{R}^n \).

**Proof:** This is an immediate consequence of the preceding theorem, since a subset of \( S \) is closed in the relative topology of \( S \) if and only if its complement in \( S \) is relatively open and \( f^{-1}(\mathbb{R}^n \sim E) = (S \sim f^{-1}(E)) \) for any subset \( E \subset \mathbb{R}^n \). That suffices for the proof.

These results show that the continuity of a mapping in a set \( S \) really is a property of the topology of \( S \). A simple consequence of either result is that if \( g : U \rightarrow V \) and \( f : V \rightarrow W \) are continuous mappings between subsets \( U,V,W \) of some Euclidean
spaces then the composition \( f \circ g : U \to W \) that takes a point \( x \in U \) to the point \((f \circ g)(x) = f(g(x))\) is also continuous; indeed if \( E \subset W \) is open then \( f^{-1}(E) \) is open since \( f \) is continuous and \((f \circ g)^{-1}(E) = g^{-1}(f^{-1}(E))\) is open since \( g \) is continuous, and consequently \((f \circ g)\) is continuous. The results in Theorem 1.6 and Corollary 1.1 involve the inverse image of a set under a mapping \( f \); the image of an open set under a continuous mapping is not necessarily open, and the image of a closed set under a continuous mapping is not necessarily closed. For instance if \( f : \mathbb{R} \to \mathbb{R} \) is the continuous mapping \( f(x) = e^{-x^2} \) then \( f(\mathbb{R}) = \{ x \mid 0 < x \leq 1 \} \), which is neither open nor closed although \( \mathbb{R} \) itself is both open and closed.

**Theorem 1.7** If \( f : S \to \mathbb{R}^n \) is a continuous mapping from a subset \( S \subset \mathbb{R}^m \) into \( \mathbb{R}^n \) then the image \( f(E) \) of any compact subset \( E \subset S \) is a compact subset of \( \mathbb{R}^n \).

**Proof:** If \( E \subset S \) is compact and \( f(E) \) is contained in a union of open sets \( U_\alpha \) then \( E \) is contained in the union of the open sets \( f^{-1}(U_\alpha) \); and since \( E \) is compact it is contained in a union of finitely many of the sets \( U_\alpha \), so \( E \) is contained in the union of the inverse images of these finitely many open sets. That suffices for the proof.

Since a compact subset of \( \mathbb{R}^n \) is closed, by the Heine-Borel Theorem, the inverse under a continuous mapping of a compact set is necessarily closed; but it is not necessarily compact. For example the inverse image of the set \( [-\frac{1}{2}, \frac{1}{2}] \subset \mathbb{R} \) under the mapping \( f : \mathbb{R} \to \mathbb{R} \) given by \( f(x) = \sin x \) is an unbounded set hence is not compact.

**Corollary 1.2** If \( f \) is a continuous function on a compact set \( U \subset \mathbb{R}^n \) then there are points \( a, b \in U \) such that

\[
(1.18) \quad f(a) = \sup_{x \in U} f(x) \quad \text{and} \quad f(b) = \inf_{x \in U} f(x).
\]

**Proof:** The image \( f(U) \subset \mathbb{R} \) is compact by the preceding theorem, hence is a closed set; so if \( \alpha = \sup_{x \in U} f(x) \) then since \( \alpha \) is a limit point of the set \( f(U) \) it must be contained in the set \( f(U) \) hence \( \alpha = f(a) \) for some point \( a \in U \), and correspondingly for \( \beta = \inf_{x \in U} f(x) \). That suffices for the proof.

**Theorem 1.8** A one-to-one continuous mapping from a compact subset \( U \subset \mathbb{R}^m \) onto a subset \( V \subset \mathbb{R}^n \) has a continuous inverse.

**Proof:** If the mapping \( f : U \to V \) is one-to-one it has a well defined inverse mapping \( g : V \to U \). To show that \( g \) is continuous it suffices to show that \( g^{-1}(E) \) is closed for any closed subset \( E \subset U \), in view of Corollary 1.1. If \( E \) is closed then by Theorem 1.3 it is compact, since \( U \) is compact; and then \( bg^{-1}(E) = f(E) \) is compact by Theorem 1.7, hence is closed by the Heine-Borel Theorem, and that suffices for the proof.

The assumption of compactness is essential in the preceding theorem. For example the mapping \( f : [0, 2\pi] \to \mathbb{R}^2 \) defined by

\[
f(t) = (\cos t, \sin t) \in \mathbb{R}^2
\]

is clearly one-to-one and continuous; but the inverse mapping fails to be continuous at the point \( f(0) = (1, 0) \). Some properties of continuity are not purely topological properties, in the sense that they cannot be stated purely in terms of open and closed sets, but are metric properties, involving the norms used to define continuity. By definition
a mapping \( f : U \to W \) between two subsets of Euclidean spaces is continuous at a point \( a \in U \) if and only if for every \( \epsilon > 0 \) there is a \( \delta > 0 \) such that \( ||f(a) - f(x)|| \leq \epsilon \) whenever \( ||x - a|| \leq \delta \). The mapping \( f \) is continuous in the set \( U \) if for each point \( a \in U \) and any \( \epsilon > 0 \) there is a \( \delta_a > 0 \), which may depend on the point \( a \), such that \( ||f(a) - f(x)|| \leq \epsilon \) whenever \( ||x - a|| \leq \delta_a \). The mapping \( f \) is uniformly continuous in \( U \) if it is possible to find values \( \delta_a \) that are independent of the point \( a \in U \).

Equivalently the mapping \( f \) is uniformly continuous in \( U \) if for any \( \epsilon > 0 \) there exists \( \delta > 0 \) such that \( ||f(x) - f(y)|| < \epsilon \) whenever \( ||x - y|| < \delta \). It should be kept in mind that the two norms are in different spaces; and it is evident from the inequalities (1.7) that they may be different norms as well. Not all continuous mappings are uniformly continuous, as for example the mapping \( f : \mathbb{R} \to \mathbb{R} \) defined by \( f(x) = x^2 \); but in some circumstances continuous mappings are automatically uniformly continuous.

**Theorem 1.9** A continuous mapping \( f : U \to \mathbb{R}^n \) from a compact subset \( U \subset \mathbb{R}^m \) into \( \mathbb{R}^n \) is uniformly continuous.

**Proof:** If \( f : U \to \mathbb{R}^n \) and \( \epsilon > 0 \) then for any point \( a \in U \) there is \( \delta_a > 0 \) such that \( ||f(x) - f(a)|| < \frac{1}{2}\epsilon \) whenever \( x \in N_{\delta_a}(a) \). The collection of neighborhoods \( N_{\frac{1}{2}\delta_a}(a) \) for all points \( a \in U \) are an open covering of \( U \), and if \( U \) is compact finitely many of these neighborhoods cover all of \( U \). If \( \delta > 0 \) is the minimum of the finitely many positive numbers \( \delta_a \) for this finite covering, and if \( x, y \in U \) are any two points such that \( ||x - y|| < \frac{1}{2}\delta \), then \( x \in N_{\frac{1}{2}\delta}(a) \) for one of these neighborhoods, and since \( ||x - y|| < \frac{1}{2}\delta \) it is also the case that \( y \in N_{\delta}(a) \). It follows that \( ||f(x) - f(y)|| \leq ||f(x) - f(a)|| + ||f(a) - f(y)|| \leq \frac{1}{2}\epsilon + \frac{1}{2}\epsilon = \epsilon \), so \( f \) is uniformly continuous, which concludes the proof.
Chapter 2

Differentiable Mappings

2.1 The Derivative

A mapping $f: U \rightarrow \mathbb{R}^n$ from an open subset $U \subset \mathbb{R}^m$ into $\mathbb{R}^n$ is **differentiable** at a point $a \in U$ if there is a linear mapping $A: \mathbb{R}^m \rightarrow \mathbb{R}^n$ described by an $n \times m$ matrix $A$ such that for all $h$ in an open neighborhood of the origin in $\mathbb{R}^m$

\[ f(a + h) = f(a) + Ah + \epsilon(h) \]  

where $\lim_{h \to 0} \frac{||\epsilon(h)||}{||h||} = 0$.

Here $h \in \mathbb{R}^m$ so $Ah \in \mathbb{R}^n$, while $f(a)$, $f(a + h)$, $\epsilon(h) \in \mathbb{R}^n$. This definition is independent of the norm chosen; for if $\lim_{h \to 0} \frac{||\epsilon(h)||}{||h||_2} = 0$ then from the inequalities (1.7) it follows that $\frac{||\epsilon(h)||}{||h||_\infty} \leq \frac{||\epsilon(h)||}{||h||_2} \frac{1}{\sqrt{n}}$ so $\lim_{h \to 0} \frac{||\epsilon(h)||}{||h||_\infty} = 0$ as well, and the converse holds similarly. If $f$ is differentiable at $a$ it is continuous at $a$, since $\lim_{h \to 0} A h = 0$ and $\lim_{h \to 0} \epsilon(h) = 0$. For example, if $m = n = 1$ it is possible to divide (2.1) by the real number $h$ and to rewrite that equation in the form

\[ \lim_{h \to 0} \frac{f(a + h) - f(a)}{h} = A \lim_{h \to 0} \frac{\epsilon(h)}{h} = 0; \]

this is a form of the familiar definition that the real-valued function $f(x)$ of the variable $x \in \mathbb{R}$ is differentiable at the point $a$ and that its derivative at that point is the real number $A$. For another example, if $m = 3$ and $n = 2$ so that $f: \mathbb{R}^3 \rightarrow \mathbb{R}^2$ then (2.1) takes the form

\[
\begin{pmatrix}
  f_1(a + h) \\
  f_2(a + h)
\end{pmatrix} = \begin{pmatrix}
  f_1(a) \\
  f_2(a)
\end{pmatrix} + \begin{pmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23}
\end{pmatrix} \begin{pmatrix}
  h_1 \\
  h_2 \\
  h_3
\end{pmatrix} + \begin{pmatrix}
  \epsilon_1(h) \\
  \epsilon_2(h)
\end{pmatrix}.
\]

**Theorem 2.1** A mapping $f: \mathbb{R}^m \rightarrow \mathbb{R}^n$ is differentiable at a point $a$ if and only if each of the coordinate functions $f_i$ of the mapping $f$ is differentiable at that point.

**Proof:** If $f$ is a differentiable mapping it follows from (2.1) that each coordinate function $f_i$ satisfies

\[ f_i(a + h) = f_i(a) + \sum_{j=1}^{m} a_{ij}h_j + \epsilon_i(h) \]  

where $\lim_{h \to 0} \frac{||\epsilon_i(h)||}{||h||} = 0,$
since \( \frac{|e_i(h)|}{\|h\|_\infty} \leq \frac{\|e(h)\|_\infty}{\|h\|_\infty} \); and this is just the condition that each of the coordinate functions \( f_i \) of the mapping \( f \) is differentiable at the point \( a \). Conversely if each of the coordinate mappings \( f_i \) is differentiable at the point \( a \) then (2.2) holds for \( 1 \leq i \leq n \). The collection of these \( n \) equations taken together form the equation (2.1) in which \( \epsilon(h) = \{e_i(h)\} \); and since \( \frac{|e(h)|}{\|h\|_\infty} \leq \max_{1 \leq i \leq n} \frac{|e_i(h)|}{\|h\|_\infty} \) it follows that \( \lim_{h \to 0} \frac{|e(h)|}{\|h\|_\infty} = 0 \), so the mapping \( f \) is differentiable, and that concludes the proof.

For the special case of a vector \( h = \{h_j\} \) for which \( h_j = 0 \) for \( j \neq k \) for some index \( k \), equation (2.2) takes the form

\[
\lim_{h_k \to 0} \frac{|e_k(h_k)|}{\|h_k\|} = 0;
\]

that is just the condition that \( f_i(x) \), viewed as a function of the variable \( x_k \) alone for fixed values \( x_j = a_j \) for the remaining variables for \( j \neq k \), is a differentiable function of that variable \( x_k \) and that its derivative at the point \( x_k = a_k \) is the real number \( a_{ik} \). The constant \( a_{ik} \) is called the partial derivative of the function \( f_i \) with respect to the variable \( x_k \) at the point \( a \), and is customarily denoted by \( a_{ik} = \partial_{x_k} f_i(a) \). It follows that the entries in the matrix \( A = \{a_{ik}\} \) are the uniquely determined partial derivatives of the coordinate functions of the mapping; this matrix is called the derivative of the mapping \( f \) at the point \( a \) and is denoted by \( f'(a) \), so

\[
f'(a) = \left\{ (f'(a))_{ik} = \partial_{x_k} f_i(a) \mid 1 \leq k \leq m, \ 1 \leq i \leq n \right\}.
\]

Generally it is a fairly straightforward matter to calculate the partial derivatives of a function; merely consider all the variables except one of them as constants, and apply the familiar techniques for calculating derivatives of a function of one variable. That can be applied to each coordinate function of a mapping, to yield the derivative of that mapping. It is evident from (2.1) that a linear combination \( c_1 f_1 + c_2 f_2 \) of two differentiable mappings \( f_1, f_2 : \mathbb{R}^m \to \mathbb{R}^n \) at a point \( a \) for any constants \( c_1, c_2 \in \mathbb{R} \) is again a differentiable mapping at that point, and it follows from (2.3) that differentiation is linear in the sense that \( \partial f_1 + c_2 f_2)'(a) = c_1 f_1'(a) + c_2 f_2'(a) \). There are various alternative notations for derivatives and partial derivatives of functions of several variables that are in common use. For instance \( Df(a) \) is often used for \( f'(a) \) and \( \partial_x f(a) \) or \( \frac{\partial f}{\partial x}(a) \) are commonly used for \( \partial_{x_k} f(a) \).

If a mapping \( f : \mathbb{R}^m \to \mathbb{R}^n \) is differentiable at a point \( a \in \mathbb{R}^m \) then it has partial derivatives \( \partial_{x_j} f_i(a) \) with respect to each variable \( x_j \) at that point; but it is not true conversely that if the coordinate functions of a mapping \( f \) have partial derivatives at a point \( a \) with respect to each variable \( x_j \) then \( f \) is a differentiable mapping. For example the mapping \( f : \mathbb{R}^2 \to \mathbb{R} \) defined by

\[
f(x) = \begin{cases} \frac{x_1 x_2}{x_1^2 + x_2^2} & \text{if } x \neq 0, \\ 0 & \text{if } x = 0 \end{cases}
\]

vanishes identically in the variable \( x_2 \) if \( x_1 = 0 \), so \( \partial_{x_2} f(0,0) = 0 \), and similarly \( \partial_{x_1} f(0,0) = 0 \). However the function is not continuous at the origin, since for instance it takes the value \( \frac{1}{2} \) whenever \( x_1 = x_2 \) except at the origin where it takes the value 0; hence it is not differentiable at the origin.
2.1. THE DERIVATIVE

Theorem 2.2 If the partial derivatives of a mapping \( f : \mathbb{R}^m \to \mathbb{R}^n \) exist at all points near \( a \) and are continuous at the point \( a \) then the mapping \( f \) is differentiable at the point \( a \).

Proof: In view of Theorem 2.1 it is enough to prove this for the special case that \( n = 1 \), in which case the mapping \( f : \mathbb{R}^m \to \mathbb{R} \) is just a real valued function; and for convenience only the case \( m = 2 \) will be demonstrated in detail, since it is easier to follow the proof in the simpler case and all the essential ideas are present. Assume that the partial derivatives \( \partial_k f_i(x) \) exist for all points \( x \) near \( a \) and are continuous at \( a = \{a_j\} \), and consider a fixed vector \( h = \{h_j\} \). When one of the variables is held fixed and \( f \) is viewed as a function of the remaining variable it is a differentiable function of a single variable. The mean value theorem for functions of a single variable asserts that if \( f(x) \) is continuous in a closed interval \([a, b]\) and is differentiable at each point of the open interval \((a, b)\) then \( f(b) - f(a) = f'(c)(b - a) \) for some point \( c \in (a, b) \); this can be applied to the function \( f(x_1, a_2) \) of the single variable \( x_1 \) in the interval between \( a_1 \) and \( a_1 + h_1 \), and to the function \( f(a_1 + h_1, x_2) \) of the single variable \( x_2 \) in the interval between \( a_2 \) and \( a_2 + h_2 \) if \( \|h\| \) is sufficiently small; as a consequence there exist values \( \alpha_1 \) between \( a_1 \) and \( a_1 + h_1 \) and \( \alpha_2 \) between \( a_2 \) and \( a_2 + h_2 \) such that

\[
\begin{align*}
    f(a_1 + h_1, a_2) - f(a_1, a_2) &= h_1 \partial_1 f(a_1, a_2) \quad \text{and} \\
    f(a_1 + h_1, a_2 + h_2) - f(a_1 + h_1, a_2) &= h_2 \partial_2 f(a_1 + h_1, a_2).
\end{align*}
\]

Then

\[
\begin{align*}
    f(a + h) - f(a) &= f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) = \\
    &= (f(a_1 + h_1, a_2 + h_2) - f(a_1 + h_1, a_2)) + (f(a_1 + h_1, a_2) - f(a_1, a_2)) \\
    &= h_2 \partial_2 f(a_1 + h_1, a_2) + h_1 \partial_1 f(a_1, a_2) \\
    &= h_2 \partial_2 f(a_1, a_2) + h_1 \partial_1 f(a_1, a_2) + \epsilon(h)
\end{align*}
\]

where

\[
\epsilon(h) = h_2 \left( \partial_2 f(a_1 + h_1, a_2) - \partial_2 f(a_1, a_2) \right) + h_1 \left( \partial_1 f(a_1, a_2) - \partial_1 f(a_1, a_2) \right).
\]

By the triangle inequality

\[
\frac{|\epsilon(h)|}{\|h\|} \leq \left| \partial_2 f(a_1 + h_1, a_2) - \partial_2 f(a_1, a_2) \right| + \left| \partial_1 f(a_1, a_2) - \partial_1 f(a_1, a_2) \right|
\]

since \( \frac{|h_1|}{\|h\|} \leq 1 \) and \( \frac{|h_2|}{\|h\|} \leq 1 \); and since the partial derivatives are assumed to be continuous at the point \( a \) it follows that \( \lim_{h \to 0} \frac{|\epsilon(h)|}{\|h\|} = 0 \). That shows that the mapping \( f : \mathbb{R}^2 \to \mathbb{R} \) is differentiable at the point \( (a_1, a_2) \), which suffices to conclude the proof.

The converse of the preceding theorem does not hold; if a mapping \( f : \mathbb{R}^m \to \mathbb{R}^n \) is differentiable at all points of an open subset \( U \subset \mathbb{R}^m \) the partial derivatives of the coordinate functions of the mapping \( f \) exist at each point of \( U \), but need not be continuous functions in \( U \). The standard example for functions of a single variable is the function

\[
f(x) = \begin{cases} 
    \sin \frac{1}{x} & \text{if } x \neq 0, \\
    0 & \text{if } x = 0, 
\end{cases}
\]
which is differentiable at all points \( x \in \mathbb{R} \) but for which the derivative is not continuous at the point 0; indeed if \( x \neq 0 \) then \( f'(x) = -\frac{1}{x^2} \cos \frac{1}{x} \) so that \( \lim_{x \to 0} f'(\frac{1}{x}) = +\infty \), but

\[
f'(0) = \lim_{x \to 0} \sin \frac{1}{x} - 0 = \lim_{x \to 0} x \sin \frac{1}{x} = 0
\]

since \( |\sin \frac{1}{x}| \leq 1 \) for \( x \neq 0 \). If a mapping \( f : U \to \mathbb{R}^n \) defined in an open subset \( U \subset \mathbb{R}^m \) is differentiable at all points \( x \in U \), then the mapping that associates to each point \( x \in U \) the matrix \( f'(x) \) is a well defined mapping \( f' : U \to \mathbb{R}^{n \times m} \) where \( \mathbb{R}^{n \times m} \) is the vector space consisting of all \( n \times m \) matrices; the mapping \( f \) is continuously differentiable or of class \( C^1 \) in \( U \) if the mapping \( f' \) is continuous.

**Corollary 2.1** A mapping \( f : U \to \mathbb{R}^n \) defined in an open subset \( U \subset \mathbb{R}^m \) is continuously differentiable if and only if all the partial derivatives of its coordinate functions exist and are continuous functions in \( U \).

**Proof:** If the partial derivatives of the mapping \( f \) exist and are continuous then it follows from the preceding theorem that \( f \) is differentiable; and since \( f'(x) = \{\partial_i f_j(x)\} \) the mapping \( f' \) is continuous, hence \( f \) is continuously differentiable. Conversely if \( f \) is continuously differentiable then the partial derivatives of all of its coordinate functions exist and are continuous in \( U \), which suffices for the proof.

### 2.2 The Chain Rule

If \( g : U \to \mathbb{R}^m \) is a mapping defined in an open neighborhood \( U \) of a point \( a \in \mathbb{R}^n \) and \( f : V \to \mathbb{R}^n \) is a mapping defined in an open neighborhood \( V \) of the point \( b = g(a) \in \mathbb{R}^m \), where \( g(U) \subset V \), the composition \( \phi = f \circ g : U \to \mathbb{R}^n \) is the mapping defined by \( \phi(x) = f(g(x)) \) for any \( x \in U \); this situation is described in the following diagram.

\[
\begin{array}{ccc}
\mathbb{R}^l & \mathbb{R}^m & \mathbb{R}^n \\
\cup & \cup & \cup \\
U & \xrightarrow{g} & V \xrightarrow{f} W \\
\end{array}
\]

(2.4)

\[
a \xrightarrow{g} b = g(a) \xrightarrow{f} f(b) = \phi(a)
\]

**Theorem 2.3** If the mapping \( g \) is differentiable at the point \( a \) and the mapping \( f \) is differentiable at the point \( b = g(a) \) then the composite function \( \phi = f \circ g \) is differentiable at the point \( a \) and \( \phi'(a) = f'(g(a)) \cdot g'(a) \).

**Proof:** Since the mapping \( f \) is differentiable at the point \( b \)

\[
f(b + h) = f(b) + f'(b)h + \varepsilon_f(h) \quad \text{where} \quad \lim_{h \to 0} \frac{\|\varepsilon_f(h)\|_\infty}{\|h\|_\infty} = 0,
\]

and since the mapping \( g \) is differentiable at the point \( a \)

\[
g(a + k) = g(a) + g'(a)k + \varepsilon_g(k) \quad \text{where} \quad \lim_{k \to 0} \frac{\|\varepsilon_g(k)\|_\infty}{\|k\|_\infty} = 0.
\]
Substituting (2.6) into (2.5) where \( b = g(a) \) and \( h = g'(a)k + \epsilon_k(k) \) leads to the result that

\[
\phi(a + k) = f(g(a + k)) = f(b + h) = f(b) + f'(b)h + \epsilon_r(h)
\]

\[
= \phi(a) + f'(b)(g'(a)k + \epsilon_k(k)) + \epsilon_r(h)
\]

hence that

\[
\phi(a + k) = \phi(a) + f'(b)g'(a)k + \epsilon(k)
\]

where \( \epsilon(k) = f'(b)\epsilon_k(k) + \epsilon_r(h) \). From the inequality (1.9) and the triangle inequality it follows that

\[
\frac{\|\epsilon(k)\|}{\|k\|} \leq \frac{\|f'(b)\epsilon_k(k)\|}{\|k\|} + \frac{\|\epsilon_r(h)\|}{\|k\|} + \frac{\|g'(a)k + \epsilon_k(k)\|}{\|k\|}
\]

\[
\leq n\|f'(b)\| \frac{\|\epsilon_k(k)\|}{\|k\|} + \frac{\|\epsilon_r(h)\|}{\|k\|} + \left( n\|g'(a)\| + \frac{\|\epsilon_k(k)\|}{\|k\|} \right)
\]

Since \( \lim_{k \to 0} \frac{\|\epsilon_k(k)\|}{\|k\|} = 0 \) and \( \lim_{h \to 0} \frac{\|\epsilon_r(h)\|}{\|k\|} = 0 \) while \( \lim_{h \to 0} k = 0 \) it follows from the preceding equation that \( \lim_{k \to 0} \frac{\|\epsilon(k)\|}{\|k\|} = 0 \), and it then follows from (2.7) that \( h = f \circ g \) is differentiable at the point \( a \) and that \( \phi'(a) = f'(g(a)) \cdot g'(a) \), which concludes the proof.

For a simple application of the chain rule, since the function \( g(y_1, y_2) = y_1y_2 \) is differentiable at any point \( y_1, y_2 \in \mathbb{R} \) and \( g'(y_1, y_2) = (y_2, y_1) \) it follows that for any differentiable mapping \( f = \{f_1, f_2\} : U \to \mathbb{R}^2 \) in an open subset \( U \subset \mathbb{R}^m \) the composition \( \phi = g \circ f : U \to \mathbb{R} \) is a differentiable mapping and

\[
\phi'(x) = g'(f(x))f'(x) = \begin{pmatrix} f_2(x) & f_1(x) \end{pmatrix} \begin{pmatrix} f'_1(x) \\ f'_2(x) \end{pmatrix} = f_2(x)f'_1(x) + f_1(x)f'_2(x) ;
\]

this is just an extension of the familiar product rule for differentiating functions from the case of functions of a single variable to the case of functions of several variables, since \( \phi(x) = f_1(x) f_2(x) \). The entries in the matrix \( \phi'(x) \) thus have the form

\[
\partial_i \phi(x) = f_2(x)\partial_i f(x) + f_1(x)\partial_i f_2(x).
\]

The corresponding argument shows that the quotient \( f_1/f_2 \) of two differentiable functions is differentiable at any point \( x \) at which \( f_2(x) \neq 0 \), and that its derivative has the familiar form. For another direct application of the chain rule, if \( f : U \to V \) is a one-to-one mapping between two open subsets \( U, V \subset \mathbb{R}^m \) and if \( g : V \to U \) is the inverse mapping then \( g \circ f : U \to U \) is the identity mapping, so that \( (g \circ f)'(x) = I \) where I is the \( m \times m \) identity matrix. If the mappings \( f \) and \( g \) are continuously differentiable then by the chain rule \( g'(f(x)) \cdot f'(x) = I \); thus the matrix \( g'(f(x)) \) is the inverse of the matrix \( f'(x) \) at each point \( x \in U \), so both matrices are nonsingular matrices at each point \( x \in U \).

An alternative notation for the chain rule is suggestive and sometimes quite useful. When mappings \( f : \mathbb{R}^l \to \mathbb{R}^m \) and \( g : \mathbb{R}^m \to \mathbb{R}^n \) are described in terms of the coordinates \( t = \{t_1, \ldots, t_l\} \in \mathbb{R}^l \) and \( x = \{x_1, \ldots, x_m\} \in \mathbb{R}^m \) and \( y = \{y_1, \ldots, y_l\} \in \mathbb{R}^n \).
the coordinate functions of the mappings \( f, g \) and \( \phi \) have the form \( y_i = f_i(x) = \phi_i(t) \) and \( x_j = g_j(t) \). The partial derivatives are sometimes denoted by

\[
(\phi')_{ik} = \partial_k \phi_i = \frac{\partial y_i}{\partial x_k}, \quad (f')_{ij} = \partial_j f_i(x) = \frac{\partial y_i}{\partial x_j}, \quad (g')_{jk} = \partial_k g_j(t) = \frac{\partial x_j}{\partial t_k}.
\]

By the preceding theorem the derivative of the composite function \( \phi = f \circ g \) is the matrix product \( \phi' = f'g' \), which in terms of the entries of these matrices is

\[
(\phi')_{ik} = \sum_{j=1}^n (f')_{ij}(g')_{jk} \quad \text{or equivalently} \quad \partial_k \phi_i = \sum_{j=1}^n \partial_j f_i \cdot \partial_k g_j; \quad \text{and in the alternative notation this takes the form}
\]

\[
\frac{\partial y_i}{\partial t_k} = \sum_{j=1}^n \frac{\partial y_i}{\partial x_j} \frac{\partial x_j}{\partial t_k}.
\]

This is the extension to mappings in several variables of the traditional formulation of the chain rule for functions of a single variable as the identity \( \frac{du}{dt} = \frac{du}{dx} \cdot \frac{dx}{dt} \); this form of the chain rule is in some ways easier to remember, and with some caution, easier to use, than the version of the chain rule in the preceding theorem. It is customary however to omit any explicit mention of the points at which the derivatives are taken; so some care must be taken to remember the the derivative \( \frac{\partial y}{\partial x} \) is evaluated at the point \( x \) while the derivatives \( \frac{\partial y}{\partial t} \) and \( \frac{\partial x}{\partial t} \) are evaluated at the point \( t \). This lack of clarity means that some caution must be taken when this notation is used.

Some care also must be taken with the chain rule in those cases where the compositions are not quite so straightforward. For instance if \( \phi(x_1, x_2) = f(x_1, x_2, g(x_1, x_2)) \) for a function \( f(x_1, x_2, x_3) \) of three variables and a function \( g(x_1, x_2) \) of two variables, the function \( \phi \) is really the composition \( \phi = f \circ G \) of the mappings \( f: \mathbb{R}^3 \rightarrow \mathbb{R} \) given by the function \( f \) and the mapping \( G: \mathbb{R}^2 \rightarrow \mathbb{R}^3 \) given by

\[
G(x_1, x_2) = \begin{pmatrix} x_1 \\ x_2 \\ g(x_1, x_2) \end{pmatrix},
\]

so by the chain rule

\[
\phi'(x) = f'(G(x))g'(x) = \begin{pmatrix} \partial_1 f(G(x)) & \partial_2 f(G(x)) & \partial_3 f(G(x)) \\ \partial_1 g(x) & \partial_2 g(x) \end{pmatrix}
\]

\[
= \begin{pmatrix} 1 & 0 & \partial_1 g(x) + \partial_2 f(G(x)) \partial_1 g(x) \partial_3 f(G(x)) + \partial_1 f(x) \partial_2 g(x) \end{pmatrix}
\]

and for the coordinate functions of the matrix \( \phi'(x) \)

\[
\partial_1 \phi(x) = \partial_1 f(G(x)) + \partial_3 f(G(x)) \partial_1 g(x), \quad \partial_2 \phi(x) = \partial_2 f(G(x)) + \partial_1 f(G(x)) \partial_2 g(x).
\]

This amounts to calculating the partial derivative \( \partial_1 \phi(x) \) as the sum of the partial derivatives of the function \( f(x_1, x_2, g_2(x_1, x_2)) \) with respect to each of its three variables, and multiplying each of these derivatives by the derivative of what is in the place of that variable with respect to the variable \( x_1 \). Some practice, checked by going
2.3. Higher Derivatives

Back to the form of the chain rule given in Theorem 2.3, may prove helpful; and if there are any doubts about an application of the chain rule they can be cleared up by identifying the function as an explicit composition of mappings. It should be noted that in this case the meaning of the expression

\[ \frac{\partial}{\partial x_1} f(x_1, x_2, x_3) \quad \text{where} \quad x_3 = g(x_1, x_2) \]

is not clear; it may mean either the derivative of the function \( f \) with respect to its first variable or the derivative of the composite function of the two variables \( x_1, x_2 \) with respect to the variable \( x_1 \), while \( \partial_1 f(x_1, x_2, x_3) \) where \( x_3 = g(x_1, x_2) \) is less ambiguous.

The chain rule also is useful in deriving information about the derivatives of functions that are defined only implicitly. For example if a function \( f(x_1, x_2) \) satisfies the equation

\[ f(x_1, x_2)^2 + x_1 f(x_1, x_2) + f(x_1, x_2) = 2x_1 + 3x_2 \]

and the initial condition that \( f(0,0) = 0 \) then the values of that function are determined implicitly but not explicitly by the preceding equation. This equation is the condition that the composition of the mapping \( F: \mathbb{R}^2 \to \mathbb{R}^3 \) defined by \( F(x_1, x_2) = (x_1, x_2, f(x_1, x_2)) \) and the mapping \( G: \mathbb{R}^3 \to \mathbb{R} \) defined by \( G(x_1, x_2, y) = y^2 + x_1 y + y - 2x_1 - 3x_2 \) is the trivial mapping \( G \circ F(x_1, x_2) = 0 \), so that \( (G \circ F)'(0,0) = 0 \); and by the chain rule

\[ \partial_1 (G \circ F) = 5 f(x_1, x_2)^4 \partial_1 f(x_1, x_2) + x_1 \partial_1 f(x_1, x_2) + f(x_1, x_2) + \partial_1 f(x_1, x_2) - 2, \]

so since \( \partial_1 (G \circ F) = 0 \) and \( f(0,0) = 0 \) the preceding equation reduces to \( \partial_1 f(0,0) = 2 \). A similar calculation yields the value of \( \partial_2 f(0,0) \).

2.3 Higher Derivatives

If \( f: U \to \mathbb{R} \) is a function defined in an open set \( U \subset \mathbb{R}^2 \) and if the partial derivative \( \partial_1 f(x) \) exists at all points \( x \in U \) the function \( \partial_1 f(x) \) may itself have partial derivatives, such as \( \partial_2 (\partial_1 f(x)) \), which for convenience is shortened to \( \partial_2 \partial_1 f(x) \); and the process may continue, leading to \( \partial_3 \partial_2 \partial_1 f(x) \) and so on. The order in which successive derivatives are taken may be significant; for example, a straightforward calculation for the function

\[ f(x_1, x_2) = \begin{cases} \frac{x_1 x_2 (x_1^2 - x_2^2)}{x_1^2 + x_2^2} & \text{if} \quad (x_1, x_2) \neq (0,0) \\ 0 & \text{if} \quad (x_1, x_2) = (0,0) \end{cases} \]

shows that \( \partial_1 \partial_2 f(0,0) = 1 \) but \( \partial_2 \partial_1 f(0,0) = -1 \). However for sufficiently regular functions the order of differentiation is irrelevant.

**Theorem 2.4** If \( f: U \to \mathbb{R} \) is a function in open subset \( U \subset \mathbb{R}^2 \), if the partial derivatives \( \partial_1 f(x), \partial_2 f(x), \partial_1 \partial_2 f(x), \partial_2 \partial_1 f(x) \) exist at all points \( x \in U \), and if the mixed partial derivatives \( \partial_1 \partial_2 f(x), \partial_2 \partial_1 f(x) \) are continuous at a point \( a \in U \), then \( \partial_1 \partial_2 f(a) = \partial_2 \partial_1 f(a) \).
Proof: By an application of the mean value theorem for functions of one variable it follows that
\[ \Delta = \left( f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) \right) - \left( f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) \right) \]
\[ = \phi(a_1 + h_1) - \phi(a_1) \quad \text{where } \phi(x_1) = f(x_1, a_2 + h_2) - f(x_1, a_2) \]
\[ = \phi'(a_1) h_1 \quad \text{where } a_1 \text{ is between } a_1 \text{ and } a_1 + h_1 \]
if \( \| h \| \) is sufficiently small, and by another application of the mean value theorem for functions of one variable it further follows that
\[ \phi'(a_1) = \partial_1 f(a_1, a_2 + h_2) - \partial_1 f(a_1, a_2) \]
\[ = \partial_2 \partial_1 f(a_1, a_2 + h_2) h_2 \quad \text{where } a_2 \text{ is between } a_2 \text{ and } a_2 + h_2 \]
if \( \| h \| \) is sufficiently small; consequently
\[ \Delta = \partial_2 \partial_1 f(a_1, a_2) h_2 h_1. \]
On the other hand it is possible to group the terms in the equation for \( \Delta \) in another way, so that
\[ \Delta = \left( f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2 + h_2) \right) - \left( f(a_1 + h_1, a_2 + h_2) - f(a_1, a_2) \right). \]
The same argument used for the first grouping when applied to this grouping amounts to interchanging the roles of the two variables, which leads to the result that
\[ \Delta = \partial_1 \partial_2 f(\beta_1, \beta_2) h_1 h_2 \]
for some points \( \beta_1 \) between \( a_1 \) and \( a_1 + h_1 \) and \( \beta_2 \) between \( a_2 \) and \( a_2 + h_2 \) if \( \| h \| \) is sufficiently small; the points \( \beta_1 \) and \( \beta_2 \) are not necessarily the same as the points \( a_1 \) and \( a_2 \) since they are derived by different applications of the mean value theorem in one variable. Comparing the two expressions for \( \Delta \) and dividing by \( h_1 h_2 \) shows that
\[ \partial_2 \partial_1 f(a_1, a_2) = \partial_1 \partial_2 f(\beta_1, \beta_2). \]
Since the functions \( \partial_1 \partial_2 f(x_1, x_2) \) and \( \partial_2 \partial_1 f(x_1, x_2) \) are both continuous and \( \alpha_1 \) and \( \beta_1 \) both approach \( a_1 \) as \( h_1 \) tends to 0, and correspondingly for the points \( a_2 \) and \( \beta_2 \), if follows in the limit that \( \partial_2 \partial_1 f(a_1, a_2) = \partial_1 \partial_2 f(a_1, a_2) \), which suffices to conclude the proof.

The preceding result shows that so long as the partial derivatives exist and are continuous functions the order in which partial derivatives are taken is immaterial; consequently the notation can be simplified, for example by writing \( \partial_1^2 \partial_2 f(x) \) in place of \( \partial_1 \partial_2 \partial_1 f(x) \) or \( \partial_2 \partial_1 \partial_1 f(x) \) or by writing \( \partial_1 \partial_2 f(x) \) in place of \( \partial_1 \partial_1 f(x) \). Sometimes the notation is simplified further by using the multi-index notation, in which \( \partial^I f(x) \) for \( I = (3, 2, 1) \) stands for \( \partial_1^2 \partial_2 \partial_1 f(x) \) for instance. Another notation frequently used is
\[ \frac{\partial^3 f(x)}{\partial x_1^2 \partial x_2} = \partial_1^2 \partial_2 f(x). \]
The definition (2.1) of differentiability involved an approximation of a function by a polynomial of degree 1 for sufficiently small values of the auxiliary variable \( h \). The existence of higher order derivatives can be interpreted correspondingly as involving an approximation of a function by polynomials of higher degree for sufficiently small values of the auxiliary variable \( h \). Since this topic is sometimes omitted in treatments of functions of a single variable the discussion here will begin with the results for functions of one variable, from which the general result follows readily.
2.3. **HIGHER DERIVATIVES**

**Theorem 2.5 (Taylor expansion in one variable)** If $f : U \rightarrow \mathbb{R}$ has all partial derivatives up to order $k + 1$ in an open neighborhood $U \subset \mathbb{R}$ of a point $a \in \mathbb{R}$ then for any $h \in \mathbb{R}$ sufficiently small

$$f(a + h) = f(a) + f'(a)h + \frac{1}{2!} f''(a)h^2 + \frac{1}{3!} f'''(a)h^3 + \ldots$$

$$\ldots + \frac{1}{k!} f^{(k)}(a)h^k + \frac{1}{(k+1)!} f^{(k+1)}(a)h^{k+1}$$

where $\alpha$ is between $a$ and $a + h$.

**Proof:** For a fixed $h \in \mathbb{R}$ sufficiently small that the closed interval from $a$ to $a + h$ is contained in $U$ let

$$R(x) = f(a + x) - f(a) - f'(a)x - \frac{1}{2!} f''(a)x^2 - \frac{1}{3!} f'''(a)x^3 \ldots$$

$$\ldots - \frac{1}{k!} f^{(k)}(a)x^k - \frac{1}{(k+1)!} cx^{k+1}$$

where $c \in \mathbb{R}$ is chosen so that $R(h) = 0$. Clearly $R(0) = 0$ as well, so it follows from an application of the mean value theorem for functions of a single variable that $R'(\alpha_1) = 0$ for some value $\alpha_1$ between 0 and $h$. Note that

$$R'(x) = f'(a + x) - f'(a) - f''(a)x - \frac{1}{2!} f'''(a)x^2 -$$

$$\ldots - \frac{1}{(k-1)!} f^{(k-1)}(a)x^{k-1} - \frac{1}{k!} cx^k,$$

which is much the same as (2.10) but for the index $k - 1$ in place of the index $k$. In particular $R'(0) = 0$, and since $R'(\alpha_1) = 0$ it follows from another application of the mean value theorem for functions of a single variable that $R''(\alpha_2) = 0$ for some $\alpha_2$ between 0 and $\alpha_1$. The process continues, with further differentiation of the expression (2.10), yielding a sequence of points $\{\alpha_1, \alpha_2, \ldots, \alpha_{k+1}\}$ where $\alpha_{k+1}$ is between 0 and $\alpha_i$, hence is between 0 and $h$, and $R^{(i)}(\alpha_i) = 0$. Finally for the case $i = k + 1$ it follows that

$$R^{(k+1)}(x) = f^{(k+1)}(a + x) - cx$$

and since $R^{(k+1)}(\alpha_{k+1}) = 0$ it follows that $c = f^{(k+1)}(a + \alpha_{k+1})$. That suffices for the proof.

The corresponding result in several variables can be deduced from the result in a single variable by an application of the chain rule.

**Theorem 2.6 (Taylor expansion in several variables)** If $f : U \rightarrow \mathbb{R}$ has continuous partial derivatives up to order $k + 1$ in an open neighborhood $U \subset \mathbb{R}^m$ of a point $a \in \mathbb{R}^m$ then for any $h = \{h_j\} \in \mathbb{R}^m$ sufficiently small

$$f(a + h) = f(a) + \sum_{j=1}^{m} \partial_j f(a)h_j + \frac{1}{2!} \sum_{j_1,j_2=1}^{m} \partial_{j_1}f(a)h_{j_1}h_{j_2} + \ldots$$

$$\ldots + \frac{1}{k!} \sum_{j_1,\ldots,j_k=1}^{m} \partial_{j_1}\ldots h_{j_k} + \frac{1}{(k+1)!} \sum_{j_1,\ldots,j_{k+1}=1}^{m} \partial_{j_1}\ldots h_{j_{k+1}}$$
where \( \alpha \) is between \( a \) and \( a + h \) on the line segment connecting them.

**Proof:** Let \( \phi(t) = a + th \) for any \( t \in \mathbb{R} \) and consider the function \( g(t) = f(\phi(t)) \), for which \( g(0) = f(\phi(0)) = f(a) \) and \( g(1) = f(\phi(1)) = f(a + h) \). By the Taylor expansion (2.9) of the function \( g(t) \)

\[
g(1) = g(0) + g'(0) + \frac{1}{2!}g''(0) + \ldots + \frac{1}{k!}g^{(k)}(0) + \frac{1}{(k+1)!}g^{(k+1)}(\tau)
\]

for some \( \tau \in (0, 1) \). By repeated applications of the chain rule

\[
g'(t) = \sum_{j_1=1}^{m} \partial_{j_1} f(a + th) h_{j_1},
\]

\[
g''(t) = \sum_{j_1, j_2=1}^{m} \partial_{j_1 j_2} f(a + th) h_{j_1} h_{j_2}
\]

and in general

\[
g^{(\nu)}(t) = \sum_{j_1, j_2, \ldots, j_{\nu}=1}^{m} \partial_{j_1 j_2 \ldots j_{\nu}} f(a + th) h_{j_1} h_{j_2} \cdots h_{j_{\nu}};
\]

substituting these values into (2.13) for \( t = 0 \) yields (2.12) for \( \alpha = a + \tau h \) and thereby concludes the proof.

As a consequence, if a function \( f(x) \) is twice differentiable near a point \( a \in \mathbb{R}^m \) and if \( h \in \mathbb{R}^m \) is sufficiently small then

\[
f(a + h) = f(a) + \sum_{j=1}^{m} \partial_j f(a) h_j + \sum_{j_1, j_2=1}^{m} \partial_{j_1 j_2} f(a) h_{j_1} h_{j_2}
\]

for some point \( \alpha \) between \( a \) and \( h \) on the line segment connecting them. The expression

\[
\sum_{j_1, j_2=1}^{m} \partial_{j_1 j_2} f(a) h_{j_1} h_{j_2}
\]

is a quadratic form in the variables \( h_j \) defined by the matrix

\[
\partial_{j_1 j_2} f(a);
\]

this matrix, which by Theorem 2.4 is a symmetric matrix, is called the **Hessian** of the function \( f(x) \) at the point \( a \). The last term in the Taylor expansions (2.9) and (2.12) is called Lagrange’s form of the remainder; there are various alternative forms for the last term that are often used, but they will not be needed in the discussion here.

### 2.4 Functions

The special case of mappings \( f : U \to \mathbb{R}^1 \) from open subsets \( U \subset \mathbb{R}^m \) into \( \mathbb{R}^1 \), the case of ordinary real-valued functions of \( m \) variables, is particularly important; and there are various special properties and traditional notation for this special case that should be considered in more detail. The derivative \( f'(x) \) of such a function is a \( 1 \times m \) matrix, so it is not a vector in the sense used for points \( x \in \mathbb{R}^m \), which are always viewed as column vectors; but the transpose \( f'(x) \) is a column vector, so it can be viewed as an ordinary vector in \( \mathbb{R}^m \). To avoid confusion and keep these distinctions clear the transpose of the derivative is often denoted by \( \nabla f(x) \), although an earlier but still used alternative notation is \( \text{grad} f(x) \); and this vector is called the **gradient**
of the function \( f(x) \). Thus for a function \( f(x_1, x_2) \) of two variables for instance

\[
f'(x) = \left( \partial_1 f(x) \quad \partial_2 f(x) \right) \quad \text{and} \quad \nabla f(x) = \text{grad} f(x) = \left( \frac{\partial f(x)}{\partial x_1} \right).
\]

Although this is really a rather trivial point, the convention that the gradient of a function defined in a subset of a vector space \( \mathbb{R}^m \) is a vector in the same sense as all other vectors in \( \mathbb{R}^m \) is sufficiently standard and convenient that it should become familiar; thus \( \nabla f(x) \) is a vector that can be used in the same context as the vector \( x \), either in the addition of vectors or in the dot or inner product of two vectors.

As a first example of this usage, a straight line through a point \( a \in \mathbb{R}^m \) in the direction of a unit vector \( u \) can be described parametrically as the set of points \( x = \phi(t) \) for \( t \in \mathbb{R} \), where \( \phi : \mathbb{R} \to \mathbb{R}^m \) is the mapping \( \phi(t) = a + tu \). If \( f : U \to \mathbb{R} \) is a differentiable function in an open set \( U \subset \mathbb{R}^m \) containing the point \( a \) the restriction of \( f \) to this straight line can be viewed as a function \( (f \circ \phi)(t) = f(a+tu) \) of the parameter \( t \in \mathbb{R} \) near the origin. The derivative of this restriction is called the directional derivative of the function \( f \) at the point \( a \) in the direction of the unit vector \( u \), and is denoted by \( \partial_u f(a) \). It follows from the chain rule that \( \partial_u f(a) = f'(a)u' \) if \( f' \) is defined in an open set \( U \subset \mathbb{R}^m \) containing the point \( a \), so it follows from the one-variable result that \( \partial f(a) = 0 \); and since that is the case for all indices \( j \) it follows that \( f'(a) = \{0, \ldots, 0\} \) so the point \( a \) is a critical point. Not all critical points are local extrema though; the origin is a critical

2.4. FUNCTIONS

23

...
point of the function $f(x) = x^3$ of a single variable but is neither a local maximum or a local minimum. The process of finding maxima or minima of functions defined in an open subset $U \subset \mathbb{R}^m$ begins with the identification of the critical points of the function, which are the candidates for at least local extrema. In many cases it is fairly easy to tell by close inspection whether a local extremum is a local maximum or local minimum; and if that fails there is an analogue for functions of several variables of the second-derivative test familiar from the examination of extrema for functions of a single variable. If $f$ is a twice continuously differentiable function with a critical point $a \in \mathbb{R}^m$ then since $f'(a) = 0$ the Taylor expansion (2.14) takes the form

$$f(a + h) = f(a) + \sum_{j_1,j_2=1}^{m} \partial_{j_1j_2} f(a) h_{j_1} h_{j_2}$$

where $a$ is a point between $a$ and $a + h$ on the line joining these two points. If the Hessian $\partial_{j_1j_2} f(a)$ is positive definite then by continuity the Hessian $\partial_{j_1j_2} f(a)$ is positive definite for all points $a \in U$ for an open neighborhood $U$ of $a$, so that

$$\sum_{j_1,j_2=1}^{m} \partial_{j_1j_2} f(a) h_{j_1} h_{j_2} \geq 0 \quad \text{for all} \quad a \in U;$$

consequently $f(a + h) \geq f(a)$ so $a$ is a local minimum of the function $f$. Similarly if the Hessian $\partial_{j_1j_2} f(a)$ is negative definite then $a$ is a local maximum. If the Hessian is neither positive definite nor negative definite the second derivative test generally provides no information; but if the Hessian has some strictly positive and some strictly negative eigenvalues then $a$ is neither a local maximum nor local minimum, but is a point called a saddle point of the function. Geometrically the graph of a function of two variables near such a point looks something like an ordinary saddle; it is concave upwards in one direction and concave downwards in another direction. The difficulty lies in determining whether the Hessian is positive or negative definite. The simplest general result is the theorem that a symmetric matrix is positive definite if and only if the principal minors (the determinants of all submatrices formed from the first $k$ rows and columns for $1 \leq k \leq n$) are all strictly positive, and the matrix is negative definite if and only if all the principal minors are all strictly negative. For the case of functions in $\mathbb{R}^m$ for sufficiently small $m$ this is fairly easy to apply: for example if $m = 2$ the principal minors of a matrix $M = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix}$ are $m_{11}$ and $m_{11}m_{22} - m_{12}m_{21}$, so it is fairly easy to see whether the critical point is a local maximum or local minimum although it does involve calculating some second derivatives. It is sometimes tempting to think that a function $f$ has a local minimum at $a \in \mathbb{R}^m$ if $\partial^2_{jj} f(a) \geq 0$ for $1 \leq j \leq m$; but that is not necessarily the case, since for example the matrix $\begin{pmatrix} 1 & 2 \\ 1 & -2 \end{pmatrix}$ has strictly positive diagonal entries but is not positive definite since the determinants of the two principal minors are $1$ and $-3$.

The condition that a function $f$ is differentiable at a point $a$ can be viewed as the condition that $f$ can be approximated near the point $a$ by an affine function, a polynomial of degree 1 in the variables in $\mathbb{R}^m$, and that the error in this approximation is fairly small; this interpretation is often quite useful in practice. It is often expressed as the condition that for small changes $\Delta x$ in the coordinates of a point in $\mathbb{R}^m$ the change $\Delta f$ in the value of the function is approximately a linear function of $\Delta x$; for (2.1) can be written

$$\Delta f(x) = f(x + \Delta x) - f(x) = f'(x)\Delta x + \epsilon(\Delta x),$$

(2.16)
2.4. FUNCTIONS

so the change $\Delta f(x)$ in the value of the function $f(x)$ is approximately equal to the linear function $f'(x)\Delta x$ of the change $\Delta x$ in the variable $x$, and the error $\epsilon(\Delta x)$ is much smaller than the change $\Delta x$ in the variable since $\lim_{\Delta x \to 0} \epsilon(\Delta x)/\|\Delta x\| = 0$. If the function $f(x)$ is twice continuously differentiable in an open neighborhood $U$ of the point $x$ and if $M = \sup_{t \in U} \|\partial_{j_1j_2}f(t)\|_\infty$ is a bound on the Hessian of the function $f(t)$ for $t \in U$ then

$$\left| \sum_{j_1,j_2=1}^m \partial_{j_1j_2}f(t)h_{j_1}h_{j_2} \right| \leq Mm^2\|h\|_\infty^2$$

for all points $t \in U$, and it follows from (2.14) that

$$(2.17) \quad \left| \epsilon(\Delta x) \right| \leq Mm^2\|\Delta x\|_\infty^2;$$

this is a better and sometimes more useful estimate for the size of the error term in (2.1). In practice it is often convenient to have a quick approximation for the values of a function $f$ in a neighborhood of a point when the value of the function at that point is known.

**Theorem 2.7 (The Mean Value Theorem)** If $f : U \to \mathbb{R}$ is a differentiable function in an open set $U \subset \mathbb{R}^m$ and if two points $a, b$ and the line $\lambda$ joining them lie in $U$ then

$$(2.18) \quad f(b) - f(a) = \nabla f(c) \cdot (b - a)$$

for some point $c$ between $a$ and $b$ on the line $\lambda$ joining these two points.

**Proof:** The function $g(t) = f(a + t(b - a))$ is a differentiable function of the variable $t$ in an open neighborhood of the interval $[0, 1]$, and by the mean value theorem for functions of one variable $g(1) - g(0) = g'(\tau)$ for a point $\tau \in (0, 1)$. Since $g(1) = f(b)$ and $g(0) = f(a)$, while by the chain rule

$$g'(t) = \sum_{j=1}^m \partial_j f(a + t(b - a))(b_j - a_j),$$

it follows that

$$f(b) - f(a) = g(1) - g(0) = g'(\tau) \cdot 1 = \sum_{j=1}^m \partial_j f(a + \tau(b - a))(b_j - a_j)$$

$$= \sum_{j=1}^m \partial_j f(c)(b_j - a_j) = \nabla f(c) \cdot (b - a)$$

where $c = a + \tau(a - b)$, and that suffices for the proof.

The mean value theorem does not extend directly to mappings $f : U \to \mathbb{R}^n$ for $n > 1$; for although the preceding theorem can be applied to each coordinate function of the mapping $f$, the points at which the derivatives of the different coordinate functions are evaluated may be different. However for some purposes an estimate is useful enough.
Corollary 2.2 (The Mean Value Inequality) If \( f : U \rightarrow \mathbb{R}^n \) is a differentiable mapping in an open set \( U \subset \mathbb{R}^m \) and if two points \( a, b \) and the line \( \lambda \) joining them lie in \( U \) then

\[
\|f(b) - f(a)\|_{\infty} \leq n\|f'(c)\|_{\infty}\|b - a\|_{\infty}
\]

(2.19) for some point \( c \) between \( a \) and \( b \) on the line \( \lambda \).

Proof: For any unit vector \( u \in \mathbb{R}^n \) the dot product \( f_u(x) = u \cdot f(x) \) is a real-valued function to which the Mean Value Theorem can be applied, so

\[
f_u(b) - f_u(a) = \nabla f_u(c) \cdot (b - a) = \sum_{j=1}^{m} \partial_j u \cdot f(c) (b_j - a_j)
\]

\[
= \sum_{i=1}^{n} \sum_{j=1}^{m} u_i \partial_j f_i(c) (b_j - a_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} u_i \cdot (f'(c))_{ij} (b_j - a_j)
\]

\[
= u \cdot (f'(c)(b - a))
\]

for some point \( c \) between \( a \) and \( b \) on the line \( \lambda \) joining these two points. If the unit vector \( u \) is chosen to lie in the direction of the vector \( f(b) - f(a) \) then \( |f_u(b) - f_u(a)| = |u \cdot (f(b) - f(a))| = ||f(b) - f(a)||_2 \) so \( ||(b) - (a)||_2 = u \cdot (bf'(c) \cdot (b - a)) \); and by the Cauchy-Schwarz inequality and (1.9)

\[
u \cdot f'(c)(b - a) \leq ||f'(c)(b - a)||_2 \leq n \|f'(c)\|_{\infty} \|b - a\|_{\infty},
\]

which suffices for the proof.

Corollary 2.3 If \( f : U \rightarrow \mathbb{R}^n \) is a differentiable mapping in a cell \( \Delta \subset \mathbb{R}^m \) such that \( \|f'(x)\|_{\infty} \leq M \) for all \( x \in U \) then the mapping \( f \) is uniformly continuous in \( \Delta \).

Proof: If \( ||f'(x)||_{\infty} \leq M \) for all \( x \in \Delta \) then for any two points \( x, y \in \Delta \) the line joining them is contained in \( \Delta \) so it follows from the Mean Value Inequality that

\[
||f(x) - f(y)||_{\infty} \leq Mn\|x - y\|_{\infty},
\]

which shows that the mapping \( f \) is uniformly continuous in \( \Delta \) and thereby concludes the proof.

One way in which to avoid the question whether a vector is a row vector or column vector is to write the vector as an explicit linear combination of basis vectors in the vector space \( \mathbb{R}^m \). If \( f(x) = x_j \) then

\[
f'(x) = (0 \ \cdots \ 0 \ 1 \ \cdots \ 0),
\]

where the entry 1 is in column \( j \); thus \( f'(x) \) is independent of \( x \) and actually is one of the standard basis vectors for the vector space \( \mathbb{R}^m \). It is tempting to avoid introducing a separate notation and to denote this function just by \( x_j \). The standard notation for its derivative then would be \( x'_j \), which is somewhat confusing since \( x'_j \) commonly is used to denote another set of variables in \( \mathbb{R}^n \); it is clearer and rather more customary to denote the derivative of the function \( x_j \) by \( dx_j \), so that

\[
dx_j = (0 \ \cdots \ 0 \ 1 \ 0 \cdots 0)
\]

(2.20)
with the entry 1 in column \( j \) and all other entries 0. With this notation the vector \( f'_j(x) \) for an arbitrary differentiable function \( f \) can be denoted correspondingly by \( df(x) \) and written in terms of the basis (2.20) as

\[
(2.21) \quad df(x) = \sum_{j=1}^{m} \partial_j f(x) \, dx_j.
\]

This form for the derivative \( f'(x) \) is called a differential form, or to be more explicit, a differential form of degree 1. More generally a mapping \( f : U \rightarrow \mathbb{R}^m \) when viewed as associating to a point \( x \in \mathbb{R}^m \) a vector also of dimension \( m \) is called a vector field on the subset \( U \subset \mathbb{R}^m \); this is familiar from physics, for the electric, magnetic and gravitational fields. A vector field \( f \) in an open subset \( U \subset \mathbb{R}^m \) with the coordinate functions \( f_j(x) \) also can be written as a linear combination of the standard basis vectors in \( \mathbb{R}^n \) and consequently as a differential form of degree 1, explicitly as

\[
(2.22) \quad \omega_f(x) = \sum_{j=1}^{m} f_j(x) \, dx_j.
\]

The discussion of differential forms will be taken up again in Chapter 5.
CHAPTER 2. DIFFERENTIABLE MAPPINGS
Chapter 3

The Rank Theorem

3.1 The Inverse Mapping Theorem

A continuous one-to-one mapping between two open subsets of \( \mathbb{R}^n \) with a continuous inverse, a mapping called a **homeomorphism** between these two sets, really identifies the two sets as far as topological properties are concerned. If the mapping and its inverse are both continuously differentiable, in which case the mapping is called a \( C^1 \) homeomorphism or sometimes a \( C^1 \) diffeomorphism, this identification also extends to properties of the derivatives of functions; and of course it is also possible to consider \( C^r \) homeomorphisms for any \( r > 0 \) to identify properties of the higher derivatives of functions on these two sets. It is a familiar result from calculus in one variable that if \( f : U \rightarrow \mathbb{R} \) is a continuously differentiable function in an open interval \( U \subset \mathbb{R} \) such that \( f'(x) \neq 0 \) for all \( x \in U \) then the mapping \( f \) is a \( C^1 \) homeomorphism from \( U \) to its image \( f(U) \). At least the local version of this result extends to mappings between open subsets of \( \mathbb{R}^n \) for \( n > 1 \). If \( f : U \rightarrow V \) is a \( C^1 \) homeomorphism between two open subsets \( U, V \subset \mathbb{R}^m \) and if \( g : V \rightarrow U \) is its inverse then \( g(f(x)) = x \) and an application of the chain rule as on page 17 shows that \( g'(f(x))f'(x) = 1 \), the identity matrix, and consequently that \( \det f'(x) \neq 0 \) at each point \( x \in U \). What replaces the condition that \( f'(a) \neq 0 \) for functions of one variable is the condition that \( \det f'(a) \neq 0 \) for mappings in several variables; the matrix \( f' \) is called the Jacobian matrix of the mapping \( f \), and its determinant is called the Jacobian of the mapping \( f \). It is convenient first to establish a special case of the result for mappings in several variables, from which the general case follows quite simply.

**Theorem 3.1** If \( f : W \rightarrow \mathbb{R}^n \) is a continuously differentiable mapping defined in an open neighborhood \( W \subset \mathbb{R}^n \) of the origin such that \( f(0) = 0 \) and \( f'(0) = I \), the identity matrix, then for any sufficiently small open subneighborhood \( U \subset W \) of the origin the restriction of the mapping \( f \) to \( U \) is a \( C^1 \) homeomorphism \( f : U \rightarrow V \) between \( U \) and an open neighborhood \( V \subset \mathbb{R}^n \) of the origin.

**Proof:** The mapping \( r(x) = f(x) - x \) is a continuously differentiable mapping such that \( r(0) = 0 \) and \( r'(0) = 0 \); therefore there is a closed cell \( \Delta \subset W \subset \mathbb{R}^n \) centered at the origin \( 0 \) sufficiently small that the following conditions are satisfied:

(i) \( \|r'(|x|)\| \leq \frac{1}{2n} \) for all \( x \in \Delta \),

(ii) \( \det f'(x) \neq 0 \) for all \( x \in \Delta \). The first step in the proof it to establish a basic
inequality on which the remainder of the proof rests. From the Mean Value Inequality (2.19) and (i) it follows that for any points \( a, b \in \Delta \) there is a point \( c \) on the line connecting those two points for which

\[
\|r(b) - r(a)\|_{\infty} \leq n \|r'(c)\|_{\infty} \|b - a\|_{\infty} \leq \frac{1}{2\sqrt{n}} \|b - a\|_{\infty};
\]

so by (1.7)

\[
\|r(b) - r(a)\|_{2} \leq \|r(b) - r(a)\|_{\infty} \leq \frac{1}{2\sqrt{n}} \|b - a\|_{\infty} \leq \frac{1}{2} \|b - a\|_{2}.
\]

Then

\[
\|f(b) - f(a)\|_{2} - \|b - a\|_{2} = \|f(b) - f(a)\|_{2} - \|b - a\|_{2} = \|r(b) - r(a)\|_{2} \leq \frac{1}{2} \|b - a\|_{2};
\]

and from this observation and the triangle inequality it follows that

\[
\|b - a\|_{2} \leq \|f(b) - f(a)\|_{2} + \|b - a\|_{2} \leq \frac{1}{2} \|b - a\|_{2} + \|f(b) - f(a)\|_{2}
\]

and consequently that

\[
(3.1) \quad \|f(b) - f(a)\|_{2} \geq \frac{1}{2} \|b - a\|_{2},
\]

which is the basic inequality.

It follows from (3.1) that the mapping \( f : \Delta \to \mathbb{R}^{n} \) is an injective mapping; for if \( f(b) = f(a) \) that inequality shows that \( \|b - a\|_{2} = 0 \). In particular \( f(x) \neq f(0) = 0 \) for any point \( x \in \partial \Delta \), the boundary of the cell \( \Delta \); so since \( \partial \Delta \) is compact there is a positive number \( d > 0 \) such that

\[
\|f(x)\|_{2} \geq d > 0 \quad \text{for all } x \in \partial \Delta.
\]

The open neighborhood \( V \) of the origin defined by

\[
V = \{ y \in \mathbb{R}^{n} \mid \|y\|_{2} < \frac{d}{2} \}
\]

then is contained in the image \( f(\Delta) \). To demonstrate that, for any point \( b \in V \) the function \( \psi(x) = \|f(x) - b\|_{2} \) is a continuous function on \( \Delta \), and it is just necessary to show that \( \psi(a) = 0 \) for some point \( a \in \Delta \). This function satisfies \( \psi(0) = \|b\|_{2} < \frac{d}{2} \) since \( b \in V \), while if \( x \in \partial \Delta \)

\[
d \leq \|f(x)\|_{2} \leq \psi(x) + \|b\|_{2} < \psi(x) + \frac{d}{2}
\]

so \( \psi(x) > \frac{d}{2} \). Since \( \psi(0) < \psi(x) \) for any point \( x \in \partial \Delta \), the function \( \psi(x) \) must take its minimum value in the closed set \( \Delta \) at some interior point \( a \in \Delta \), which is the obvious candidate to be a zero of the function \( \psi(x) \). The square of this function, the function \( \psi(x)^{2} = \sum_{i=1}^{n} (f_{i}(x) - b_{i})^{2} \), then also takes its minimum value at the point \( a \in \Delta \), and hence that point is a critical point of \( \psi(x)^{2} \); a zero of all of its partial derivatives \( \partial \psi(x)^{2} = \sum_{i=1}^{n} 2(f_{i}(x) - b_{i}) \partial f_{i}(x) \); thus

\[
0 = \sum_{i=1}^{n} 2(f_{i}(a) - b_{i}) \partial f_{i}(a).
\]

However it
3.1. THE INVERSE MAPPING THEOREM

follows from (ii) that \( f'(a) \) is a nonsingular matrix, so the only row vector \( v \) such that \( v f'(a) = 0 \) is the zero vector; hence \( f_i(a) - b_i = 0 \) for all indices \( i \) so \( \psi(a) = 0 \).

What has been shown so far is that the mapping \( f : \Delta \to \mathbb{R}^n \) is an injective mapping, and that \( V \subset f(\Delta) \). Consequently the subset \( U = f^{-1}(V) \subset \Delta \) is an open neighborhood of the point \( a \), and the restriction of the mapping \( f \) to the open set \( U \) is a one-to-one mapping \( f : U \to V \) from \( U \) onto \( V \); so there is a well defined inverse mapping \( g = f^{-1} : V \to U \) from \( V \) onto \( U \). To show that this mapping \( g \) is continuously differentiable, first if \( a, a + h \in V \) and \( g(a) = x, g(a + h) = x + k \) then \( f(x) = a \) and \( f(x + k) = a + h \). By the basic inequality (3.1)

\[
||g(a + h) - g(a)||_2 = ||(x + k) - x||_2 \leq 2||f(x + k) - f(x)||_2 = 2||(a + h) - a||_2 = 2||h||_2,
\]

so that \( g \) is continuous. Since \( f \) is differentiable

\[
h = f(x + k) - f(x) = f'(x)k + \epsilon_f(k)
\]
and that can be rewritten

(3.2)

\[
g(a + h) - g(a) = k = f'(x)^{-1}h + \epsilon_g(h)
\]

where \( \epsilon_g(h) = -f'(x)^{-1}\epsilon_f(k) \). It follows from (1.9) that \( ||f'(x)^{-1}h||_2 \leq A||h||_2 \) where \( A = n||f'(x)^{-1}||_\infty \), or by (1.7) equivalently \( ||f'(x)^{-1}h||_2 \leq B||h||_2 \) where \( B = \sqrt{n}A \), and similarly \( ||\epsilon_g(h)||_2 = || - f'(x)^{-1}\epsilon_f(k)||_2 \leq B||\epsilon_f(k)||_2 \); altogether then

(3.3)

\[
||\epsilon_g(h)||_2 \leq B \frac{||\epsilon_f(k)||_2}{||h||_2} \leq 2B \frac{||\epsilon_f(k)||_2}{||h||_2} \leq 2
\]

since by the fundamental inequality (3.1) yet again

\[
\frac{||k||_2}{||h||_2} = \frac{||(x + k) - x||_2}{||f(x + k) - f(x)||_2} \leq 2.
\]

Since \( \lim_{k \to 0} \frac{||\epsilon_f(k)||_2}{||h||_2} = 0 \) while \( \lim_{h \to 0} k = 0 \) as a consequence of the continuity of the mapping \( g \); it follows from (3.3) that \( \lim_{h \to 0} \frac{||\epsilon_g(h)||_2}{||h||_2} = 0 \) and consequently from (3.2) that the mapping \( g \) is differentiable, and moreover that \( g'(a) = f'(x)^{-1} \) so that the derivative \( g'(a) \) is also continuous; that suffices for the proof.

Corollary 3.1 (Inverse Mapping Theorem) If \( f : W \to \mathbb{R}^n \) is a continuously differentiable mapping in an open neighborhood \( W \subset \mathbb{R}^n \) of a point \( a \in \mathbb{R}^n \) such that \( \det f'(a) \neq 0 \) then for any sufficiently small open subneighborhood \( U \subset W \) of the point \( a \) the restriction of the mapping \( f \) to \( U \) is a \( C^1 \) homeomorphism \( f : U \to V \) between \( U \) and an open neighborhood \( V \) of the image point \( b = f(a) \). If \( g : V \to U \) is the inverse mapping then \( g'(y) = f'(g(y))^{-1} \) for each point \( y \in V \).

Proof: Introduce the invertible affine mappings \( \phi, \psi : \mathbb{R}^n \to \mathbb{R}^n \) defined by

\[
\phi(x) = f'(a)^{-1}x + a, \quad \psi(x) = x - b.
\]

The composition \( F = \psi \circ f \circ \phi \) is then a continuously differentiable mapping from an open neighborhood of the origin \( 0 \in \mathbb{R}^n \) into \( \mathbb{R}^n \) such that \( F(0) = 0 \), and by the chain rule \( F'(0) = \phi'(b)f'(a)\phi'(0) = I \cdot f'(a)f'(a)^{-1} = I \). The preceding theorem shows that the mapping \( F \) is invertible, and that its inverse \( G \) is a continuously differentiable
mapping from an open neighborhood of the origin into $\mathbb{R}^n$. The composition $g = \phi \circ G \circ \psi$ is a continuously differentiable mapping from an open neighborhood of $b$ into $\mathbb{R}^n$, and since $f = \psi^{-1} \circ G \circ \phi^{-1}$ it follows that $g \circ f = \phi \circ G \circ \psi \circ \phi^{-1} \circ F \circ \phi^{-1}$ is the identity mapping so that $g$ is the inverse mapping to $f$. Finally since $y = f(g(y))$ for any point $y \in V$ it follows from the chain rule that $I = f'(g(y))g'(y)$ where $I$ is the identity matrix, so that $g'(y) = f'(g(y))^{-1}$, and that concludes the proof.

One consequence of the Inverse Mapping Theorem is that if $f : U \rightarrow \mathbb{R}^n$ is a continuously differentiable mapping defined in an open subset $U \subset \mathbb{R}^n$ such that $\det f'(x) \neq 0$ for all points $x \in U$ then $f$ is an open mapping, in the sense that the image of an open subset $U_0 \subset U$ is an open subset $f(U_0) \subset \mathbb{R}^n$. The inverse image of any open set under a continuous mapping is always open; but the image of an open subset under a general continuous mapping is not necessarily open, so open continuous mappings are a special subclass of continuous mappings. If $f : U \rightarrow \mathbb{R}^n$ is an invertible continuously differentiable mapping from an open subset $U \subset \mathbb{R}^n$ with coordinates $(t_1, \ldots, t_n)$ onto an open subset $V \subset \mathbb{R}^n$ with coordinates $(x_1, \ldots, x_n)$, points in $V$ can be described either in terms of the coordinates $(x_1, \ldots, x_n)$ in $\mathbb{R}^n$ or alternatively in terms of the coordinates $(t_1, \ldots, t_n) \in U \subset \mathbb{R}^n$; for this reason the parameters $(t_1, \ldots, t_n)$ are often described as being another local coordinate system in $V$, and the mapping $f$ is called a local change of coordinates in $V$. The advantage of a change of coordinates is that it may be possible to describe the local geometry much more simply in terms of the coordinates $(t_1, \ldots, t_n)$ than in terms of the initial coordinates $(x_1, \ldots, x_n)$. For example, an arc of a circle of radius 1 centered at the origin in the plane with a coordinate system $(x_1, x_2)$ can be described locally as a straight line segment $r = 1$ in terms of polar coordinates $(r, \theta)$. It should be emphasized though that the Inverse Mapping Theorem is only a local result; a mapping $f : U \rightarrow \mathbb{R}^n$ for which $\det f'(x) \neq 0$ at all points $x \in U \subset \mathbb{R}^n$ need not be invertible if $n > 0$. For example the derivative of the mapping $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by $f(r, \theta) = (r \cos \theta, r \sin \theta)$ is the matrix

$$f'(r, \theta) = \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}$$

for which $\det f'(r, \theta) = r$. The inverse Mapping Theorem asserts that the mapping $f$ is locally invertible near any point $(r, \theta) \in \mathbb{R}^2$ for which $r \neq 0$; but $f$ is clearly not a one-to-one mapping in the open subset $U = \{(r, \theta) | r > 0\}$ since $f(r, \theta) = f(r, \theta + 2n\pi)$ for any integer $n$. The problem of finding general necessary and sufficient conditions for a mapping to be globally invertible is a famous and difficult one$^1$. Points at the Jacobian of a differentiable mapping $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$ vanishes are known as singular.

$^1$The Jacobian Problem is the question whether a mapping $f : \mathbb{C}^n \rightarrow \mathbb{C}^n$ from the n-dimensional complex vector space to itself for $n > 1$ defined by polynomial equations such that $\det f'(x) = 1$ for all points $x \in \mathbb{C}^n$ is necessarily an invertible mapping from $\mathbb{C}^n$ onto itself. This is a long standing open problem in mathematics, having been raised by O. H. Keller in 1939. It is known to be true for polynomial mappings $f : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ of degree at most 100, but not in general, at least at the time of this writing. It is known to be false for finite fields, rather than the complex numbers. It is also false for functions more general than polynomials; the mapping $f : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ defined by $f(x, y) = (e^x, e^{-x}y)$ has the derivative $f'(x, y) = 1$ at all points $(x, y)$ but its image excludes the axis $(0, y)$ and the mapping is not one-to-one. A famous example due to Fatou and Bieberbach, and discussed for example in the book Several Complex Variables by S. Bochner and W. T. Martin, Princeton University Press, 1948, is an injective mapping $f : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ given by functions that are everywhere
3.1. THE INVERSE MAPPING THEOREM

points of the mapping. Mappings can be very complicated indeed near their singular points. For a simple example, consider again the mapping $f : \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(r, \theta) = (r \cos \theta, r \sin \theta)$ at the origin $r = \theta = 0$ where $f'(0, 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, as sketched in the accompanying Figure 3.1. A cell centered at the origin is mapped to the bow-tie region, and the mapping is locally one-to-one except along the axis $r = 0$; that entire axis is mapped to a single point, the origin. The general study of the geometry of singularities of mappings is quite extensive, including what is known as the theory of catastrophes².

Figure 3.1: The mapping $f : (r, \theta) \to (r \cos \theta, r \sin \theta)$ in an open neighborhood of the origin.

An important and sometimes somewhat confusing concept that arises in the consideration of one-to-one continuously differentiable mappings is that of orientation. If $f : U \to \mathbb{R}^n$ is a continuously differentiable mapping from an open subset $U \subset \mathbb{R}^n$ into $\mathbb{R}^n$ and if $\det f'(a) \neq 0$ at a point $a \in U$ then either $\det f'(a) > 0$, in which case the mapping $f$ is said to be orientation preserving at the point $a$, or $\det f'(a) < 0$, in which case the mapping $f$ is said to be orientation reversing near the point $a$. If $U$ is a connected open set and the Jacobian of the mapping $f$ is nowhere vanishing then the mapping $f$ is either preserving at all points of $U$ or orientation reversing at all points of $U$. In the case $n = 1$ an orientation preserving mapping is a monotonically increasing mapping, and an orientation reversing mapping is a monotonically decreasing mapping; the orientation of $\mathbb{R}^1$, identified with the set of real numbers, is the direction of increasing numbers. If $n > 1$ each coordinate axis in $\mathbb{R}^n$ is naturally oriented in the direction in which the variable increases, and the orientation of $\mathbb{R}^n$ itself is defined by the choice of an order of the coordinate axes. For example in the case $n = 2$ with the variables $x_1, x_2$ it is customary to specify an orientation as either clockwise or counterclockwise, where clockwise determines the order $x_1, x_2$ of the axes while counterclockwise determines the reversed order $x_2, x_1$ of the axes; and in the case $n = 3$ the right-hand rule familiar from physics specifies the order of the axes as that determined by the order of the thumb and first two fingers on the right hand, when these fingers are viewed as lying along three coordinate axes. When the variables are denoted by $x_1, x_2, \ldots, x_n$ the natural order is the corresponding order of the axes of these variables, first the $x_1$ axis, then the $x_2$ axis, and so on. However for variables such as $r, \theta$ in $\mathbb{R}^2$ or $\alpha, \beta, \gamma$ in $\mathbb{R}^3$ there is no natural order so it is necessary to specify an order of the axes. Two different orders though may lead to the same orientation. A

analytic and with the derivative $f'(x) = 1$ at all points $x \in \mathbb{C}^2$; but the image omits an open subset of $\mathbb{C}^2$. A survey of the problem and some approaches, and some examples of some false proofs, can be found in the paper by H. Bass, E.H. Connell, D. Wright in the Bulletin of the American Mathematical Society, volume 7, 1982. If it is just assumed that $\det f'(x) \neq 0$ for all points $x \in \mathbb{C}^2$, the “strong” Jacobian problem, there is a counterexample for real mappings $f : \mathbb{R}^2 \to \mathbb{R}^2$ by S Pinchuk in Mathematische Zeitschrift, volume 217, 1994.

permutation of the variables is a linear transformation $M$ described by a permutation matrix $M$, a matrix that has a single entry of 1 in each row and column but all other entries 0; the linear transformation described by a permutation matrix, a permutation of the coordinates, is orientation preserving if $\det M = +1$ but orientation reversing if $\det M = -1$. Thus for $n = 3$ for example the orders $x_1, x_2, x_3$ and $x_2, x_3, x_1$ and $x_3, x_1, x_2$ determine the same orientation since these orders arise from the first by repeated applications of the permutation matrix $M_+ = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ for which $\det M_+ = +1$; on the other hand the orders $x_2, x_1, x_3$ and $x_3, x_2, x_1$ and $x_1, x_3, x_2$ arise from the three previous orders by applications of the permutation matrix $M_- = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ for which $\det M_- = -1$, the permutation that switches the first two coordinates. A permutation that interchanges any two consecutive vectors is orientation reversing, since $\det\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = 1$, and any permutation can be written as a successive application of such interchanges; so beginning with any chosen order of the coordinates, those orders that arise by an even number of such interchanges beginning with the initially chosen order have the same orientation as the initially chosen order, while those that arise by an odd number of such interchanges have the opposite orientation. Thus there are always just two possible orientations of the vector space $\mathbb{R}^n$ for any $n > 0$. For example, the vector space $\mathbb{R}^2$ with coordinates $r, \theta$ can be given the orientation $r, \theta$ and the vector space $\mathbb{R}^2$ with coordinates $x, y$ can be given the orientation $x, y$. The mapping $f : \mathbb{R}^2(r, \theta) \rightarrow \mathbb{R}^2(x, y)$ that takes a point $(r, \theta)$ to the point $f(r, \theta) = (r \cos \theta, r \sin \theta)$ has the derivative $f'(r, \theta) = r$, and consequently

- if $r > 0$ then $f'(r, \theta) > 0$ and $f$ is orientation preserving,
- if $r < 0$ then $f'(r, \theta) < 0$ and $f$ is orientation reversing,
- if $r = 0$ then $f'(r, \theta) = 0$ and $f$ is singular,

as indicated in the accompanying Figure 3.2. Orientation can be described alternately in terms of the unit vectors $dx_i$ along the coordinate axes rather than in terms of the coordinate axes themselves; and the equivalence of different choices of the order of the unit vectors can be handled through a special case of the exterior algebra generated by these vectors as discussed in the Appendix. Associate to any order of the basis vectors $dx_i$ such as $dx_{i_1}, dx_{i_2}, \ldots, dx_{i_n}$ a formal vector denoted by

$$dx_{i_1} \wedge dx_{i_2} \wedge \cdots \wedge dx_{i_n},$$

with the convention that when the order of two adjacent indices is reversed this vector changes sign; thus

$$dx_{i_1} \wedge \cdots \wedge dx_{i_k} \wedge dx_{i_{k+1}} \wedge \cdots \wedge dx_{i_n} = -dx_{i_1} \wedge \cdots \wedge dx_{i_{k+1}} \wedge dx_{i_k} \wedge \cdots \wedge dx_{i_n}.$$
This vector is not to be interpreted as a vector in the space $\mathbb{R}^n$, but as a basis vector in the abstract one-dimensional vector space consisting of the products $tdx_1 \land dx_2 \land \cdots \land dx_n$ for all $t \in \mathbb{R}$. All the orders of the axes, or equivalently all the orders of the basis vectors $dx_i$, that determine the same vector (3.5) describe the same orientation of $\mathbb{R}^n$, and those that determine the negative of the vector (3.5) describe the opposite orientation; thus an orientation of $\mathbb{R}^n$ can be specified by the choice of a vector $dx_1 \land dx_2 \cdots \land dx_n$. For example, in $\mathbb{R}^2$ with the coordinates $r, \theta$ one orientation is specified by the vector $dr \land d\theta$ and the other possible orientation is specified by the vector $d\theta \land dr = -dr \land d\theta$.

Locally any $C^1$ homeomorphism can be decomposed into a composition of simpler $C^1$ homeomorphisms; that can be useful in reducing the examination of local properties of $C^1$ homeomorphisms to an examination of the properties of some simpler homeomorphisms. The permutations of the variables in $\mathbb{R}^n$ and translations in $\mathbb{R}^n$ are two particularly simple $C^1$ homeomorphisms; a somewhat less simple but useful $C^1$ homeomorphism is one that changes only a single variable, so for which the coordinate functions $f_i(x)$ are such that $f_i(x) = x_i$ for $i \neq j$ while $f_j(x)$ is a function of all the variables.

**Theorem 3.2** A $C^1$ mapping $f : U \to \mathbb{R}^n$ defined in an open neighborhood $U$ of the origin in $\mathbb{R}^n$ and such that $\det f'(0) \neq 0$ can be written in any sufficiently small open subneighborhood $V \subset U$ of the origin as the composition of $C^1$ homeomorphisms are either (i) translations, or (ii) permutations of the variables in $\mathbb{R}^n$, or (iii) mappings that change only a single variable.

**Proof:** The composition of the mapping $f$ and the translation $T y = y - f(0)$ yields a mapping $T f$ such that $T f(0) = 0$. Suppose that the coordinate functions $f_i$ of the mapping $T f$ have the property that $f_i(x) = x_i$ whenever $i < r$ for some index $r$; for convenience such a mapping will be called a mapping of type $r$, but just for the course of the present proof. Any mapping is of type 1, so there is no loss of generality in assuming that the mapping $T f$ is of type $r$ for some $r \geq 1$; but only the identity mapping is of type $r$ for any $r > n$. It is also convenient to simplify the notation in the proof by writing points $x \in \mathbb{R}^n$ in the form

$$x = \begin{pmatrix} x_I \\ x_r \\ x_{II} \end{pmatrix} \quad \text{where} \quad \begin{cases} x_I \in \mathbb{R}^{r-1} & I = (1, \ldots, r-1) \\ x_r \in \mathbb{R}^1 \\ x_{II} \in \mathbb{R}^{n-r-1} & II = (r+1, r+2, \ldots, n) \end{cases}$$

and to write the mapping $T f$ correspondingly in the form

$$T f(x) = \begin{pmatrix} x_I \\ f_I(x) \\ f_{II}(x) \end{pmatrix} \quad \text{where} \quad \begin{cases} x_I \in \mathbb{R}^{r-1} & I = (1, \ldots, r-1) \\ f_I(x) \in \mathbb{R}^1 \\ f_{II}(x) \in \mathbb{R}^{n-r-1} & II = (r+1, r+2, \ldots, n). \end{cases}$$

In these terms the derivative of the mapping $T f$ is the $n \times n$ matrix

$$(T f)'(x) = \begin{pmatrix} I_{r-1} & 0 & 0 \\ \partial_I f_I(x) & \partial_I f_{II}(x) & \partial_{II} f_I(x) \\ \partial_I f_{II}(x) & \partial_{II} f_{II}(x) & \partial_{II} f_{II}(x) \end{pmatrix}$$

in which $I_{r-1}$ is the $(r-1) \times (r-1)$ identity matrix while

$$\partial_I f_i(x) = \begin{pmatrix} \partial_1 f_i(x) & \cdots & \partial_{r-1} f_i(x) \end{pmatrix}, \quad a \ 1 \times (r-1) \text{ matrix},$$
36  

CHAPTER 3. THE RANK THEOREM

\[ \partial f_{11}(x) = \left\{ \partial_i f_i(x) \mid 1 \leq i \leq n, 1 \leq j \leq r-1, \right\} \]  

an \((n - r - 1) \times (r - 1)\) matrix,  

and similarly for the other components of the matrix (3.7). Since \( \det(TT') = \det f'(0) \neq 0 \), not all the entries in column \( r \) of this matrix can be zero; so if \( P_r \) is a permutation matrix that interchanges rows \( r \) and \( j \) for a suitable index \( r < j \leq n \) the derivative \((PTT)'(x)\) of the mapping \( P_rTf \) also has the form (3.7) but with the additional property that \( \partial_j f_r(0) \neq 0 \). Let \( \phi : U \longrightarrow \mathbb{R}^n \) be the mapping defined by

\[
\phi \left( \begin{array}{c} x_I \\ x_r \\ x_{II} \end{array} \right) = \left( \begin{array}{c} x_I \\ f_r(x) \\ x_{II} \end{array} \right) ,
\]

for which

\[
(3.8) \quad \phi(x) = \left( \begin{array}{ccc} I_{r-1} & 0 & 0 \\ \partial f_r(x) & \partial f_r(x) & \partial f_r(x) \\ 0 & 0 & I_{n-r-1} \end{array} \right) .
\]

Since \( \det \phi'(0) = \partial_j f_r(0) \neq 0 \) it follows from the Inverse Mapping Theorem that the mapping \( \phi \) is invertible in a sufficiently small open neighborhood \( U_r \subset U \) of the origin.  

The inverse mapping of course has the corresponding form

\[
\phi^{-1} \left( \begin{array}{c} y_I \\ y_r \\ y_{II} \end{array} \right) = \left( \begin{array}{c} y_I \\ g_r(y_I, y_r, y_{II}) \\ y_{II} \end{array} \right)
\]

for some function \( g_r \); and since \( \phi^{-1} \circ \phi \) is the identity mapping then in particular

\[
x_r = g_r(x_I, f_r(x), x_{II}) ,
\]

hence

\[
\phi_r^{-1}P_rTf(x) = \left( \begin{array}{c} x_I \\ g_r(x_I, f_r(x), x_{II}) \\ f_r(x) \end{array} \right) = \left( \begin{array}{c} x_I \\ x_r \\ f_r(x) \end{array} \right).
\]

Thus the mapping \( \phi_r^{-1}P_rTf \) is of type \( r + 1 \). This procedure can be repeated, and finally the composition

\[
(\phi_{n-1}^{-1}P_{n-1}) \circ (\phi_{n-2}^{-1}P_{n-2}) \circ \cdots \circ (\phi^{-1}P_rTf)
\]

is a mapping of type \( n + 1 \) so is the identity mapping; consequently

\[
f = T^{-1}P_r^{-1} \phi_r \circ \cdots \circ P_{n-1}^{-1} \phi_{n-1} \circ P_n^{-1} \phi_n ,
\]

which exhibits the mapping \( f \) in a sufficiently small neighborhood of the origin as the composition of mappings of type (i), (ii) and (iii) as in the statement of the theorem, thereby concluding the proof.

3.2 The Implicit Function Theorem

The Inverse Mapping Theorem can be applied to examine mappings between spaces of different dimensions; the general result will be discussed in Section 3.3, but a special case of particular importance and usefulness will be discussed in this section.
3.2. THE IMPLICIT FUNCTION THEOREM

Theorem 3.3 If \( f : U \rightarrow \mathbb{R}^n \) is a continuously differentiable mapping defined in an open neighborhood \( U \) of the origin \( 0 \in \mathbb{R}^m \) where \( m > n \), and if \( f(0) = 0 \) and the \( n \times n \) matrix consisting of the first \( n \) columns of the \( n \times m \) matrix \( f'(0) \) is of rank \( n \), then after shrinking the neighborhood \( U \) if necessary there is a \( C^1 \) homeomorphism \( \phi : V \rightarrow U \) between an open neighborhood \( V \) of the origin \( 0 \in \mathbb{R}^m \) and the open neighborhood \( U \), where

\[
\phi(t_1, \ldots, t_m) = \{g_1(t), \ldots, g_n(t), t_{n+1}, \ldots, t_m\}
\]

for some continuously differentiable functions \( g_i(t) \) of the variables \( t \in V \) such that \( g_i(0) = 0 \) and

\[
(f \circ \phi)(t_1, \ldots, t_m) = \{t_1, \ldots, t_n\}.
\]

Proof: The assertion of the theorem can be summarized as the existence of the mapping \( \phi \) in the diagram

\[
\begin{array}{ccc}
\mathbb{R}^n & \mathbb{R}^m & \mathbb{R}^n \\
\uparrow & \uparrow & \uparrow \\
V & \phi & U \\
\downarrow & \downarrow & \downarrow \\
& \mathbb{R}^n & \mathbb{R}^n
\end{array}
\]

with the properties (3.9) and (3.10). To simplify the notation write points \( x \in \mathbb{R}^m \) in the form

\[
x = \begin{pmatrix} x_I \\ x_{II} \end{pmatrix}
\]

where \( x_I \in \mathbb{R}^n \) \( I = (1, 2, \ldots, n) \) \( x_{II} \in \mathbb{R}^{m-n} \) \( II = (n+1, n+2, \ldots, m) \)

and write the mapping \( f \) correspondingly in the form

\[
f(x) = \begin{pmatrix} f_I(x) \\ f_{II}(x) \end{pmatrix}
\]

where \( f_I(x) \in \mathbb{R}^n \) \( I = (1, 2, \ldots, n) \) \( f_{II}(x) \in \mathbb{R}^{m-n} \) \( II = (n+1, n+2, \ldots, m) \).

The \( n \times m \) matrix \( f'(x) \) then can be written

\[
f'(x) = \begin{pmatrix} \partial_i f(x) & \partial_{II} f(x) \end{pmatrix}
\]

in terms of the \( n \times n \) matrix \( \partial_I f(x) \) having the entries \( \left( \partial_I (f(x)) \right)_{ij} = \partial_j f_i(x) \) for \( 1 \leq i, j \leq n \) and the \( n \times (m-n) \) matrix \( \partial_{II} f(x) \) with entries \( \left( \partial_{II} (f(x)) \right)_{ij} = \partial_j f_i(x) \) for \( 1 \leq i \leq n \) and \( n+1 \leq j \leq m \). Introduce the continuously differentiable mapping \( \psi : U \rightarrow \mathbb{R}^m \) defined by

\[
\psi(x) = \psi \begin{pmatrix} x_I \\ x_{II} \end{pmatrix} = \begin{pmatrix} f(x) \\ x_{II} \end{pmatrix}
\]

and note that

\[
\psi'(x) = \begin{pmatrix} \partial_i f(x) & \partial_{II} f(x) \\ \partial_{II} x_I & \partial_{II} x_{II} \end{pmatrix} = \begin{pmatrix} \partial_i f(x) & \partial_{II} f(x) \\ 0 & I_{m-n} \end{pmatrix}
\]

where \( I_{m-n} \) is the identity \( (m-n) \times (m-n) \) matrix; consequently \( \det \psi'(0) = \det \partial_I f(0) \neq 0 \), so it follows from the inverse mapping theorem that after shrinking
the neighborhood $U$ if necessary the mapping $\psi$ is a one-to-one invertible mapping from $U$ onto an open neighborhood $V$ of the image $\psi(0) = 0 \in \mathbb{R}^m$. If $\phi = \psi^{-1}$ then

$$\phi(t) = \psi\left(\begin{pmatrix} g_I(t) \\ t_{II} \end{pmatrix}\right)$$

for a mapping $g_I : \mathbb{R}^m \longrightarrow \mathbb{R}^n$ for which $g_I(0) = 0$ since $\phi(0) = 0$, which is a mapping of the form (3.9); and

$$\begin{pmatrix} t_I \\ t_{II} \end{pmatrix} = (\psi \circ \phi)\begin{pmatrix} t_I \\ t_{II} \end{pmatrix} = \psi\left(\begin{pmatrix} g_I(t) \\ t_{II} \end{pmatrix}\right) = \begin{pmatrix} f(\phi(t)) \\ t_{II} \end{pmatrix}$$

so $f(\phi(t)) = t_I$, which demonstrates (3.10) and thereby concludes the proof.

**Corollary 3.2** If $f : U \longrightarrow \mathbb{R}^n$ is a continuously differentiable mapping defined in an open subset $U \subset \mathbb{R}^m$ where $m > n$, and if $\text{rank } f'(a) = n$ at a point $a \in U$, then there is a local coordinate system in an open neighborhood of the point $a$ such that the mapping $f - f(a)$ is a linear mapping described by a matrix of rank $n$ in terms of these coordinates.

**Proof:** By a translation and permutation of the coordinates in $U$ the mapping $f - f(a)$, which takes the point $a \in U$ to the origin in $\mathbb{R}^n$, can be changed to the form of a mapping that satisfies the hypotheses of the preceding theorem; then by that theorem, after a further change of the coordinates in an open subneighborhood of the point $a$, the mapping $f - f(a)$ will be a linear mapping of the form (3.10) in the new coordinates, a linear mapping described by the the $n \times m$ matrix $(I \ 0)$ where $I$ is the $n \times n$ identity matrix. That suffices for the proof.

**Corollary 3.3** If $f : U \longrightarrow \mathbb{R}^n$ is a continuously differentiable mapping defined in an open subset $U \subset \mathbb{R}^m$ where $m > n$, and if $\text{rank } f'(a) = n$ at a point $a \in U$, then there is a local coordinate system in an open neighborhood $U_a$ of the point $a$ in terms of which the set $\{ x \in U_a \ | \ f(x) = f(a) \}$ is a linear subspace of dimension $m - n$.

**Proof:** It follows from the preceding Corollary 3.2 that there is a local coordinate system in an open neighborhood $U_a$ of the point $a$ such that the mapping $f - f(a)$ is a linear mapping of rank $n$ in terms of these coordinates; consequently the set of points in $U_a$ at which the mapping $f - f(a)$ vanishes is a linear subspace of dimension $m - n$, which suffices for the proof.

The set $V = \{ x \in U \ | \ f(x) = 0 \}$ of zeros of a mapping $f : U \longrightarrow \mathbb{R}^n$ defined in an open subset $U \subset \mathbb{R}^m$ is called the **zero locus** of the mapping $f$. If the mapping is continuous its zero locus is a relatively closed subset of $U$, since it is the inverse image of the closed set $0 \in \mathbb{R}^n$ under a continuous mapping. The preceding corollary implies that this zero locus is a particularly simple subset of $U$ if the mapping is $C^1$ and its derivative $f'(x)$ is of maximal rank at each point $x \in V$. A relatively closed subset $V \subset U$ of an open set $U \subset \mathbb{R}^n$ is called a **$k$-dimensional submanifold** of $U$ if for any point $a \in V$ there are an open neighborhood $U_a \subset U$ and a local change of coordinates in $U_a$ such that $V \cap U_a$ is a linear subspace of dimension $k$ in terms of these coordinates. If the change of coordinates is described by a continuously differentiable mapping the subset $V$ is said somewhat more precisely to be a $C^1$ submanifold; similarly it is said to be a $C^2$ submanifold if the change of coordinates is described by a $C^2$ mapping, and
so on. If the local coordinates are described by only a continuous mapping, the set $V$ is just a topological submanifold. For example the circle

$$(3.11) \quad V = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1 \right\} \subset \mathbb{R}^2$$

can be defined as the zero locus of the function $f(x) = x_1^2 + x_2^2 - 1$ in the plane, and since $f'(x) = (2x_1, 2x_2)$ is a matrix of rank 1 at each point $x \in V$, at which not both $x_1 = 0$ and $x_2 = 0$, the zero locus $V$ is a one-dimensional submanifold of the plane. Indeed it can be described locally as the linear space $r = 1$ in terms of polar coordinates $(r, \theta)$ in the plane in a neighborhood of each point of $V$; but of course there is no single coordinate neighborhood in which the entire circle can be described by the equation $r = 1$, since the other coordinate $\theta$ is not a single-valued function on the full circle. The condition that a subset $V \subset U$ of an open set $U \subset \mathbb{R}^m$ be a submanifold is a local condition; so if the subset $V$ is a submanifold in a neighborhood of each of its points, it is a submanifold. It is natural though to consider two $C^1$ submanifolds $V_1 \subset U_1, V_2 \subset U_2$ of open sets $U_1, U_2 \subset \mathbb{R}^m$ as being equivalent if there is a $C^1$ homeomorphism $\phi : U_1 \to U_2$ such that $\phi(V_1) = V_2$; two submanifolds $V_1$ and $V_2$ that are equivalent in this sense are called homeomorphic submanifolds. More precisely they are called $C^1$ homeomorphic submanifolds in this case, and are called $C^r$ homeomorphic submanifolds if the homeomorphism $\phi$ is a $C^r$ homeomorphism; if $\phi$ is only a topological homeomorphism the submanifolds are called just topologically homeomorphic submanifolds. This equivalence relation really involves not just the point set $V$ but the particular way in which that set is imbedded in the open set $U \subset \mathbb{R}^m$. For example a circle in $\mathbb{R}^3$ and a knotted curve in $\mathbb{R}^3$ are both submanifolds of $\mathbb{R}^3$; but they are not homeomorphic topological submanifolds, since there is no homeomorphism of $\mathbb{R}^3$ to itself that transforms a circle into a knotted curve.

**Theorem 3.4** A relatively closed subset $V \subset U$ of an open set $U \subset \mathbb{R}^m$ is a $k$-dimensional $C^1$ submanifold of $U$ if and only if in an open neighborhood $U_a$ of any point $a \in V$ there are $m - k$ continuously differentiable functions $f_i(x)$ for $1 \leq i \leq m - k$ such that the gradients $\nabla f_i(a)$ are linearly independent and

$$(3.12) \quad V \cap U_a = \left\{ x \in U_a \mid f_1(x) = \cdots = f_{m-k}(x) = 0 \right\}.$$  

**Proof:** If $V$ is a $k$-dimensional $C^1$ submanifold of $U$ then by definition there is a $C^1$ change of coordinates in an open neighborhood $U_a$ of any point $a \in V$ such that in terms of the new coordinates $t_1, \ldots, t_m$ in the neighborhood $U_a$ the intersection $V \cap U_a$ is the $k$-dimensional linear subspace described by the equations $t_1 = \cdots = t_{m-k} = 0$. The coordinates $t_i$ are $C^1$ functions $t_i = f_i(x)$ of the original coordinates $x_1, \ldots, x_m$ in the neighborhood $U_a$, and the Jacobian matrix of the change of coordinates is the non-singular $m \times m$ matrix $(\nabla f_1(x) \cdots \nabla f_m(x))$. Hence in terms of the original coordinates the intersection $V \cap U_a$ is described by the equations $f_1(x) = \cdots = f_{m-k}(x) = 0$ where the gradients $\nabla f_1(x), \ldots, \nabla f_{m-k}(x)$ are linearly independent at each point $x \in U_a$. Conversely if the relatively closed subset $V \subset U$ is described by the equations (3.12) in an open neighborhood $U_a$ of a point $a \in U$ for some $C^\infty$ functions.

---

3The study of knots is an old subject in mathematics, going back to work of P. G. Tait in the 1870’s, but it is still a very active subject, with recent ties to mathematical physics and representation theory among other topics. Classical knot theory focuses on the classification of knots and the algebraic invariants that distinguish different knots. Further information can be found in the book by R. H. Crowell and R. H. Fox, *Introduction to Knot Theory*, Springer, 1963.
f_1, \ldots, f_{m-k} in U_a and if the gradients \nabla f_i(a) are linearly independent then after relabeling the variables in \mathbb{R}^m it can be assumed that the \((m-k) \times (m-k)\) matrix \(\partial_i f_j(a)\) for \(1 \leq i, j \leq m-k\) is nonsingular. It then follows from Theorem 3.3 that after shrinking the neighborhood \(U_a\) further if necessary there is a \(C^1\) homeomorphism \(\phi : W \rightarrow U_a\) between an open neighborhood \(W\) of the origin in \(\mathbb{R}^m\) and the open neighborhood \(U_a\) such that \((f \circ \phi)(t) = \{t_1, \ldots, t_{m-k}\}\). The mapping \(\phi\) thus is a local change of coordinates in the open subset \(U_a\), and the functions \(f_i(x)\) describe the new local coordinates \(t_i = f_i(x)\); since the subset \(V\) is characterized by the condition that \(f_i(x) = 0\) for \(1 \leq i \leq m-k\) then in terms of the new local coordinates \(t_1, \ldots, t_m\) the subset \(V \cap U_a\) is characterized by the condition that \(t_1 = \cdots = t_{m-k} = 0\), so the intersection \(V \cap U_a\) is a linear subspace of dimension \(k\) in terms of these new coordinates. That is the case for any point \(a \in V\) so \(V\) is a \(C^1\) submanifold of \(U\), and that suffices for the proof.

This characterization of submanifolds is the simplest to apply, and is the one most commonly used to describe explicit submanifolds of subsets of Euclidean spaces. If the gradients \(\nabla f_i(x), \ldots, \nabla f_{m-k}(x)\) are linearly dependent at some points the set of zeros is not necessarily a submanifold at these exceptional points but may have singularities of one sort or another. The study of such singularities has been and still is a very active area of mathematical research. However it should be noted that the zero locus of a collection of functions \(f_1(x), \ldots, f_{m-k}(x)\) may still be a \(k\)-dimensional submanifold even though their gradients are linearly dependent at all points. For instance the zero locus of the function \(f(x) = (x_1^2 + x_2^2 - 1)^2\) is also a submanifold of \(\mathbb{R}^2\), a circle; but \(f'(x) = 2(x_1^2 + x_2^2 - 1)(2x_1 2x_2)\) and consequently \(f'(x) = 0\), the zero vector, at each point of the zero locus of the function \(f\). Some more more

\[4\] Some classical examples of singularities are isolated singularities of algebraic curves, the topological properties of which were investigated by K. Brauner (Abh. Math. Sem. Hamburg, vol. 6 (1928), pages 8 - 54). A polynomial function \(p(z_1, z_2)\) of two complex variables amounts to a pair of real functions of four real variables, the real and imaginary parts of the variables and of the function; so the set \(V\) of points at which \(p(z_1, z_2) = 0\) is a candidate to be a submanifold of dimension 2 in \(\mathbb{R}^4\). For a linear polynomial \(p(z_1, z_2) = a_1z_1 + a_2z_2\) the set \(V\) is a linear submanifold, and the intersection \(V \cap S^3\) of the set \(V\) and the sphere \(S^3 = \{z \in \mathbb{C}^2 | |z_1|^2 + |z_2|^2 = 1\}\) is a one-dimensional submanifold of the sphere, a linear slice of the sphere so just an ordinary circle. If the polynomial \(p(z_1, z_2)\) has no constant term but a nontrivial linear term then again the set \(V\) is a 2-dimensional submanifold near the origin, and the intersection \(V \cap S^3\) is a submanifold of the sphere that is \(C^1\) homeomorphic to a circle in the sphere. The set \(V\) of zeros of the polynomial \(p(z_1, z_2) = z_1^2 + z_2^2\) is a 2-dimensional submanifold near the origin except at the origin itself, which is a singularity; actually this polynomial splits as the product \(p(z_1, z_2) = (z_1 + i z_2)(z_1 - i z_2)\) of two linear functions both of which vanish at the origin, so near the origin the set \(V\) is the union of two linear spaces of dimension 2, the linear spaces \(z_1 + i z_2 = 0\) and \(z_1 - i z_2 = 0\), which meet only at a single point, the origin. In this case the intersection \(V \cap S^3\) is a submanifold consisting of two disjoint circles which are not linked in the three-dimensional sphere. For the polynomial \(p(z_1, z_2) = z_1^2 + z_2^2\) the origin is again a singularity, and in this case the intersection \(V \cap S^3\) as a submanifold of the sphere is a nontrivial knot in the sphere, which is fairly easy to describe. In general the intersection \(V \cap S^3\) is a submanifold of the sphere that is a knot of a special type, and its topological structure determines a good deal of the structure of the set \(V\) at its singular point. There are analogous topological sets that arise in the examination of isolated singularities of the sets of zeros of polynomials in several variables, a fascinating subject in its own right; these examples can be used to introduce the exotic differentiable structures on higher dimensional spheres, as discovered by J. W. Milnor. A more detailed discussion can be found in Milnor’s book Singular Points of Complex Hypersurfaces, Annals of Mathematics Studies, number 61, Princeton University Press, 1968.
Theorem 3.5 (Implicit Function Theorem) If \( f : \Delta \rightarrow \mathbb{R}^{m-k} \) is a continuously differentiable mapping defined in an open neighborhood of the origin \( 0 \in \mathbb{R}^m \), where the neighborhood \( \Delta \) is a cell that is the product \( \Delta = \Delta_I \times \Delta_{II} \) of cells \( \Delta_I \subset \mathbb{R}^{m-k} \) and \( \Delta_{II} \subset \mathbb{R}^k \), and if \( f(0) = 0 \) and the \((m-k) \times (m-k)\) matrix \((\partial_1 f(0) \cdots \partial_{m-k} f(0))\) is of rank \( m-k \), then after shrinking the neighborhood \( \Delta \) if necessary the zero locus of the mapping \( f \) is a submanifold \( V \subset \Delta \) that can be described by a set of equations of the form

\[
V \cap \Delta = \left\{ x \in U_a \left| x_i = g_i(x_{m-k+1}, \ldots x_m), \ 1 \leq i \leq m-k \right. \right\}
\]

for some continuously differentiable functions \( g_i \) in the cell \( \Delta_{II} \subset \mathbb{R}^k \).

**Proof:** It follows from Theorem 3.3 that after shrinking the neighborhood \( \Delta \) sufficiently there is a \( C^1 \) change of coordinates in \( \Delta \) such that the original coordinates \( x_1, \ldots, x_m \) are related to the new coordinates \( t_1, \ldots, t_m \) by a mapping \( x = \phi(t) \) of the form (3.9), so that

\[
x_i = g_i(t_1, \ldots, t_m) \quad \text{for } 1 \leq i \leq m-k \text{ and}
\]

\[
x_i = t_i \quad \text{for } m-k+1 \leq i \leq m
\]

for some \( C^1 \) functions \( g_i(t) \); and by (3.10) in terms of the new coordinates the functions \( f_i(x) \) are given by

\[
f_i(\phi(t_1, \ldots, t_m)) = t_i \quad \text{for } 1 \leq i \leq m-k.
\]

A point \( x \in U_a \) which is described in terms of the new local coordinates as \( x = \phi(t) \) is contained in the submanifold \( V \) if and only if \( f_i(\phi(t)) = t_i = 0 \) for \( 1 \leq i \leq m-k \), or in view of (3.14) if and only if

\[
x_i = g_i(0, \ldots, 0, t_{m-k+1}, \ldots, t_m) \quad \text{for } 1 \leq i \leq m-k \text{ and}
\]

\[
x_i = t_i \quad \text{for } m-k+1 \leq i \leq m,
\]

which can be rewritten

\[
x_i = g_i(0, \ldots, 0, x_{m-k+1}, \ldots, x_m) \quad \text{for } 1 \leq i \leq m-k,
\]

and that suffices for the proof.

The result of the preceding theorem is illustrated in the accompanying Figure 3.3. The zero locus of the function \( f \) is the image \( X = \phi(L) \subset \Delta \) of the linear subspace \( L \) defined by the equations \( t_1 = \cdots = t_{m-k} = 0 \) under the \( C^1 \) mapping \( \phi : \mathbb{R}^m \longrightarrow \mathbb{R}^m \) that describes the change of coordinates in an open neighborhood of the point \( a \). The mapping \( \phi \) does not change the last \( k \) coordinates but only the first \( m-k \) coordinates of the points; so the mapping preserves the “height” of points above the coordinate axis of the first \( m-k \) coordinates, but moves the points by changing their first \( m-k \) coordinates. Points on the linear subspace \( L \) are described by the parameters \( t_i = x_i \) for \( m-k+1 \leq i \leq m \) while the coordinates \( t_i \) are held constant for \( 1 \leq i \leq m-k \); points on the set \( X \) are also described by the parameters \( t_i = x_i \) for \( m-k+1 \leq i \leq m \), but the coordinates \( x_i \) for \( 1 \leq i \leq m-k \) vary, depending on the parameters \( t_i = x_i \) for \( m-k+1 \leq i \leq m \).
Figure 3.3: The bijective mapping $\phi$ takes the linear subspace described by $f(\phi(t)) = 0$ to the set described by $f(x) = 0$, preserving the last $m - n$ coordinates but shifting the first $n$ coordinates of these subsets.

**Corollary 3.4** A relatively closed subset $V \subset U$ of an open set $U \subset \mathbb{R}^m$ is a $k$-dimensional $C^1$ submanifold of $U$ if and only if for any point $a \in V$, after a permutation of the coordinates in $\mathbb{R}^m$ if necessary, there is an open cell $\Delta_a = \Delta_{I,a} \times \Delta_{II,a}$ centered at the point $a \in V$, a product of cells $\Delta_{I,a} \in \mathbb{R}^{m-k}$ and $\Delta_{II,a} \in \mathbb{R}^k$, such that the intersection $V \cap \Delta_a$ is described by the equations

$$V \cap U_a = \left\{ x \in U_a \mid x_i = g_i(x_{m-k+1}, \ldots, x_m), \ 1 \leq i \leq m-k \right\}$$

for some continuously differentiable functions $g_i$ in the cell $\Delta_{II,a} \in \mathbb{R}^k$.

**Proof:** If $V$ is a $k$-dimensional $C^1$ submanifold of $U$ then by the preceding theorem the subset $V$ can be described in an open neighborhood of any point $a \in V$ by the equations (3.19). Conversely if $V$ is defined by the equations (3.19) these equations can be rewritten $f_i(x) = x_i - g_i(x_{m-k+1}, \ldots, x_m) = 0$ for $1 \leq i \leq m-k$, and since $\nabla f_i(x)$ are clearly linearly independent vectors it follows from Theorem 3.4 that $V$ is a $k$-dimensional $C^1$ submanifold. That suffices for the proof.

In the Implicit Function Theorem, the equation $f(x) = 0$ can be viewed as a description of relations among the coordinates $x_1, \ldots, x_m$ in $\mathbb{R}^m$, which implicitly describe some of the coordinates as functions of the remaining coordinates; the theorem provides conditions determining which coordinates really can be viewed as explicit functions of the remaining coordinates, at least locally. For example, if $f : U \to \mathbb{R}^1$ is a continuously differentiable function in an open subset $U \subset \mathbb{R}^m$ and if $\partial_1 f(a) \neq 0$ at some point $a \in U$ then by the Implicit Function Theorem there are an open neighborhood $\Delta \subset U$ of the point $a$ which is a product $\Delta = \Delta_I \times \Delta_{II}$ of cells $\Delta_I \subset \mathbb{R}^1$ and $\Delta_{II} \subset \mathbb{R}^{m-1}$ and a mapping $g : \Delta_{II} \to \Delta_I$ such that $f(x_1, \ldots, x_m) = 0$ for a point $\{x_1, \ldots, x_m\} \in \Delta$ if and only if $x_1 = g(x_2, \ldots, x_m)$. The equation $f(x_1, x_2, \ldots, x_m) = 0$ defines $x_1$ implicitly as a function of the remaining variables, and in an open neighborhood of any point $a$ at which $\partial_1 f(a) \neq 0$ this equation also defines $x_1$ explicitly as a function $x_1 = g(x_2, \ldots, x_m)$ of the remaining variables. It is obvious that some condition on the function $f$ is necessary for this to be true; for instance without any condition the function $f(x_1, \ldots, x_m)$ might be independent of the variable $x_1$, in which case it can say nothing at all about that variable. The example in which $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 1$ illustrates another way in which some condition on the function $f$ is necessary. The equation $f(x) = 0$ can be viewed as
defining \( x_1 \) as the function \( x_1 = \pm \sqrt{1 - x_2^2 - x_3^2} \), where the square root has two well defined locally continuous values so long as \( x_2^2 + x_3^2 < 1 \); but \( x_1 \) is not a well defined function in an open neighborhood of any point at which \( x_2^2 + x_3^2 = 1 \), or equivalently at which \( \partial_1 f(x_1, x_2, x_3) = 0 \). If the variable \( x_1 \) is defined explicitly by \( x_1 = g(x_2, x_3) \) then \( f(g(x_2, x_3), x_2, x_3) = 0 \) identically in the variables \( x_2, x_3 \); and the chain rule can be applied to calculate the derivatives of the function \( g(x_2, x_3) \), since

\[
0 = \frac{\partial}{\partial x_2} f(g(x_2, x_3), x_2, x_3)) = \frac{\partial f}{\partial x_2} g(x_2, x_3) \frac{\partial g}{\partial x_2}(x_2, x_3) + \frac{\partial f}{\partial x_3} g(x_2, x_3) \frac{\partial g}{\partial x_3}(x_2, x_3))
\]

Thus the partial derivatives of \( g \) can be expressed in terms of the partial derivatives of \( f \), and similarly for higher derivatives. Of course the same argument can be applied to another variable \( x_i \), expressing it as a function of the remaining variables provided that \( \partial_i f(x) \neq 0 \). If \( f_1, f_2 \) are continuously differentiable functions in in an open subset \( U \subset \mathbb{R}^4 \) and if

\[
(3.20) \quad \det \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_1}(a) \\ \frac{\partial f_1}{\partial x_2}(a) & \frac{\partial f_2}{\partial x_2}(a) \end{pmatrix} \neq 0
\]

then the set of points at which \( f(x) = 0 \) near \( a \) can be described explicitly as the set of points at which \( x_1 \) and \( x_2 \) are explicitly given functions of the remaining variables, say \( x_1 = g(x_3, x_4) \) and \( x_2 = g_2(x_3, x_4) \). Thus there are the identities \( f_1(g_1(x_3, x_4), x_2, x_3, x_4) = f_2(g_2(x_3, x_4), x_2, x_3, x_4) = 0 \); and the partial derivatives of these identities yield two linear equations relating the partial derivatives of the functions \( g_1(x_3, x_4) \) and \( g_2(x_3, x_4) \), which can be solved explicitly in view of (3.20).

### 3.3 The Rank Theorem

The Inverse Mapping Theorem can be applied to describe mappings between spaces in situations more general than those to which the Implicit Function applies, in particular to describe mappings \( f : \mathbb{R}^m \rightarrow \mathbb{R}^n \) where \( m < n \).

**Theorem 3.6 (Rank Theorem)** If \( f : U \rightarrow \mathbb{R}^n \) is a continuously differentiable mapping defined in an open subset \( U \subset \mathbb{R}^m \) and if \( \text{rank} f'(x) = r \) for all points in an open neighborhood of a point \( a \in U \) then in suitable local coordinates near the points \( a \in \mathbb{R}^m \) and \( f(a) \in \mathbb{R}^n \) the mapping \( f \) is a linear mapping of rank \( r \).

**Proof:** This theorem involves a change of local coordinates in both the domain and range of the mapping \( f \), so requires the existence at least locally of \( \mathcal{C}^1 \) homeomorphisms \( \psi \) and \( \theta \) in the diagram

\[
\mathbb{R}^m \xrightarrow{\psi} \mathbb{R}^m \xrightarrow{f} \mathbb{R}^n \xrightarrow{\theta} \mathbb{R}^n
\]

so that the composition \( \theta \circ f \circ \psi : \mathbb{R}^m \rightarrow \mathbb{R}^n \) is a linear mapping of rank \( r \). If \( \text{rank} f'(a) = r \) then by suitably relabeling the variables in the domain \( \mathbb{R}^m \) and range \( \mathbb{R}^n \) of the mapping \( f \) it can be assumed that the leading \( r \times r \) submatrix of \( f'(a) \) is a nonsingular \( r \times r \) matrix; and by continuity that is also the case for the matrix \( f'(x) \) for all points \( x \in U_a \subset \mathbb{R}^m \) for a sufficiently small open neighborhood \( U_a \) of the point \( a \). As before it is convenient to write points \( x \in \mathbb{R}^m \) in the form \( x = (x_i) \) where
\( x_l \in \mathbb{R}^r \) and \( x_{ll} \in \mathbb{R}^{m-r} \), and correspondingly to write points \( y \in \mathbb{R}^n \) in the form \( y = (y_{ll}) \) where \( y_l \in \mathbb{R}^r \) and \( y_{ll} \in \mathbb{R}^{n-r} \). The mapping \( f \) can then be written

\[
f \begin{pmatrix} x_l \\
x_{ll} \end{pmatrix} = \begin{pmatrix} f_l(x) \\
f_{ll}(x) \end{pmatrix}
\]

where \( x_l, f_l(x) \in \mathbb{R}^r \) while \( x_{ll}, f_{ll}(x) \in \mathbb{R}^{n-r} \); and the matrix \( f'(x) \) then can be decomposed into the submatrices

\[
f'(x) = \begin{pmatrix} \partial_l f_l(x) & \partial_{ll} f_l(x) \\
\partial_{ll} f_l(x) & \partial_{lll} f_l(x) \end{pmatrix}
\]

where \( \partial_l f_l(x) \) is an \( r \times r \) matrix which is nonsingular for all points \( x \in U_a \). In these terms, introduce the mapping \( \phi : U_a \rightarrow \mathbb{R}^m \) defined by

\[
\phi \begin{pmatrix} x_l \\
x_{ll} \end{pmatrix} = \begin{pmatrix} f_l(x) \\
x_{ll} \end{pmatrix},
\]

so that

\[
\phi'(x) = \begin{pmatrix} \partial_l f_l(x) & \partial_{ll} f_l(x) \\
0 & I_{m-r} \end{pmatrix}
\]

where \( I_{m-r} \) is the \( (m-r) \times (m-r) \) identity matrix; since the matrix \( \partial_l f_l(x) \) is nonsingular for all points \( x \in U_a \) it follows that the matrix \( \phi'(x) \) also is nonsingular for all points \( x \in U_a \). If the neighborhood \( U_a \) is sufficiently small it follows from the Inverse Mapping Theorem that the restriction of the mapping \( \phi \) to \( U_a \) is a \( C^1 \) homeomorphism \( \phi : U_a \rightarrow V \) between the open sets \( U_a \) and \( V \subset \mathbb{R}^m \); and clearly the inverse mapping \( \psi : V \rightarrow U_a \) has the form

\[
\psi \begin{pmatrix} t_l \\
t_{ll} \end{pmatrix} = \begin{pmatrix} g_{ll}(t) \\
t_{ll} \end{pmatrix} \quad \text{for a } C^1 \text{ mapping } g_{ll} : V \rightarrow \mathbb{R}^r.
\]

Since \( \phi(\psi(t)) = t \) it follows that \( f_l(\psi(t)) = t_l \); consequently the composite mapping \( h(t) = (f \circ \psi)(t) \) has the form

\[
h(t) = \begin{pmatrix} t_l \\
h_{ll}(t) \end{pmatrix} \quad \text{where } h_{ll}(t) = f_{ll}(\psi(t)) \in \mathbb{R}^{n-r},
\]

and

\[
h'(t) = \begin{pmatrix} I_r & 0 \\
\partial_l h_{ll}(t) & \partial_{ll} h_{ll}(t) \end{pmatrix}
\]

where \( I_r \) is the \( r \times r \) identity matrix. On the other hand since \( h = f \circ \psi \) then by the chain rule \( h'(t) = f'(\psi(t)) \cdot \psi'(t) \); the \( m \times m \) matrix \( \psi'(t) \) is nonsingular for all \( t \in U_a \), while by hypothesis \( \text{rank } f'(\psi(t)) = r \) for all \( t \in U_a \), and consequently \( \text{rank } h'(t) = r \) for all points \( t \in U_a \) as well. However it is evident from (3.22) that \( \text{rank } h'(t) = r \) if and only if \( \partial_l h_{ll}(t) = 0 \) for all points \( t \in U_a \); so the mapping \( h_{ll} \) is a function just of the variables \( t_l \in \mathbb{R}^r \). The mapping \( \theta : \mathbb{R}^n \rightarrow \mathbb{R}^n \) defined by

\[
\theta \begin{pmatrix} y_l \\
y_{ll} \end{pmatrix} = \begin{pmatrix} y_l \\
y_{ll} - h_{ll}(y_l) \end{pmatrix}
\]

has the derivative

\[
\theta'(y) = \begin{pmatrix} I_r & 0 \\
-\partial_l h_{ll}(y_l) & I_{n-r} \end{pmatrix},
\]
so after restricting the neighborhood $U_a$ further if necessary this mapping can be viewed as a local change of coordinates in that neighborhood by the Inverse Mapping Theorem; and it is evident from the definition of this mapping and the explicit form (3.21) of the mapping $h$, in which $h_{1I}(t)$ is a function of the variables $t_I$ alone, that

$$(\theta \circ h)(t) = (\theta \circ f \circ \psi)(t) = \begin{pmatrix} t_I \\ 0 \end{pmatrix},$$

which is a linear mapping of rank $r$. That suffices for the proof.

**Corollary 3.5** If $f : U \rightarrow \mathbb{R}^n$ is a continuously differentiable mapping defined in an open subset $U \subset \mathbb{R}^n$ and if $\text{rank} f'(x) = r$ for all points $x \in U$ then

(i) the image $f(U_a)$ of any sufficiently small open neighborhood $U_a \subset U$ of a point $a \in U$ is a $C^1$ submanifold of dimension $k$ of an open neighborhood of the point $f(a) \in \mathbb{R}^n$;

(ii) for any point $b \in f(U) \subset \mathbb{R}^n$ the inverse image $f^{-1}(b)$ is a $C^1$ submanifold of dimension $n - k$ in $U$.

**Proof:** It follows from the preceding theorem that after suitable changes of coordinates in open neighborhoods $U_a$ of the point $a \in U \subset \mathbb{R}^n$ and $V_b$ of the point $b = f(a) \in \mathbb{R}^n$ the mapping $f$ is a linear mapping of rank $k$; consequently the image $f(U_a)$ is a linear subspace of dimension $k$ in terms of the new coordinates in $\mathbb{R}^n$ and the inverse image $f^{-1}(y)$ of any point $y \in f(U_a) \subset \mathbb{R}^n$ is a linear subspace of dimension $n - k$ in terms of the new coordinates in $\mathbb{R}^m$. Consequently the image $f(U_a)$ is a $k$-dimensional submanifold of $V_b$ in terms of the original coordinates in $\mathbb{R}^n$, and the inverse image $f^{-1}(y) \cap U_a$ of any point $y \in f(U_a)$ is a $k$-dimensional $C^1$ submanifold of $U_a$ in terms of the original coordinates in $\mathbb{R}^m$. The latter statement holds for any point in $f^{-1}(y)$; so since the inverse image $f^{-1}(y) \subset U$ is a relatively closed subset of $U$ which is locally a submanifold of $U$ it must be a submanifold of $U$, and that suffices for the proof.

Part (i) of the preceding corollary can only be expected to hold locally. For instance it is quite possible that the image of an open interval $I \subset \mathbb{R}$ under a continuously differentiable mapping $\phi : I \rightarrow \mathbb{R}^n$ is a curve that intersects itself at some points, so that although the image is locally a one-dimensional submanifold the full image has singularities; and the image of the open interval need not even be a closed subset of $\mathbb{R}^n$. The situation in higher dimensions of course is even more complicated; but the local result is of considerable use. For the Inverse Mapping Theorem and the Implicit Function Theorem the hypothesis only required that the mapping be of maximal rank at a point $a$, or equivalently that there is a square submatrix of maximal size in the Jacobian matrix of the mapping so that the determinant of that matrix at the point $a$ is nonzero; and since the Jacobian of a $C^1$ mapping is a continuous function it is necessarily nonzero at all points near enough to $a$. However if the rank of the Jacobian matrix is not maximal at a point $a$ its rank at nearby points may well be greater at points arbitrarily close to $a$. That is the reason that the hypothesis in the rank theorem requires that the rank be locally constant. In some applications of the Rank Theorem it is still the case that the rank is maximal, so if the rank is maximal at a point is is maximal at all nearby points.

**Corollary 3.6** If $\phi : U \rightarrow \mathbb{R}^n$ is a continuously differentiable mapping from an open set $U \subset \mathbb{R}^k$ into $\mathbb{R}^n$ where $k < n$ such that $\text{rank} \, \phi'(x) = k$ at each point $x \in U$ then the image under $\phi$ of any sufficiently small open neighborhood of any point $a \in U$ is a $C^1$ submanifold of an open neighborhood of the point $\phi(a) \in \mathbb{R}^n$. 
Proof: If rank \(\phi'(a) = k\) at a point \(a \in U\) for a mapping \(\phi : \mathbb{R}^k \rightarrow \mathbb{R}^n\) where \(k < n\) then rank \(\phi'(x) = k\) for all points \(x\) near \(k\), so the desired result follows immediately from the preceding theorem, and that suffices for the proof.

For example a mapping \(f : [0, 1] \rightarrow \mathbb{R}^m\) can be viewed as a \textbf{parametrized curve} \(\gamma\) in the vector space \(\mathbb{R}^m\), beginning at the point \(f(0)\) and ending at the point \(f(1)\); and if \(f'(t) \neq 0\) at each point \(t \in [0, 1]\) then by Corollary 3.6 the image of an open neighborhood of any point \(t_0 \in (0, 1)\) is a 1-dimensional submanifold of an open neighborhood of the point \(f(t_0) \in \mathbb{R}^m\). That the mapping \(f\) is differentiable at the point \(t_0\) means that

\[
f(t) = f(t_0) + f'(t_0)(t - t_0) + \epsilon(t) \quad \text{where} \quad \lim_{t \to t_0} \frac{\|\epsilon(t)\|}{\|t - t_0\|} = 0
\]

for all points \(t\) sufficiently near \(t_0\). The parametrized curve described by \(f_0(t) = f(t_0) + f'(t_0)(t - t_0)\) is a line in \(\mathbb{R}^m\) passing through the point \(f(t_0)\), called the \textbf{tangent line} to the curve \(\gamma\) at the point \(f(t_0) = a_0\), and is of course another submanifold of \(\mathbb{R}^m\); the vector \(f'(t_0)\) is called the \textbf{tangent vector} to the curve \(\gamma\) at the point \(f(t_0) = a_0\). It follows from \(2.1\) and the uniqueness of the derivative of a function that the tangent line to the curve at the point \(a_0\) is the parametrized line through that point that is the best linear approximation to the curve near that point. It is more common to change the parametrization of the tangent line so that the point \(a\) corresponds to the parameter value \(t = 0\), so to describe the tangent line as the parametric curve \(f_0(t) = f(t_0) + f'(t_0)t\). A continuously differentiable mapping \(\phi : U \rightarrow \mathbb{R}^n\) from an open subset \(U \subset \mathbb{R}^m\) into \(\mathbb{R}^n\) takes a curve \(\gamma \subset U\) described parametrically by a mapping \(f : [0, 1] \rightarrow U\) to the curve \(\phi(\gamma) \subset \mathbb{R}^n\) described parametrically by the mapping \(\phi \circ f : [0, 1] \rightarrow \mathbb{R}^n\). By the chain rule the tangent vector to the image curve \(\phi(\gamma)\) at the point \(\phi(a_0)\) is the vector \((\phi \circ f)'(t_0) = \phi'(a_0)f'(t_0)\). Thus the derivative \(\phi'(a_0)\) of the mapping \(\phi\) at the point \(a_0\) can be interpreted as the linear mapping that takes tangent vectors to curves through the point \(a_0\) to tangent vectors to the image curves through the point \(\phi(a_0)\).

If \(V\) is a \(C^1\) submanifold of an open subset \(U \subset \mathbb{R}^m\) then through each point \(a \in V\) it is possible to pass a continuously differentiable curve contained entirely in the set \(V\) near any point \(a \in V\); for in some local coordinates near the point \(a\) the subset is a linear space, and there are straight lines through that point contained within that linear subspace. Moreover if \(\dim V = k\) there are curves with tangent vectors spanning a linear subspace of dimension \(k\); for in some local coordinates near the point \(a\) the subset is a linear subspace of dimension \(k\), there are \(k\) linearly independent straight lines through any point of a linear space of dimension \(k\). For a submanifold \(V \subset U\) of an open set \(U \subset \mathbb{R}^m\) defined by equations

\[
(3.23) \quad V = \{ x \in U \mid f_1(x) = \cdots = f_{n-k}(x) = 0 \}
\]

where the vectors \(\nabla f_i(x)\) are linearly independent at each point \(x \in V\), the linear subspace of dimension \(k\) consisting of the tangent vectors at the point \(a\) of all continuously differentiable curves contained in \(V\) and passing through that point is called the \textbf{tangent space} to the submanifold \(V\) at the point \(a\) and is denoted by \(T_a(V)\). For a curve described as the image of a continuously differentiable mapping \(\phi : I \rightarrow V\) of an interval \(I \subset \mathbb{R}\) into \(V\) with \(\phi(0) = a\), since \(f_i(\phi(t)) = 0\) identically in \(t\) it follows from the chain rule that \(f'_i(a)\phi'(0) = 0\), or equivalently that \(\nabla f_i(a) \cdot \phi'(0) = 0\); thus the tangent vector to the curve is perpendicular to all of the gradient vectors \(\nabla f_i(a)\), so the tangent space to \(V\) at the point \(a \in V\) is the linear subspace consisting of all
points \( x \in \mathbb{R}^m \) such that the vector \( x - a \) is perpendicular to the vectors \( \nabla f_i(a) \); The equation of the tangent space consequently is

\[
T_a(V) = \left\{ x \in \mathbb{R}^m \mid (x - a) \cdot \nabla f_i(a) = 0 \quad \text{for} \quad 1 \leq i \leq n \right\}.
\]

The tangent space is the linear subspace of \( \mathbb{R}^n \) that is the best local approximation to the submanifold \( V \) near that point, just as is the tangent vector to a curve. For example, if \( V \subset \mathbb{R}^3 \) is the surface defined by

\[
V = \left\{ x \in \mathbb{R}^3 \mid f(x) = x_1^2 + x_2^3 + x_3^4 - 1 = 0 \right\}
\]

then \( f'(1, 1, 1) = (2 \ 3 \ 4) \) so the equation of the tangent plane to \( V \) at the point \( (1, 1, 1) \) is \( \{2, 3, 4\} \cdot \{x_1 - 1, x_2 - 1, x_3 - 1\} = 0 \) or equivalently \( 2x_1^3x_2 + 4x_3 = 9 \).

To determine the extrema of a \( \mathcal{C}^1 \) function \( g \) defined on a \( k \)-dimensional submanifold \( V \subset \mathbb{R}^m \) it is always possible theoretically to introduce local coordinates in an open neighborhood \( U \) of any point \( a \in V \) so that \( V \cap U \) is a \( k \)-dimensional linear subspace in terms of those coordinates; the restriction of the function \( g \) to \( V \) then is locally just a function on a \( k \)-dimensional linear subspace of \( \mathbb{R}^m \), so can be viewed as a function on an open subset of \( \mathbb{R}^k \) and its critical points are candidates for points at which the function has an extremal value. However that can be a rather arduous calculation in all but the simplest situations, since determining the local coordinates in terms of which the submanifold is linear can be quite difficult to do explicitly. However the implicit function theorem provides a much simpler and widely used method for finding local extrema on a submanifold. For the submanifold \( V \) defined by (3.23) where the vectors \( \nabla f_i(x) \) are linearly independent at all points \( x \in V \), a point \( a \in V \) will be a local extremum of the restriction of the continuously differentiable function \( g(x) \) to the submanifold \( V \) if and only if it is a local extrema of the restriction of the function \( g(x) \) to any continuously differentiable curve lying in \( V \) and passing through the point \( a \). Then for any \( \mathcal{C}^1 \) parametrized curve \( \phi : U \rightarrow \mathbb{R}^m \) in an open neighborhood \( U \subset \mathbb{R}^1 \) of the origin \( t = 0 \) passing through the point \( a \), so that \( \phi(0) = a \), and contained in \( V \), so that \( f_i(\phi(t)) = 0 \) for all \( t \in U \), the origin \( t = 0 \) is a local extrema for the composite function \( g(\phi(t)) \) and consequently

\[
0 = \frac{d}{dt}g(\phi(t))\bigg|_{t=0} = g'(a)\phi'(0) = \nabla g(a) \cdot \phi'(0);
\]

thus \( \nabla g(a) \) must be perpendicular to the tangent line to the curve, and since that is the case for all curves in \( V \) the vector \( \nabla g(a) \) must be perpendicular to the tangent space \( T_a(V) \) to the submanifold \( V \) at the point \( a \). However the tangent space is characterized in (3.24) as the plane perpendicular to the gradients \( \nabla f_i(a) \), so the vector \( \nabla g(a) \) must be contained in the span of the vectors \( \nabla f_i(a) \), consequently must be expressible as a linear combination

\[
\nabla g(a) = \sum_{i=1}^{n} \lambda_i \nabla f_i(a)
\]

for some parameters \( \lambda_i \in \mathbb{R} \). This set of equations, combined with the equations (3.12) for the submanifold \( V \), are an explicit set of equations that can determine the local extrema of the function \( g(x) \), a technique called the method of Lagrange multipliers.

As an example of the use of this method, find the minimum of \( x + y + z \) subject to the restrictions that \( \frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 1 \) and \( x, y \) and \( z \) are positive, where \( a, b, \) and \( c \) be fixed and
positive. The submanifold $V$ is defined by the equation $f(x, y, z) = \frac{x}{a} + \frac{y}{b} + \frac{z}{c} - 1 = 0$, and the function of interest is $g(x, y, z) = x + y + z$, where

$$\nabla g = (1, 1, 1) \quad \nabla f = \left( \frac{-a}{x^2}, \frac{-b}{y^2}, \frac{-c}{z^2} \right)$$

The Lagrange formula is the equation $\nabla g = \lambda \cdot \nabla f$; the first component of this vector equation is $\lambda = -x^2/a$, and upon substituting this into the other two components there result the equations

\[
y = x \sqrt{\frac{b}{a}} \quad \text{and} \quad z = x \sqrt{\frac{c}{a}}
\]

Since $a, b$ and $c$ are fixed there results a one-parameter set of solutions, parametrized by $x$. When restricted to the set $S = \{(x, y, z) : f(x, y, z) = 0, x > 0, y > 0, z > 0\}$ this is

\[
\frac{a}{x} + \frac{b}{x \sqrt{b/a}} + \frac{c}{x \sqrt{c/a}} = \frac{\sqrt{a}}{x} \left( \sqrt{a} + \sqrt{b} + \sqrt{c} \right) = 1
\]

\[
\Rightarrow (x, y, z) = \left( \sqrt{a} + \sqrt{b} + \sqrt{c} \right) \left( \sqrt{a}, \sqrt{b}, \sqrt{c} \right).
\]

Thus the extremum associated to this point is $f(x, y, z) = \left( \sqrt{a} + \sqrt{b} + \sqrt{c} \right)^2$. To see that this is the minimum, the set $S$ is noncompact but the function $g$ must have a minimum on any compact subset of $S$. The function $g$ increases whenever one of the coordinates $x$, $y$, or $z$ becomes very large. Since $S$ is contained entirely in the first quadrant, it is only necessary to minimize $f$ on a reasonable compact subset near the origin, which is the solution found.

When finding the extrema of a continuously differentiable function $g(x)$ on a set such as the closed semidisc

\[
D = \left\{ x \in \mathbb{R}^2 \mid x_1^2 + x_2^2 \leq 1, x_2 \geq 0 \right\}
\]

in the plane, the procedure is first to examine critical points in the interior of the semidisc, then to examine the two pieces of the boundary of the semidisc that are local submanifolds, the semicircle $x_1^2 + x_2^2 = 1$, $x_2 \geq 0$ and the axis $-1 \leq x_1 \leq 1$, $x_2 = 0$, and then the two remaining points ($-1, 0$) and $(1, 0)$ that are points where the boundary fails to be a submanifold. In general, the procedure is to examine critical points in the interior of the region, and then to examine the pieces of the boundary that are submanifolds and apply the method of Lagrange multipliers in all these cases.

If $f : U \rightarrow \mathbb{R}^m$ is a continuously mapping defined in an open subset $U \subset \mathbb{R}^n$ where $m < n$ then rank $f'(a) \leq n$ at any point $a \in U$; and if rank $f'(x) = n$ at a point $x = a \in U$ then it is also the case that rank $f'(x) = n$ for all points $x \in U$ for an open neighborhood $V$ of the point $a$, since if some $n \times n$ submatrix of $f'(a)$ has a nonzero determinant then by continuity that submatrix of $f'(x)$ will also have a nonzero determinant for all points $x$ near enough $a$. The Rank Theorem asserts that the mapping $f$ is equivalent to a linear mapping of rank $n$ in an open neighborhood of each point $a \in V$, and consequently the image of the mapping $f$ is a full open neighborhood of the point $f(a)$. In particular the image of the mapping $f$ is not contained within any proper linear subspace of the image space $\mathbb{R}^n$, so the coordinate functions $f_i$ of the mapping $f$ are linearly independent functions at each point $a \in U$. On the other
3.3. THE RANK THEOREM

hand if \( \text{rank } f(x) = r \) at all points \( x \in U \), then by the Rank Theorem, the mapping \( f \) will be a linear mapping of rank \( r \) in suitable local coordinates in open neighborhoods of each point \( a \in U \subset \mathbb{R}^m \) and its image \( f(a) \subset \mathbb{R}^n \); that means that the image of the restriction of the mapping \( f \) to an open neighborhood \( V_a \) of any point \( a \in U \) is contained in a submanifold of dimension \( r \) in an open neighborhood \( V_f(a) \subset \mathbb{R}^n \) of the image point \( f(a) \). This submanifold is described as the set of common zeros of \( n-r \) linear functions in some local coordinates near the point \( f(a) \), and these functions will at least be some continuously differentiable functions \( h_i \) near the point \( f(a) \) in the initial coordinates in \( \mathbb{R}^n \). Therefore the component functions \( f_i \) of the mapping \( f \) satisfy a collection of \( n-r \) nontrivial equations \( h_i(f_1(x), f_2(x) \ldots, f_n(x)) = 0 \) for \( 1 \leq i \leq n-r \); that is usually expressed by saying that the functions \( f_i \) are functionally dependent.

Submanifolds have a rather more intrinsic structure as well. If \( V \) is a topological submanifold of an open subset \( U \subset \mathbb{R}^m \) it has the relative topology induced by the topology of \( U \), in which the open subsets of \( V \) are the intersections of \( V \) with the open subsets of \( U \). Continuous functions on \( V \) can be defined as usual items of the open subsets of \( V \). If \( V_1 \subset U_1 \) and \( V_2 \subset U_2 \) are topological submanifolds of open subsets \( U_1 \subset \mathbb{R}^m \) and \( U_2 \subset \mathbb{R}^m \) a mapping \( \phi : V_1 \rightarrow V_2 \) is continuous in the relative topologies if and only if \( \phi^{-1}(W) \subset V_1 \) is an open subset of \( V_1 \) whenever \( W \subset V_2 \) is an open subset of \( V_2 \); and the submanifolds \( V_1 \) and \( V_2 \) are weakly homeomorphic if and only if there is a continuous one-to-one mapping \( \phi : V_1 \rightarrow V_2 \) with a continuous inverse. This must be distinguished from the notion that the two submanifolds \( V_1 \) and \( V_2 \) are homeomorphic, as defined on page 39; that refers to the way in which the submanifolds are imbedded in the open subsets \( U_i \), so the submanifolds are equivalent if and only if there is a homeomorphism \( \psi : U_1 \rightarrow U_2 \) such that \( \psi(V_1) = V_2 \). Two submanifolds that are weakly homeomorphic need not be submanifolds of subsets \( U_i \) in spaces \( \mathbb{R}^m \) of the same dimension. For instance, a circle in the plane and a knotted curve in space are generally weakly homeomorphic, when the knotted curve is described as the image of a one-to-one \( C^1 \) mapping of the circle into \( \mathbb{R}^3 \) with a nontrivial derivative; but they are certainly not homeomorphic as submanifolds, since they are imbedded in spaces of different dimensions. Weak equivalence of submanifolds is an equivalence relation in the customary sense; and a weak equivalence class is called a manifold, in this case a topological manifold. A given submanifold \( V \subset U \) in an open subset \( U \subset \mathbb{R}^m \) thus determines a manifold, an equivalence class of submanifolds under weak equivalence that includes the submanifold \( V \). It is perhaps more common to speak of submanifolds being equivalent as manifolds, rather than as weakly equivalent submanifolds.

As might be expected, it is also possible to define manifolds even more intrinsically, without considering their embeddings. If \( V \subset U \) is a \( k \)-dimensional topological submanifold of an open set \( U \subset \mathbb{R}^m \) then by definition each point \( a \in V \) is contained in an open subset \( U_a \subset \mathbb{R}^m \) for which there is a homeomorphism \( \phi_a : U_a \rightarrow W_a \) between the open subset \( U_a \subset \mathbb{R}^n \) and another open subset \( W_a \subset \mathbb{R}^m \) such that \( \phi_a(V \cap U_a) \) is a linear subspace \( L_a \subset W_a \), which can be viewed as an open subset \( L_a \subset \mathbb{R}^k \); the restriction \( \phi_a : V \cap U_a \rightarrow L_a \) is a homeomorphism between the relatively open subset \( V_a = V \cap U_a \subset V \) and the open subset \( L_a \subset \mathbb{R}^k \), which can be considered as describing a system of local coordinates in the relatively open neighborhood \( V_a \) of the point \( a \in V \), the local coordinates in the open subset \( L_a \subset \mathbb{R}^k \). The corresponding construction can be carried out for any other topological submanifold \( V' \subset U' \) that is homeomorphic to \( V \) through a homeomorphism \( \psi : U \rightarrow U' \) such that \( \psi(V) = V' \); if \( a' = \psi(a) \) there is an open subset \( U'_a \) with a homeomorphism \( \phi'_a : U'_a \rightarrow W'_a \) such
that the restriction $\phi_{a'} : V' \cap U'_{a'} \longrightarrow L'_{a'}$ is a homeomorphism between the relatively open subset $V' = V' \cap U'_{a'} \subset V'$ and the open subset $L'_{a'} \subset \mathbb{R}^k$, so the composition $\phi_{a'} \circ \psi : V_p \longrightarrow L'_{a'}$ can be considered as describing another system of local coordinate in the relatively open neighborhood $V_a$ of the point $a \in V$. These two sets of local coordinates near $a$ are related by the homeomorphism $\phi_{a'} \circ \psi \circ (\phi_a)^{-1} : V_a \longrightarrow L'_{a'}$. Thus the system of local coordinates in neighborhoods of a point $a \in V$ provided by equivalent topological $k$-dimensional submanifolds of $\mathbb{R}^n$ are all homeomorphic to one another. In particular if $V_a$ and $V_b$ are any two overlapping systems of local coordinates in $V$ then they determine homeomorphic systems of local coordinates in an open neighborhood of any point $c \in V_a \cap V_b$; thus the manifold $V$ can be described by a collection of relatively open coordinate neighborhoods $V_a \subset V$ which determine equivalent coordinate neighborhoods in any intersection $U_a \cap U_b$. Weak homeomorphisms between two submanifolds representing the same manifold can be described entirely in terms of these local coordinate neighborhoods. The same construction can be carried out for $C^1$ submanifolds, providing systems of local coordinates that are related by $C^1$ homeomorphisms in the intersections of coordinate neighborhoods, and of course also for $C^r$ submanifolds for any $r$. The notion of a manifold can be extended to those defined more abstractly just by coverings of a topological space $V$ by coordinate neighborhoods. Traditionally a $k$-dimensional manifold is defined as a topological space $M$ such that each point $\text{pin} M$ has an open neighborhood $U_p \subset M$ with a one-to-one continuous mapping $f_p : U_p \longrightarrow V_p$ onto an open subset $V_p \subset \mathbb{R}^k$; the sets $U_p$ are called coordinate neighborhoods on the manifold $M$, and the parameters in the image set $V_p \subset \mathbb{R}^k$ are called local coordinates on the manifold $M$ in $U_p$. Two $k$-dimensional manifolds $M$ and $M'$ are homeomorphic if there is a continuous invertible one-to-one mapping $\phi : M \longrightarrow M'$; this is an equivalence relation in the traditional sense, identifying manifolds that are topologically the same. It is customary to impose at least some restrictions on the topology of the set $M$ to avoid odd and irregular cases\footnote{For example the real axis together with one additional point $0'$, where an open neighborhood of the zero $0$ is defined as the set of real numbers $\{ x \in \mathbb{R} \mid |x| \leq \epsilon \}$ and an open neighborhood of the additional point $0'$ is defined to be that point together with the set of real numbers $\{ x \in \mathbb{R} \mid 0 < |x| \leq \epsilon \}$ is a one-dimensional manifold; but the points $0$ and $0'$ do not have disjoint open neighborhoods}; thus the topology is usually assumed to be a Hausdorff topology, defined as a topology such that any two points of $M$ have disjoint open neighborhoods. It is possible to speak of continuous functions on a manifold, since the manifold has a topology. However it is not possible to speak of differentiable functions, since two overlapping coordinate neighborhoods neighborhoods each look like pieces of $\mathbb{R}^k$, so that the notion of a function being differentiable in each of those coordinate neighborhoods makes sense; but the notions of differentiability in the two coordinate neighborhoods may be quite different, since it is only required that the mappings $f_p : U_p \longrightarrow V_p$ be continuous. However if the manifold $M$ has the additional property that in any intersection $U_p \cap U_q$ of coordinate neighborhoods of $M$ the composition

$$f_p^{-1} f_q : f_q(U_p \cap U_q) \longrightarrow f_p(U_p \cap U_q)$$

is a $C^r$ mapping between the open subsets $f_q(U_p \cap U_q) \subset \mathbb{R}^k$ and $f_q(U_p \cap U_q) \subset \mathbb{R}^k$, then the notion of a function being $C^r$ is well defined; a manifold $M$ with this property is called a differentiable manifold of class $C^r$, or just a $C^r$ manifold for short. The Whitney Embedding Theorem\footnote{The general result was proved by H. Whitney, Annals of Mathematics, vol. 37 (1936), pages 645-680. The result is discussed in a number of articles and books, for example in the} shows that any abstractly defined manifold can be
realized by a submanifold of some space $\mathbb{R}^n$, so that it is possible to define manifolds either as equivalence classes of submanifolds or as abstractly defined manifolds.

---

Chapter 4
Integration

4.1 The Riemann Integral

The definition of the Riemann integral over cells in $\mathbb{R}^n$ is essentially the same as the familiar definition of the Riemann integral over intervals in $\mathbb{R}^1$. The content of a closed cell

$$\Delta = \left\{ \mathbf{x} \in \mathbb{R}^n \mid a_j \leq x_j \leq b_j \text{ for } 1 \leq j \leq n \right\}$$

is defined by

$$(4.1) \quad |\Delta| = \prod_{j=1}^{n} (b_j - a_j),$$

and is also considered as the content $|\Delta|$ of the corresponding open cell $\Delta$. The content of a cell or interval in $\mathbb{R}^1$ is its length, the content of a cell in $\mathbb{R}^2$ is its area, and the content of a cell in $\mathbb{R}^3$ is its volume; so the content of a cell is just a generalization of the area or volume to arbitrary dimensions. The cell can be decomposed as a union of smaller closed cells by adjoining finitely many points of subdivision

$$a_j = c_{j,0} < c_{j,1} < \cdots < c_{j,n_j} = b_j$$

of each of its sides and writing $\Delta$ as the union of the closed subcells

$$\left\{ \mathbf{x} \in \mathbb{R}^n \mid c_{j,k-1} \leq x_j \leq c_{j,k} \text{ for } 1 \leq j \leq n, 1 \leq k \leq n_j \right\};$$

such a decomposition $\Delta = \bigcup \Delta_i$ of the cell is called a partition $\mathcal{P}$ of the cell. The subcells of the partition are not disjoint, but only have portions of their boundaries in common. For the case $n = 1$ this is the decomposition of an interval, a cell in $\mathbb{R}^1$, as a union of subintervals; and for $n = 2$ it is illustrated in the accompanying Figure 4.1. It is clear that $|\Delta| = \sum_i |\Delta_i|$ for any partition of a cell $\Delta$. If $f$ is a bounded function on a cell $\Delta \subset \mathbb{R}^n$, its least upper bound and greatest lower bound in the set $\Delta$ are well defined finite real numbers; the upper sum of the function $f$ for the partition $\mathcal{P}$ is defined by

$$(4.2) \quad S^*(f, \mathcal{P}) = \sum_{\Delta_i} \sup_{\mathbf{x} \in \Delta_i} f(\mathbf{x}) |\Delta_i|,$$
and similarly the lower sum is defined by

\[(4.3) \quad S_*(f, \mathcal{P}) = \sum \inf_{x \in \Delta_i} f(x)|\Delta_i|.\]

It is clear that

\[(4.4) \quad S_*(f, \mathcal{P}) \leq S^*(f, \mathcal{P})\]

for any partition \(\mathcal{P}\) of the cell \(\mathbf{\Delta}\). Adding further points of subdivision to the sides of the cell leads to a finer decomposition of the cell as a union of subcells, called a refinement of the partition. If \(\mathcal{P}_2\) is a refinement of a partition \(\mathcal{P}_1\) of a cell \(\Delta\) then

\[(4.5) \quad S^*(f, \mathcal{P}_2) \leq S^*(f, \mathcal{P}_1) \quad \text{and} \quad S_*(f, \mathcal{P}_2) \geq S_*(f, \mathcal{P}_1);\]

for if a cell \(\Delta_i\) in the partition \(\mathcal{P}_1\) is decomposed into a union \(\mathbf{\Delta}_i = \bigcup \mathbf{\Delta}_{ij}\) of cells \(\mathbf{\Delta}_{ij}\) in the partition \(\mathcal{P}_2\) then

\[\sup_{x \in \mathbf{\Delta}_{ij}} f(x) \leq \sup_{x \in \mathbf{\Delta}_i} f(x) \quad \text{and} \quad \sum_j |\mathbf{\Delta}_{ij}| = |\mathbf{\Delta}_i|\]

and consequently

\[S^*(f, \mathcal{P}_2) = \sum_{i,j} \sup_{x \in \mathbf{\Delta}_{ij}} f(x)|\mathbf{\Delta}_{ij}|\]

\[\leq \sum_{i,j} \sup_{x \in \mathbf{\Delta}_i} f(x)|\mathbf{\Delta}_{ij}| = \sum_i \sup_{x \in \mathbf{\Delta}_i} f(x)|\mathbf{\Delta}_i| = S^*(f, \mathcal{P}_1),\]

and correspondingly for the lower sums with the reversed inequalities. It follows that if \(\mathcal{P}_1\) and \(\mathcal{P}_2\) are any two partitions of a cell \(\Delta\) then

\[(4.6) \quad S_*(f, \mathcal{P}_1) \leq S^*(f, \mathcal{P}_2);\]

for by considering all the points of subdivision of the sides of the cell \(\mathbf{\Delta}\) from both partitions as defining a further partition \(\mathcal{P}'\), a common refinement of both of these partitions, then by (4.4) and (4.5)

\[(4.7) \quad S_*(f, \mathcal{P}_1) \leq S_*(f, \mathcal{P}') \leq S^*(f, \mathcal{P}) \leq S^*(f, \mathcal{P}_2).\]
4.1. RIEMANN INTEGRAL

In view of (4.7) the upper sums $S^*(f, \mathcal{P})$ for all of the partitions $\mathcal{P}$ of a cell $\Delta$ are bounded below by any lower sum $S_*(f, \mathcal{P}_1)$, and the lower sums $S_*(f, \mathcal{P})$ for all of the partitions $\mathcal{P}$ of a cell $\Delta$ are bounded above by any upper sum $S^*(f, \mathcal{P}_2)$; so it is possible to define the upper integral and lower integral of the function $f$ over the cell $\Delta$ by

$$
\int_\Delta^* f = \inf_{\mathcal{P}} S^*(f, \mathcal{P}) \quad \text{and} \quad \int_\Delta^- f = \sup_{\mathcal{P}} S_*(f, \mathcal{P}),
$$

where the infimum and supremum are extended over all partitions $\mathcal{P}$ of the cell $\Delta$, and clearly

$$
\int_\Delta^- f \leq \int_\Delta^* f.
$$

The function $f$ is said to be Riemann integrable over the cell $\Delta$ if the upper and lower integrals are actually equal; this common value is called the Riemann integral of the function $f$ over the cell $\Delta$ and is denoted by $\int_\Delta f$, so if $f$ is Riemann integrable over $\Delta$ then

$$
\int_\Delta f = \int_\Delta^- f = \int_\Delta^* f.
$$

It is evident from the definition that any constant function $c$ is Riemann integrable, and that $\int_\Delta c = c |\Delta|$, for any cell $\Delta$. However not all bounded functions are Riemann integrable. For example, if $f$ is the function on the real line $\mathbb{R}$ defined by

$$
f(x) = \begin{cases} 
0 & \text{if } x \text{ is rational}, \\
1 & \text{if } x \text{ is irrational},
\end{cases}
$$

then $\sup_{x \in \Delta} f(x) = 1$ and $\inf_{x \in \Delta} f(x) = 0$ for any partition $\mathcal{P}$ of a cell $\Delta$ into a union $\Delta = \bigcup_i \Delta_i$, and consequently $S^*(f, \mathcal{P}) = |\Delta|$ and $S_*(f, \mathcal{P}) = 0$; and since that is the case for any partition $\mathcal{P}$ it follows that $\int_\Delta^- f = |\Delta|$ and $\int_\Delta^* f = 0$, so the function $f$ is not Riemann integrable over a nontrivial cell $\Delta$. However it is fairly clear that continuous functions are Riemann integrable. Indeed if $f$ is continuous in $\Delta$ then it is uniformly continuous since $\Delta$ is compact, so for any $\epsilon > 0$ there is a partition $\mathcal{P}$ of the cell as a union $\Delta = \bigcup_i \Delta_i$, where the cells $\Delta_i$ are sufficiently small that $\sup_{x \in \Delta_i} f(x) - \inf_{x \in \Delta_i} f(x) \leq \epsilon$ for each cell $\Delta_i$. Then

$$
S^*(f, \mathcal{P}) - S_*(f, \mathcal{P}) = \sum_i \left( \sup_{x \in \Delta_i} f(x) |\Delta_i| - \inf_{x \in \Delta_i} f(x) |\Delta_i| \right) 
\leq \sum_i \epsilon |\Delta_i| = \epsilon |\Delta|;
$$

that is the case for all partitions hence $\int_\Delta^- f - \int_\Delta^* f \leq \epsilon |\Delta|$ for any $\epsilon > 0$, consequently $\int_\Delta^- f = \int_\Delta^* f$ so $f$ is Riemann integrable. This proof can be tweaked to yield a somewhat more general result that is a complete characterization of Riemann integrable functions, a result basically due to Riemann but put into modern terms with simpler proofs by Lebesgue. The oscillation of a bounded function $f$ in a subset $S \subseteq \mathbb{R}^n$ is defined by

$$
o_f(S) = \sup_{x \in S} f(x) - \inf_{x \in S} f(x);
$$
this is well defined, since the supremum and infimum of any bounded collection of real numbers \( \{ f(x) \} \) are well defined. In view of (4.12) it is clear that

\[
(4.14) \quad S^*(f, \mathcal{P}) - S_*(f, \mathcal{P}) = \sum_i o_f(\Delta_i)|\Delta_i|
\]

for any partition \( \mathcal{P} \), so the notion of oscillation is closely related to integrability. It is evident that if \( S_1 \subset S_2 \) then \( o_f(S_1) \leq o_f(S_2) \). In particular if a function \( f \) is defined and bounded in an open neighborhood \( N_r(a) \) of a point \( a \) then \( o_f(N_r(a)) \leq o_f(N_{r_2}(a)) \) whenever \( r_1 \leq r_2 \) and \( r_2 \) is sufficiently small that these neighborhoods are contained in \( U \), so the limit

\[
(4.15) \quad o_f(a) = \lim_{r \to 0} o_f(N_r(a))
\]

is well defined; it is called the oscillation of the function \( f(x) \) at the point \( a \).

**Lemma 4.1** A bounded function \( f \) in an open neighborhood of a point \( a \in \mathbb{R}^n \) is continuous at the point \( a \) if and only if \( o_f(a) = 0 \).

**Proof:** If \( o_f(a) = 0 \) then for any \( \epsilon > 0 \) there is a \( \delta > 0 \) such that \( o_f(N_\delta(a)) < \epsilon \), so in particular \( |f(x) - f(a)| < \epsilon \) whenever \( x \in N_\delta(a) \) and consequently \( f \) is continuous at the point \( a \). Conversely if \( f \) is continuous at \( a \) then for any \( \epsilon > 0 \) there is a \( \delta > 0 \) such that \( |f(x) - f(a)| < \frac{\epsilon}{2} \) whenever \( x \in N_\delta(a) \); therefore \( |f(a)| - \frac{\epsilon}{2} \leq |f(x)| \leq |f(a)| + \frac{\epsilon}{2} \) for any \( x \in N_\delta(a) \), hence \( o_f(N_\delta(a)) < \epsilon \) and consequently \( o_f(a) = 0 \). That suffices for the proof.

**Lemma 4.2** If \( f(x) \) is a bounded function in a subset \( S \subset \mathbb{R}^n \) then

\[
E = \left\{ x \in S \mid o_f(x) < \epsilon \right\}
\]

is a relatively open subset of \( S \) for any \( \epsilon > 0 \).

**Proof:** If \( a \in E \) then \( o_f(a) < \epsilon \), so by definition \( o_f(N_\delta(a) \cap E) < \epsilon \) if \( \delta \) is sufficiently small. If \( b \in N_{\frac{\epsilon}{3}}(a) \cap E \) then \( N_{\frac{\epsilon}{3}}(b) \cap E \subset N_\delta(a) \cap E \subset \mathbb{R}^n \) so \( o_f(N_{\frac{\epsilon}{3}}(b) \cap E) \leq o_f(N_\delta(a) \cap E) < \epsilon \) and therefore \( N_{\frac{\epsilon}{3}}(b) \cap E \subset E \). That shows that \( E \) is a relatively open subset of \( S \) and thereby concludes the proof.

A subset \( S \subset \mathbb{R}^n \) has **content zero** if for any \( \epsilon > 0 \) there are **finitely** many open cells \( \Delta_i \) such that \( S \subset \bigcup \Delta_i \) and \( \sum_i |\Delta_i| < \epsilon \); the set \( S \) has **measure zero** if for any \( \epsilon > 0 \) there are **countably** many open cells \( \Delta_i \) such that \( S \subset \bigcup \Delta_i \) and \( \sum_i |\Delta_i| < \epsilon \). Clearly any set of content zero is also a set of measure zero; and a compact set of measure zero is also of content zero, since if \( S \) is compact and \( S \subset \bigcup \Delta_i \) where \( \sum_i |\Delta_i| < \epsilon \) then \( S \) is already contained in the union of finitely many of these cells \( \Delta_i \).

**Lemma 4.3** If countably many subsets \( E_i \subset \mathbb{R}^n \) each have measure zero then their union \( E = \bigcup_{i=1}^\infty E_i \) also has measure zero.

**Proof:** Since \( E_i \) has measure zero then for any \( \epsilon > 0 \) there are countably many open cells \( \Delta_{ij} \) for \( j = 1, 2, \ldots \) such that \( E_i \subset \bigcup_{j=1}^\infty \Delta_{ij} \) and \( \sum_{j=1}^\infty |\Delta_{ij}| < \frac{\epsilon}{2} \); then \( E = \bigcup_{i=1}^\infty E_i \subset \bigcup_{i=1}^\infty \bigcup_{j=1}^\infty \Delta_{ij} \), which is also a countable union of open cells, and \( \sum_{i=1}^\infty \sum_{j=1}^\infty |\Delta_{ij}| = \sum_{i=1}^\infty \sum_{j=1}^\infty |\Delta_{ij}| \leq \sum_{i=1}^\infty \frac{\epsilon}{2} = \epsilon \), which suffices for the proof.
4.1. RIEMANN INTEGRAL

For example, the set \( \mathbb{Q} \) of rational numbers does not have content zero, since no finite number of finite open intervals can cover all the rational numbers; but it does have measure zero as a consequence of the preceding lemma, since each individual point of course has measure zero. There are noncountable sets of measure zero as well, for any proper submanifold of an open subset \( U \subset \mathbb{R}^n \) is a set of measure zero and is a noncountable set if its dimension is greater than 0. Indeed if \( V \subset U \) is a submanifold of dimension \( k \) where \( 1 \leq k < n \) then by Corollary 3.4 for any point \( a \in V \) after relabeling the coordinates if necessary the intersection \( V \cap \Delta \) of the submanifold \( V \) with a sufficiently small cell \( \Delta_a = \Delta_a' \times \Delta_a'' \) centered at the point \( a \), where \( \Delta_a' \) is in the subspace \( \mathbb{R}^{n-k} \) consisting of the first \( n-k \) coordinates \( x_1, \ldots, x_{n-k} \) and \( \Delta_a'' \) is in the subspace \( \mathbb{R}^k \) consisting of the last \( k \) coordinates \( x_{n-k+1}, \ldots, x_n \), is described by the equations

\[
V \cap \Delta_a = \left\{ x \in U \mid x_i = g_i(x_{n-k+1}, \ldots, x_n) \text{ for } 1 \leq i \leq n-k \right\}
\]

for some continuously differentiable functions \( g_i(x') \) of the variables \( x' \in \Delta_a' \), as in the accompanying Figure 4.2. Each point \( x \in V \) can be enclosed in a small cell \( \Delta_1 = \Delta_1' \times \Delta_1'' \) where \( |\Delta_1| < \epsilon \), so that \( \sum |\Delta_i| < \epsilon |\Delta''| \), from which it is evident that the set \( V \cap \Delta_a \) is of measure 0. The entire submanifold \( V \) can be covered by countably many such cells \( \Delta_a \), so by Lemma 4.3 the set \( V \) is also of measure 0.

Figure 4.2: Submanifold of a cell, a set of measure 0 \( \Delta \)

A function \( f \) is **continuous almost everywhere** in a set \( S \subset \mathbb{R}^n \) if the subset \( E \subset S \) of points where \( f \) fails to be continuous is a set of measure zero. For example, the function \( f \) on \( \mathbb{R} \) defined by

\[
f(x) = \begin{cases} 
\frac{1}{q} & \text{if } x = \frac{p}{q} \text{ where } p, q \text{ are coprime integers}, \\
0 & \text{if } x \text{ is irrational},
\end{cases}
\]

(4.16)

is continuous at the irrational numbers but is discontinuous at any rational number, so it is continuous almost everywhere. It is an amusing point, but a bit harder to prove, that there are no functions that are continuous at the rational numbers but discontinuous at any irrational number\(^1\).

---

\(^1\)This result follows from an application of the Baire Category Theorem, which is discussed in most of the standard text books on real analysis.
CHAPTER 4. INTEGRATION

Theorem 4.1 (Theorem of Lebesgue) A bounded function in a cell $\overline{\Delta} \subset \mathbb{R}^n$ is Riemann integrable over $\overline{\Delta}$ if and only if it is continuous almost everywhere in $\overline{\Delta}$.

Proof: As preliminary observations, for any bounded function $f$ in a cell $\overline{\Delta}$ and for any $\epsilon > 0$ let

$$E_{\epsilon} = \{ x \in \overline{\Delta} | f(x) < \epsilon \} \quad \text{and} \quad F_{\epsilon} = \{ x \in \overline{\Delta} | f(x) \geq \epsilon \},$$

so $F_{\epsilon} = \overline{\Delta} \sim E_{\epsilon}$. By Lemma 4.2 the set $E_{\epsilon}$ is relatively open in $\overline{\Delta}$, so its complement $F_{\epsilon}$ is a relatively closed subset of $\overline{\Delta}$, and since $\overline{\Delta}$ is compact it follows that the closed subset $F_{\epsilon} \subset \overline{\Delta}$ is compact. By Lemma 4.1 the function $f$ is continuous at a point $a \in \overline{\Delta}$ if and only if $f(a) = 0$; so the set of points in $\overline{\Delta}$ at which $f$ is not continuous is the set $F = \{ x \in \overline{\Delta} | f(x) > 0 \}$, and $F_{\epsilon} \subset F$ for any $\epsilon > 0$ while $F = \bigcup_{n=1}^{\infty} F_{1/n}$.

To turn to the proof itself, first suppose that the function $f$ is bounded and continuous almost everywhere in $\overline{\Delta}$. The set $F$ then is of measure zero, as are the subsets $F_{\epsilon}$ and since the sets $F_{\epsilon}$ are compact, they even are of content zero. Therefore there are finitely many open cells $\Delta_j \subset \mathbb{R}^n$ such that $F_{\epsilon} \subset \bigcup_j \Delta_j$ and $\sum_j |\Delta_j| < \epsilon$. The complement $K = \overline{\Delta} \sim \bigcup_j \Delta_j$ is a closed and consequently compact subset of $\overline{\Delta}$, and if $a \in K$ then $a \notin F_{\epsilon}$ so $f(a) < \epsilon$; hence for any point $a \in K$ there is an open cell $\Delta''_{jk}$ containing that point such that $f(\Delta'_{jk}) < \epsilon$. Since $K$ is compact finitely many of these open cells $\Delta''_{jk}$ serve to cover the compact set $K$. Let $P$ be the partition of $\overline{\Delta}$ determined by the points of subdivision of the finitely many cells $\Delta_j$ and $\Delta''_{jk}$. Let $I''$ be the set of those indices $i$ such that the cell $\Delta_i$ is contained within some of the cells $\Delta_j$, and let $I''$ be the remaining set of indices. Thus $F_{\epsilon} \subset \bigcup_{i \in I''} \Delta_i$ and $\sum_{i \in I''} |\Delta_i| < \epsilon$, while if $i \in I'$ then $\Delta_i \subset \Delta''_{jk}$ for some of the cells $\Delta''_{jk}$ so $f(\Delta_i) < \epsilon$. If $|f(x)| \leq M$ for all points $x \in \overline{\Delta}$ then $f(D) \leq 2M$ for any subset $D \subset \overline{\Delta}$ and consequently

$$S^*(f, P) - S_*(f, P) = \sum_{i \in I''} f(\overline{\Delta}_i) |\overline{\Delta}_i| = \sum_{i \in I'} f(\overline{\Delta}_i) |\overline{\Delta}_i| + \sum_{i \in I''} f(\overline{\Delta}_i) |\overline{\Delta}_i|$$

$$\leq 2M \sum_{i \in I'} |\overline{\Delta}_i| + \epsilon \sum_{i \in I''} |\overline{\Delta}_i| \leq 2M \epsilon + |\overline{\Delta}_i| \epsilon.$$

Since that is the case for any $\epsilon$ it follows that the function $f$ is Riemann integrable.

Next suppose that the function $f$ is Riemann integrable over the cell $\overline{\Delta}$. For any $\eta > 0$ there is a partition $P$ of the cell $\overline{\Delta}$ such that

$$\sum_i f(\Delta_i) |\Delta_i| = S^*(f, P) - S_*(f, P) < \eta.$$

Let $I'$ be the set of those indices $i$ such that $\Delta_i \cap F_{\epsilon} \neq \emptyset$ and let $I''$ be the remaining indices; thus $F_{\epsilon} \subset \bigcup_{i \in I'} \overline{\Delta}_i$, and if $i \in I'$ there is a point $a_i \in \Delta_i$ such that $f(a_i) \geq \epsilon$, and consequently $f(\Delta_i) \geq \epsilon$. Then

$$\eta > \sum_{i \in I'} f(\Delta_i) |\Delta_i| + \sum_{i \in I''} f(\Delta_i) |\Delta_i| \geq \sum_{i \in I'} f(\Delta_i) |\Delta_i| \geq \sum_{i \in I'} \epsilon |\Delta_i|$$

so that $\sum_{i \in I'} |\Delta_i| < \eta/\epsilon$, and since $F_{\epsilon} \subset \bigcup_{i \in I'} \overline{\Delta}_i$ for all $\eta > 0$ it follows that $F_{\epsilon}$ is a set of measure zero. Finally since $F \subset \bigcup_{n=1}^{\infty} F_{1/n}$ and all the sets $F_{1/n}$ are of measure zero the set $F$ also is of measure zero, which suffices for the proof.

The preceding characterization of Riemann integrable functions is expressed in terms of basic properties of the function itself, rather than in terms of the construction.
of the Riemann integral, so it is in many ways a more useful characterization. Consequently a function defined in a subset \( S \subset \mathbb{R}^n \) will be called **integrable** if it is bounded and continuous almost everywhere in \( S \); thus a function is integrable in a cell if and only if it is Riemann integrable in that cell. However it is possible to extend the notion of the integral of an integrable function to more general sets than just cells. A bounded subset \( D \subset \mathbb{R}^n \) is called a **Jordan domain** if its boundary \( \partial D \) is a set of measure zero; since \( \partial D \) is a closed bounded and hence compact set, then actually \( \partial D \) is a set of content zero for any Jordan domain \( D \). This definition does not impose any further restrictions on the set \( D \); so a Jordan domain may be open or closed or even a more general point set\(^2\). An open or closed cell \( \Delta \) is a Jordan domain, since as noted earlier \( \partial \Delta \) is a set of measure zero for any cell \( \Delta \). The intersection and union of two Jordan domains are again Jordan domains. To verify that first for the union, if \( D_1 \) and \( D_2 \) are Jordan domains then \( \overline{D_1 \cup D_2} \subset \overline{D_1} \cup \overline{D_2} \) and \( \sim (D_1 \cup D_2) = \sim D_1 \cap \sim D_2 \subset \sim D_1 \cap \sim D_2 \) and consequently \( \partial (D_1 \cup D_2) \subset (\overline{D_1} \cap \overline{D_2}) \cup (\overline{D_1} \cap \overline{D_2}) \), which is a set of measure zero since \( \overline{D_1} \cap \overline{D_2} \) and \( \overline{D_1} \cap \overline{D_2} \) are sets of measure zero. The corresponding argument shows that the intersection of Jordan domains is also a Jordan domain. If \( D \subset \overline{\Delta} \) is a Jordan domain then its complement \( \overline{\Delta} \sim D \) is also a Jordan domain since \( \partial (\overline{\Delta} \sim D) = \partial \overline{\Delta} \cup \partial D \).

A Jordan domain \( D \subset \mathbb{R}^n \) is a bounded subset of \( \mathbb{R}^n \) by definition, so \( D \subset \overline{\Delta} \) for some cell \( \Delta \subset \mathbb{R}^n \). An integrable function \( f \) in the Jordan domain \( D \) can be extended to a function \( \tilde{f} \) in the cell \( \overline{\Delta} \) by setting

\[
\tilde{f}(x) = \begin{cases} 
  f(x) & \text{if } x \in D \\
  0 & \text{if } x \in \overline{\Delta} \sim D.
\end{cases}
\]

The extended function \( \tilde{f} \) of course is bounded in \( \overline{D} \). It is continuous almost everywhere in the interior of \( D \), since the function \( f \) is continuous almost everywhere in \( D \) by assumption, and it is also continuous in the complement \( \overline{\Delta} \sim \overline{D} \) of the closure of \( D \), where it vanishes identically by definition; and since \( \partial D \) is a set of measure 0, by the definition of a Jordan domain, it follows that the function \( \tilde{f} \) is also continuous almost everywhere in \( \overline{\Delta} \). Consequently the function \( \tilde{f} \) is integrable in the cell \( \Delta \), so it is possible to define the integral of the initial function \( f \) over the Jordan domain \( D \) as

\[
\int_D f = \int_{\overline{\Delta}} \tilde{f}.
\]

It is evident from the definition of the Riemann integral that the value of the integral \( \int_D f \) is independent of the choice of the cell \( \overline{\Delta} \); for if \( \overline{\Delta} \subset \overline{\Delta'} \) then in calculating the integral \( \int_{\overline{\Delta'}} \tilde{f} \) it is always possible to take partitions such that each cell \( \overline{\Delta'}_i \) in the partition is either contained in \( \overline{\Delta} \) or is disjoint from \( \overline{\Delta} \) or is one of a finite number of cells containing \( \partial \Delta \) and of total content less than any given \( \epsilon \). As a special case, if \( f \) is an integrable function in a closed cell \( \Delta \) its restriction \( f|D \) to a Jordan domain \( D \subset \overline{\Delta} \) is an integrable function in \( D \) and the extension \( f|D \) can be identified with the product \( \chi_D f \), where \( \chi_D \) is the **characteristic function** of the set \( D \), the function

\(^2\)It is more common to call such sets Jordan measurable sets; but since the interest here is not in Jordan measure as such the modified terminology seems more appropriate. The term domain is sometimes used just to denote an open set, but it is convenient to use that term somewhat more generally.
defined by

\[ \chi_D(x) = \begin{cases} 
  1 & \text{if } x \in D, \\
  0 & \text{if } x \notin D.
\end{cases} \]

Thus if \( f \) is an integrable function in a closed cell \( \overline{\Delta} \) then its integral over any Jordan domain \( D \subset \overline{\Delta} \) can be defined equivalently by

\[ \int_D f = \int_{\overline{\Delta}} \chi_D f. \]

In particular the content of a Jordan domain \( D \subset \Delta \) can be defined as the integral of the constant function 1 over the set \( D \), so by the integral

\[ |D| = \int_{\overline{\Delta}} \chi_D. \]

When \( D = \overline{\Delta} \) this definition coincides with the previous definition of the content of a cell, since for any partition \( \mathcal{P} \) of a cell \( \overline{\Delta} \) it follows immediately from the definition of the upper sum that \( S^*(\chi_{\overline{\Delta}}, \mathcal{P}) = \sum_i |\Delta_i| = |\overline{\Delta}| \) and consequently that \( \int_{\overline{\Delta}} \chi_{\overline{\Delta}} = |\overline{\Delta}| \).

**Theorem 4.2** If \( f_1 \) and \( f_2 \) are integrable functions in a Jordan domain \( D \) then so is any linear combination \( f = c_1 f_1 + c_2 f_2 \) for real constants \( c_1, c_2 \) and

\[ \int_D f = c_1 \int_D f_1 + c_2 \int_D f_2. \]

**Proof:** It is clear that any linear combination of functions that are bounded and continuous almost everywhere in a Jordan domain \( D \) is also bounded and continuous almost everywhere in \( D \), so any linear combination of integrable functions over \( D \) is integrable over \( D \). The integrals over \( D \) of integrable functions are defined as the integrals over any cell \( \Delta \) containing \( D \) of the extended functions \( 4.17 \); so it is sufficient just to prove the theorem for linear combinations of integrable functions over a cell \( \Delta \). It is also clear from the definition of the Riemann integral over a cell that the integral of the product \( cf \) of a constant \( c \) and an integrable function \( f \) is the product of \( c \) and the integral of \( f \); so finally it is sufficient just to show that the integral over any cell \( \Delta \) of the sum of two integrable functions is the sum of their integrals. If \( \overline{\Delta} = \bigcup \Delta_i \) is a partition \( \mathcal{P} \) of the cell \( \overline{\Delta} \) then for any bounded functions \( f_1 \) and \( f_2 \)

\[ \sup_{x \in \Delta_i} (f_1(x) + f_2(x)) \leq \sup_{x \in \Delta_i} f_1(x) + \sup_{x \in \Delta_i} f_2(x); \]

therefore \( S^*(f_1 + f_2, \mathcal{P}) \leq S^*(f_1, \mathcal{P}) + S^*(f_2, \mathcal{P}), \) and since that is the case for any partition it follows that \( \int_{\overline{\Delta}} (f_1 + f_2) \leq \int_{\overline{\Delta}} f_1 + \int_{\overline{\Delta}} f_2 \). The corresponding argument for the lower sums shows that \( \int_{\overline{\Delta}} (f_1 + f_2) \leq \int_{\overline{\Delta}} f_1 + \int_{\overline{\Delta}} f_2 \). Altogether then

\[ \int_{\overline{\Delta}} f_1 + \int_{\overline{\Delta}} f_2 \leq \int_{\overline{\Delta}} (f_1 + f_2) \leq \int_{\overline{\Delta}} (f_1 + f_2) \leq \int_{\overline{\Delta}} f_1 + \int_{\overline{\Delta}} f_2; \]

and since \( f_1 \) and \( f_2 \) are integrable \( \int_{\overline{\Delta}} f_1 = \int_{\overline{\Delta}} f_1 \) and \( \int_{\overline{\Delta}} f_2 = \int_{\overline{\Delta}} f_2 \) so it follows from the preceding inequalities that

\[ \int_{\overline{\Delta}} (f_1 + f_2) = \int_{\overline{\Delta}} f_1 + \int_{\overline{\Delta}} f_2. \]
which suffices for the proof.

Integrability is really required for the preceding result to hold. For the function (4.11) for example \( \int_{\Delta} f = |\Delta| \) and \( \int_{\bar{\Delta}} f = 0 \), and it is clear that the same argument applied to the function \( g(x) = 1 - f(x) \) shows that \( \int_{\Delta} g = |\Delta| \) and \( \int_{\bar{\Delta}} g = 0 \); consequently \( \int_{\Delta} f + \int_{\bar{\Delta}} g = 2|\Delta| \) while \( \int_{\bar{\Delta}} f + \int_{\bar{\Delta}} g = 0 \). On the other hand \( f + g = 1 \) so \( \int_{\Delta} (f + g) = \int_{\Delta} (f + g) = \int_{\Delta} (f + g) = |\Delta| \). Thus it is neither true that \( \int_{\Delta} (f + g) = \int_{\bar{\Delta}} f + \int_{\bar{\Delta}} g \) nor that \( \int_{\Delta} (f + g) = \int_{\Delta} f + \int_{\Delta} g \).

**Theorem 4.3** Let \( f \) and \( g \) be integrable functions over a Jordan domain \( D \).

(i) If \( f(x) \geq 0 \) for all \( x \in D \) then \( \int_{D} f \geq 0 \).

(ii) If \( f(x) \geq g(x) \) for all \( x \in D \) then \( \int_{D} f \geq \int_{D} g \).

(iii) \( \int_{D} f \leq \int_{D} |f| \).

**Proof:** Again it is sufficient just to prove these assertions for integrable functions over a cell \( \Delta \). If \( f(x) \geq 0 \) for all \( x \in \Delta \) then \( S(f, \mathcal{P}) \geq 0 \) for any partition \( \mathcal{P} \) of the cell \( \Delta \) so (i) holds. Then (ii) follows by applying (i) to the difference \( f - g \) and using the preceding Theorem 4.2, while (iii) follows by applying (ii) to the functions \( |f| \) and \( \pm f \) since \( |f(x)| \geq f(x) \) and \( |f(x)| \geq -f(x) \) for all \( x \in \Delta \), and that suffices for the proof.

A mapping \( f : D \to \mathbb{R}^n \) from a Jordan domain \( D \subset \mathbb{R}^m \) into \( \mathbb{R}^n \) is said to be **integrable** if each of its coordinate functions \( f_i \) is integrable over \( D \), and the integral \( \int_{D} f \) is defined to be the vector having the coordinates \( \int_{D} f_i \). The two preceding theorems extend quite directly from functions to mappings, with the appropriate modification for Theorem 4.3.

**Theorem 4.4** (i) Any linear combination \( f = c_1 f_1 + c_2 f_2 \) of integrable mappings \( f_1 \) and \( f_2 \) from a Jordan domain \( D \subset \mathbb{R}^m \) into \( \mathbb{R}^n \) is integrable and

\[
\int_{D} (c_1 f_1 + c_2 f_2) = c_1 \int_{D} f_1 + c_2 \int_{D} f_2.
\]

(ii) If \( f \) is an integrable mapping from a Jordan domain \( D \subset \mathbb{R}^m \) into \( \mathbb{R}^n \) then

\[
\left\| \int_{D} f \right\| \leq \int_{D} \left\| f \right\|.
\]

**Proof:** (i) This is an immediate consequence of Theorem 4.2, since it just involves the individual coordinate functions of the mapping \( f \).

(ii) If \( u \in \mathbb{R}^n \) is a unit vector in the direction of the vector \( \int_{D} f \in \mathbb{R}^n \) then it follows from the Cauchy-Schwarz inequality and Theorems 4.2 and 4.3 for the function \( u \cdot f \) that \( \left\| \int_{D} f \right\| = |u \cdot \int_{D} f| = |\int_{D} u \cdot f| \leq \int_{D} |u \cdot f| \leq \int_{D} \left\| f \right\| \), which suffices for the proof.

**Theorem 4.5** If \( f \) is a bounded function in a Jordan domain \( D \subset \mathbb{R}^n \) such that \( f(x) = 0 \) for all points \( x \in (D \sim E) \), where \( E \subset D \) is a set of content 0, then \( f \) is integrable over \( D \) and \( \int_{D} f = 0 \).

**Proof:** Since \( E \) is of content 0 the function \( f \) clearly is integrable in \( D \). The Jordan domain \( D \) is bounded, so \( \bar{D} \subset \Delta \) for some cell \( \Delta \); the extended function \( f \) on \( \Delta \) then also is zero for points \( x \in (\Delta \sim E) \), so it is integrable over \( \Delta \), as is its absolute value \( |f| \). Since \( E \) has content 0 then for any \( \epsilon > 0 \) it is possible to find a partition \( \mathcal{P} \) of the cell \( \Delta \) as a union \( \Delta = \bigcup_{i \in I'} \Delta_i \) so that \( E \subset \bigcup_{i \in I'} \Delta_i \), for a subset \( I' \subset I \) of the
indices, where \( \sum_{i \in I'} |\Delta_i| < \epsilon \), while \( \Delta \cup E = \emptyset \) if \( i \in I'' \) for the complementary set of indices \( I'' \). Then \( |f(x)| < M \) for all points \( x \in \Delta_i \) if \( i \in I' \), where \( M \) is the bound of the function \( f \), while \( |\tilde{f}(x)| = 0 \) for all points \( x \in \Delta_i \) if \( i \in I'' \) and consequently

\[
0 \leq \int_\Delta |\tilde{f}| = \int_\Delta |\tilde{f}| \leq S^*(|\tilde{f}|, P) \leq \sum_{i \in I'} M |\Delta_i| \leq M \epsilon.
\]

Since that is the case for any \( \epsilon > 0 \) it follows that \( \int_\Delta |\tilde{f}| = 0 \) so \( \int_D f = \int_\Delta \tilde{f} = 0 \) by Theorem 4.3 (iii), which suffices for the proof.

The preceding result fails if the exceptional set is only of measure 0 rather than of content 0; for a function \( f \) such that \( f(x) = 1 \) at a countable dense set of points of a Jordan domain \( D \) and \( f(x) = -1 \) at another countable dense set of points of \( D \) while \( f(x) = 0 \) otherwise does have the property that it vanishes outside of a set of measure 0, but it is not integrable over \( D \).

**Corollary 4.1** If \( f \) is an integrable function over a Jordan domain \( D \), and if \( g \) is a bounded function in \( D \) such that \( g(x) = f(x) \) at all points \( x \in (D \sim E) \) for a subset \( E \subset D \) of content zero, then \( g \) is also integrable over \( D \) and \( \int_D g = \int_D f \).

**Proof:** The integrable function \( f \) is bounded and continuous in the complement of a set \( E_0 \) of measure 0 in \( D \), hence the function \( g \) is bounded and continuous in the complement of the set \( E \cup E_0 \) of measure 0 in \( D \), so \( g \) is integrable in \( D \). Moreover \( g = f + (g - f) \) where \( g - f \) is integrable and vanishes outside the set \( E \) of content 0, so by the preceding theorem \( \int_D (g - f) = 0 \) and therefore \( \int_D g = \int_D f \), which suffices for the proof.

A consequence of the preceding corollary is that an integrable function \( f \) in a Jordan domain \( D \) can be modified on any subset of content 0 in \( D \) in any arbitrary way, so long as it remains bounded, and the modified function remains integrable with the same integral as the function \( f \). The integral of a function \( f \) thus is not so much a property of the individual function \( f \), but rather is a property of an equivalence class of functions where two integrable functions in \( D \) are equivalent if and only if they are equal outside a set of content 0 in \( D \). In particular when considering integrals or integrability of functions over cells or Jordan domains the values of a function at the boundary of the set are immaterial, provided of course that the function remains bounded; so as far as integration is concerned, it is immaterial whether closed or open Jordan domains are considered, and of course the same is true for cells.

**Theorem 4.6** If two Jordan domains \( D_1, D_2 \) are disjoint except possibly for a subset

\[
E = D_1 \cap D_2 \subset \partial D_1 \cup \partial D_2
\]

and if \( f \) is an integrable function in the union \( D_1 \cup D_2 \) then

\[
\int_{D_1 \cup D_2} f = \int_{D_1} f + \int_{D_2} f;
\]

in particular

\[
|D_1 \cup D_2| = |D_1| + |D_2|.
\]
4.2. FUBINI'S THEOREM

Proof: The union $D_1 \cup D_2$ of two Jordan domains $D_1$, $D_2$ is also a Jordan domain, as noted on page 59, so it is contained in a cell $\Delta$; and then $\int_{D_1 \cup D_2} f = \int_{\Delta} \chi_{D_1 \cup D_2} f$.

The characteristic function of the union of the two Jordan domains is $\chi_{D_1 \cup D_2}(x) = \chi_{D_1}(x) + \chi_{D_2}(x) - \chi_E(x)$, where the intersection $E = D_1 \cap D_2$ is also a Jordan domain; consequently

$$\int_{D_1 \cup D_2} f = \int_{\Delta} (\chi_{D_1} + \chi_{D_2} - \chi_E)f$$

$$= \int_{\Delta} \chi_{D_1} f + \int_{\Delta} \chi_{D_2} f - \int_{\Delta} \chi_E f$$

$$= \int_{D_1} f + \int_{D_2} f$$

since $E \subseteq (\partial D_1 \cup \partial D_2)$ and the sets $\partial D_1$ and $\partial D_2$ are sets of measure 0 by the definition of a Jordan domain so that $|E| = 0$ and therefore $\int_{\Delta} \chi_E f = 0$ by Theorem 4.5. That demonstrates (4.26), which holds in particular for $f = 1$ and therefore implies (4.27), thereby concluding the proof.

A subset $D \subseteq \Delta$ has content zero if and only if $\int_{\Delta} \chi_D = 0$, since for any partition $\mathcal{P}$ of the cell $\Delta$ as the union $\Delta = \bigcup_{i \in I} \Delta_i$, it is clear that

$$S^*(f, \mathcal{P}) = \sum_{\{i \in I \mid D \cap \Delta_i \neq \emptyset\}} |\Delta_i|.$$  

That suggests that the notion of the content of a Jordan domain might be extended to the notion of the content of an arbitrary subset $D \subseteq \Delta$ by setting $|D| = \int_{\Delta} \chi_D$. However that does lead to some problems, since as already observed the linearity of integration requires that functions be integrable; so it is not necessarily the case that $|D_1 \cup D_2| = |D_1| + |D_2|$ for general disjoint subsets $D_1, D_2 \subseteq \Delta$. The Lebesgue integral does provide an extension of the notion of the content to more general sets than Jordan domains; but there are limits even there to the additivity of the sizes of arbitrary sets.

4.2 Fubini’s Theorem

The explicit calculation of integrals in several variables can be reduced to a succession of calculations of integrals in a single variable. A cell $\Delta \in \mathbb{R}^n = \mathbb{R}^{n'} \times \mathbb{R}^{n''}$ where $n = n' + n''$ can be written as a product $\Delta = \Delta' \times \Delta''$ of cells $\Delta' \subseteq \mathbb{R}^{n'}$ and $\Delta'' \subseteq \mathbb{R}^{n''}$, so a function $f$ in the cell $\Delta \subseteq \mathbb{R}^n$ can be written as a function $f(x', x'')$ of variables $x' \in \Delta' \subseteq \mathbb{R}^{n'}$ and $x'' \in \Delta'' \subseteq \mathbb{R}^{n''}$. For any fixed point $x'' \in \Delta''$ the function $f(x', x'')$ can be considered as a function of the variable $x' \in \mathbb{R}^{n'}$ alone; so if $f$ is bounded it is possible to consider the upper integral of this function of the variable $x'$, denoted by $\int_{x' \in \Delta'} f(x', x'')$; this upper integral then is a well defined bounded function of the variable $x'' \in \Delta''$, so it is possible to consider the upper integral of this function of the variable $x''$, denoted by $\int_{x'' \in \Delta''} \left( \int_{x' \in \Delta'} f(x', x'') \right)$ and called an iterated integral. Of course it is possible to use the lower rather than upper integral for one or another or both of these integrals, leading to iterated integrals such as

$$\int_{x'' \in \Delta''} \left( \int_{x' \in \Delta'} f(x', x'') \right);$$

and it is possible to proceed in the other order, integrating first with respect to the variable $x''$ and then the variable $x'$, leading to iterated integrals such as $\int_{x' \in \Delta'} \left( \int_{x'' \in \Delta''} f(x', x'') \right)$.
Theorem 4.7 (Fubini’s Theorem) If $f$ is an integrable function in a product cell $\Delta = \Delta' \times \Delta''$ the integrals $\int_{x'' \in \Delta''} f(x', x'')$ and $\int_{x' \in \Delta'} f(x, x'')$ are integrable functions of the variable $x'' \in \Delta''$ and

$$\int_{\Delta} f = \int_{x'' \in \Delta''} \left( \int_{x' \in \Delta'} f(x', x'') \right).$$

Proof: If $P'$ is a partition $\Delta' = \bigcup_i \Delta_i'$ of the cell $\Delta'$ and $P''$ is a partition $\Delta'' = \bigcup_j \Delta_j''$ of the cell $\Delta''$ these two separate partitions determine a partition $P = P' \times P''$ of the cell $\Delta$ and

$$S^*(f, P) = \sum_{i, j} \sup_{x' \in \Delta_i', x'' \in \Delta_j''} f(x', x'') \left| \Delta_i' \times \Delta_j'' \right|$$

$$= \sum_j \left( \sum_i \sup_{x' \in \Delta_i', x'' \in \Delta_j''} f(x', x'') \left| \Delta_i' \right| \right) \left| \Delta_j'' \right|.$$

For any fixed point $x''_0 \in \Delta_j''$

$$\sum_i \sup_{x' \in \Delta_i'} f(x', x''_0) \left| \Delta_i' \right| \geq \sum_{i, j} \sup_{x' \in \Delta_i', x'' \in \Delta_j''} f(x', x'') \left| \Delta_i' \times \Delta_j'' \right| = S^* \left( f(x', x''_0), P' \right) \geq \int_{x' \in \Delta'} f(x', x''_0).$$

That is the case for any fixed point $x''_0 \in \Delta_j''$ so the same results hold for the supremum over all such points, hence

$$S^*(f, P) \geq \sum_j \sup_{x''_0 \in \Delta_j''} \int_{x' \in \Delta'} f(x', x''_0) \left| \Delta_j'' \right| = S^* \left( f(x', x''_0), P'' \right) \geq \int_{x'' \in \Delta''} \left( \int_{x' \in \Delta'} f(x', x'') \right);$$

consequently

$$\int_{\Delta} f \geq \int_{x'' \in \Delta''} \left( \int_{x' \in \Delta'} f(x', x'') \right).$$

The corresponding argument for the lower integral leads to the reversed inequality, so altogether there is the chain of inequalities

$$\int_{\Delta} f \geq \int_{x'' \in \Delta''} \left( \int_{x' \in \Delta'} f(x', x'') \right) \geq \int_{x'' \in \Delta''} \left( \int_{x' \in \Delta'} f(x', x'') \right) \geq \int_{x' \in \Delta'} f(x', x'').$$

Since the function $f$ is integrable $\int_{\Delta} f = \int_{\Delta'} f$ and consequently all of the preceding inequalities must be equalities, which concludes the proof.

Corollary 4.2 If $f$ is a continuous function in a product cell $\Delta = \Delta' \times \Delta''$ then

$$\int_{\Delta} f(x) = \int_{x'' \in \Delta''} \left( \int_{x' \in \Delta'} f(x', x'') \right) = \int_{x' \in \Delta'} \left( \int_{x'' \in \Delta''} f(x', x'') \right).$$
4.2. FUBINI'S THEOREM

Proof: If $f$ is continuous in $\overline{\Delta}$ then for each fixed point $x'' \in \overline{\Delta}$ the function $f(x', x'')$ is a continuous function of $x' \in \overline{\Delta}$ so the upper integral or lower integral in (4.28) can be replaced by the ordinary integral; and the same holds for iterated integrals in the reverse order, so that suffices for the proof.

It is customary also to set

$$\int_{x'' \in \overline{\Delta}} \int_{x' \in \overline{\Delta}} f(x', x'') \, dx' \, dx'' = \int_{x' \in \overline{\Delta}} \left( \int_{x'' \in \overline{\Delta}} f(x', x'') \right) \, dx'$$

as another commonly used notation for an iterated integral, where the inner integral, that with respect to the inner variable $x'$, is calculated first, and the result is then integrated over the outer integral, that with respect to the outer variable $x''$. Integrals over Jordan domains $D \subset \overline{\Delta}$ are calculated similarly, by reducing them to integrals over cells; but some care often is required to keep the regions of integration straight. The process can be repeated, with more iterations; so an integral over a Jordan domain in $\mathbb{R}^n$ can be reduced to the calculation of $n$ separate integrals of functions of a single variable.

Figure 4.3: Boundary of a cell in $\mathbb{R}^2$

For examples of the application of Fubini’s Theorem, first let $D \subset \mathbb{R}^2$ be the region bounded by the lines $x_1 + x_2 = 0$, $x_1 - 2x_2 = 0$, and $x_2 = 1$ as in the accompanying figure. An application of Fubini’s Theorem shows that for any continuous function $f(x)$ in $D$

$$\int_D f(x) = \int_0^1 \int_{-2x}^{2x} f(x_1, x_2) \, dx_1 \, dx_2,$$

thus reducing the integral to an iteration of integrals of functions of a single variable. Alternatively for an iterated integral in the other order note that the region $D$ can be split into region $D_1$ consisting of those points $x \in D$ for which $-1 \leq x_1 \leq 0$ and the region $D_2$ consisting of those points $x \in D$ for which $0 \leq x_1 \leq 2$, and then an application of Fubini’s Theorem shows that for any continuous function $f(x)$ in $D$

$$\int_D f(x) = \int_{-1}^0 \int_{-x}^{x} f(x_1, x_2) \, dx_1 \, dx_2 + \int_0^2 \int_{x}^{2x} f(x_1, x_2) \, dx_1 \, dx_2.$$

The order in which Fubini’s Theorem is applied can make a difference in the calculation. For example the iterated integral

$$\int_0^1 \int_{\sqrt{x^2}}^1 \cos(x^2 + 1) \, dx_2 \, dx_1$$
involves a rather complicated trigonometric integral with respect to the variable $x_2$, but the equal iterated integral involves only the simple integrations

\[
\int_0^1 \int_0^{\sqrt{x_2}} \cos(x_2^2 + 1) dx_1 dx_2 = \int_0^1 \int_0^{x_2^2} \cos(x_2^2 + 1) dx_1 dx_2 = \int_0^1 x_2^2 \cos(x_2^2 + 1) dx_2.
\]

There are some cases in which an integrable function $f(x', x'')$ in a product $\Delta = \Delta' \times \Delta''$ is not an integrable function of one variable when the other variable is held fixed. For instance the function $f(x_1, x_2)$ of variables $x_1, x_2 \in \mathbb{R}$ defined in a cell $\Delta = \Delta' \times \Delta'' \subset \mathbb{R}^2$ by

\[
f(x', x'') = \begin{cases} 
0 & \text{if either } x' \text{ or } x'' \text{ is irrational}, \\
\frac{1}{q} & \text{if both } x' \text{ and } x'' \text{ are rational and } x'' = \frac{p}{q}
\end{cases}
\]

for coprime integers $p$ and $q$ actually is independent of the variable $x'$, and as a function of the variable $x''$ it is continuous whenever $x''$ is irrational but discontinuous whenever $x''$ is rational. As a function of two variables its points of discontinuity are the products $\Delta' \times \frac{p}{q}$ for all rational points $\frac{p}{q}$, a countable union of intervals; and since intervals are sets of measure zero in $\mathbb{R}^2$ the points of discontinuity form a set of measure zero, so the function $f$ is integrable. Of course it is also integrable as a function of the variable $x''$ for each fixed point $x' \in \Delta'$; but for any fixed point $x''$ the function takes the values 0 or $\frac{1}{q}$ on dense subsets of the interval $\Delta'$ so the function is not integrable as a function of the variable $x'$. Fortunately in this case the integration can be taken first with respect to the variable $x''$, and the simpler version of Fubini’s Theorem holds.

It should be noted in Fubini’s Theorem it is necessary to assume that a function $f$ on the product cell $\Delta = \Delta' \times \Delta'' \subset \mathbb{R}^2$ is integrable; there are functions that are integrable in the variable $x' \in \Delta'$ for each fixed point $x'' \in \Delta''$, and conversely are integrable in the variable $x'' \in \Delta''$ for each fixed point $x' \in \Delta'$, but are not integrable in the product cell $\Delta$. For example let $f(x', x'')$ be a function on the cell $\Delta' \times \Delta'' \subset \mathbb{R}^2$ that takes the value 1 at the center of the square, that takes the value 1 at each of four points in the subsquares when the square is split in half on each line, but where these points are chosen so that no two lie on the same horizontal or vertical line, and so on; and takes the value 0 otherwise. Thus this function takes each of the values 1 and 0 on dense subsets of the square, so it not integrable; but on each horizontal or vertical line the function takes the value 1 at a single point and is 0 otherwise, so is integrable.
4.3 Limits and Improper Integrals

It is familiar that pointwise limits of continuous functions need no longer be continuous; for example the functions \( f_\nu(x) = x^\nu \) in the closed interval \([0, 1]\) are continuous and

\[
f(x) = \lim_{\nu \to \infty} f_\nu(x) = \begin{cases} 
0 & \text{if } 0 \leq x < 1, \\
1 & \text{if } x = 1,
\end{cases}
\]

which fails to be continuous at the point \( x = 1 \). It is even the case that limits of integrable functions need no longer be integrable; for example it is a straightforward matter to verify that

\[
\lim_{m \to \infty} \lim_{n \to \infty} (\cos \pi m!x)^{\frac{1}{n}} = \begin{cases} 
0 & \text{if } x \text{ is irrational,} \\
1 & \text{if } x \text{ is rational,}
\end{cases}
\]

which is not integrable\(^3\). However a stronger notion of convergence does preserve continuity. A sequence of mappings \( f_\nu : U \to \mathbb{R}^n \) from a subset \( U \subset \mathbb{R}^m \) into \( \mathbb{R}^n \) is uniformly convergent to a mapping \( f \) in \( U \) if for any \( \epsilon > 0 \) there is an integer \( N \) such that \( \|f_N(x) - f(x)\| < \epsilon \) for all \( x \in U \) whenever \( \nu > N \). A standard notation is to write \( f_\nu \to f \) to indicate that the sequence of functions \( f_\nu \) converges pointwise to the function \( f \), and to write \( f_\nu \Rightarrow_U f \) to indicate that the sequence of functions \( f_\nu \) converges uniformly to the function \( f \) in the set \( U \). It is evident from the inequalities \((1.7)\) that the notion of uniform convergence is independent of the norm \( \|x\| \) is used in the definition.

**Theorem 4.8** The limit of a uniformly convergent sequence of continuous mappings from a subset \( U \subset \mathbb{R}^m \) into \( \mathbb{R}^n \) is continuous in \( U \).

**Proof:** If \( f_\nu \Rightarrow_U f \) then for any \( \epsilon > 0 \) there is an integer \( N \) sufficiently large that \( \|f_N(x) - f(x)\| < \frac{1}{3}\epsilon \) for all \( x \in U \). For any point \( a \) there is a \( \delta > 0 \) such that \( \|f_N(a) - f_N(x)\| < \frac{1}{3}\epsilon \) whenever \( \|x - a\| < \delta \), since the mapping \( f_N \) is continuous at the point \( a \). Then \( \|f(a) - f_N(a)\| \leq \|f(a) - f_N(a)\| + \|f_N(a) - f(x)\| + \|f_N(x) - f(x)\| < \frac{1}{3}\epsilon + \frac{1}{3}\epsilon + \frac{1}{3}\epsilon = \epsilon \) whenever \( \|x - a\| < \delta \), so the mapping \( f \) is continuous at the point \( a \); and since that is the case at any point \( a \in U \) the mapping \( f \) is continuous in \( U \), which concludes the proof.

---

\(^3\)It is amusing, although somewhat more difficult to demonstrate, that it does require two consecutive limiting processes to yield this function, which cannot be realized as a simple limit of continuous functions. R. Baire introduced a classification of functions, under which continuous functions are of class 0, functions that are not continuous but are pointwise limits of continuous functions are of class 1, functions that are not of class 1 but are pointwise limits of functions of class 1 are of class 2, and so on. He showed that there are functions of all these “Baire classes”. This is discussed in some detail in C. Carathéodory’s book *Vorlesungen über Reelle Funktionen*, 1935 (English translation *Real Functions*, Chelsea 1946) and F. Hausdorff’s book *Grundzüge der Mengenlehre*, 1937 (English translation *Set Theory*, AMS Chelsea 1957.)

The subject is often treated in books on real analysis through a set of problems.
CHAPTER 4. INTEGRATION

It is easy to see that \( \lim_{n \to \infty} f_n(x) = 0 \) for all points \( x \in [0, 1] \) but that \( \int_0^1 f_n(x) \, dx = 1 \).

**Theorem 4.9** If \( f_\nu \) is a uniformly convergent sequence of integrable functions in a cell \( \Delta \subset \mathbb{R}^n \) then its limit \( f \) is an integrable function in \( \Delta \) and

\[
\lim_{\nu \to \infty} \int_\Delta f_\nu = \int_\Delta f.
\]

**Proof:** Since the functions \( f_\nu \) converge uniformly to \( f \) there is an integer \( N \) such that

\[
|f(x) - f_N(x)| < 1 \quad \text{for all } x \in \Delta;
\]

the function \( f_N \) is integrable by hypothesis so it is bounded by some constant \( M \) in \( \Delta \), and then

\[
|f(x)| \leq |f(x) - f_N(x)| + |f_N(x)| \leq 1 + M
\]

for all \( x \in \Delta \), so the limit function is bounded in \( \Delta \). The set \( E_\nu \subset \Delta \) of points at which the function \( f_\nu \) fails to be continuous is a set of measure zero, since the functions \( f_\nu \) are integrable by hypothesis; so the union \( E = \bigcup_{\nu=1}^\infty E_\nu \) is also a set of measure zero, and all the functions \( f_\nu \) are continuous in \( \Delta \sim E \). The functions \( f_\nu \) converge uniformly to \( f \) in \( \Delta \sim E \) so it follows from the preceding Theorem 4.8 that the limit function \( f \) is also continuous in \( \Delta \sim E \), hence that \( f \) is integrable. For any \( \epsilon > 0 \) there is an \( N \) such that \( |f(x) - f_N(x)| < \epsilon \) for all \( x \in \Delta \) whenever \( \nu > N \); and then for any \( \nu > N \) since the functions \( f \) and \( f_\nu \) are integrable it follows from Theorem 4.3 that

\[
\left| \int_\Delta (f - f_\nu) \right| \leq \int_\Delta |f - f_\nu| \leq \int_\Delta \epsilon = \epsilon |\Delta|
\]

and consequently that \( \lim_{\nu \to \infty} \int_\Delta f_\nu = \int_\Delta f \), which concludes the proof.

Even uniform limits of differentiable functions are not necessarily differentiable; for example the functions \( f_n(x) = \frac{1}{n} \sin n^2 x \) converge uniformly to zero, but \( f'_n(0) \neq 0 \).

**Theorem 4.10** If \( f_\nu \) are differentiable functions in a cell \( \Delta \subset \mathbb{R}^n \) which converge at a point \( a \in \Delta \), and if the derivatives \( f'_\nu \) converge uniformly to a function \( g \) in \( \Delta \), then the functions \( f_\nu \) actually converge uniformly in \( \Delta \) to a differentiable function \( f \) in \( \Delta \) for which \( f' = g \).

**Proof:** For any \( \epsilon > 0 \) there is an \( N \) sufficiently large that \( \|f'_\mu(x) - f'_\nu(x)\|_\infty < \epsilon \) for all \( x \in \Delta \) and \( \|f'_\mu(a) - f'_\nu(a)\|_\infty < \epsilon \) whenever \( \mu, \nu > N \). It then follows from the Mean Value Inequality, Corollary 2.2, that for any point \( x \in \Delta \) there is a point \( y \in \Delta \) such that whenever \( \mu, \nu > N \)

\[
\|f_\mu(x) - f_\nu(x)\|_\infty \leq \|f'_\mu(y) - f'_\nu(y)\|_\infty \|x - a\|_\infty < n \epsilon \|x - a\|_\infty
\]

hence that

\[
\|f_\mu(x) - f_\nu(x)\|_\infty \leq \|f_\mu(a) - f_\nu(a)\|_\infty + n \epsilon \|x - a\|_\infty < (1 + nd) \epsilon
\]

where \( d = \sup_{x \in \Delta} \|x - a\|_\infty \); hence the functions \( f_\nu \) converge uniformly to a function \( f \) in \( \Delta \). For any point \( x \in \Delta \) and any \( h \in \mathbb{R}^n \) the function

\[
(4.30) \quad \eta_\epsilon(h) = \begin{cases} \frac{1}{\|h\|_\infty} (f_\nu(x + h) - f_\nu(x) - f'_\nu(x) h) & \text{if } h \neq 0, \\ 0 & \text{if } h = 0, \end{cases}
\]
is a continuous function of the variable $h$ in an open neighborhood of the origin, since $f_\nu$ is differentiable. If $h \neq 0$ then
\[
\|\eta_\mu(h) - \eta_\nu(h)\|_\infty \leq \frac{1}{\|h\|_\infty} \| (f_\mu(x + h) - f_\mu(x)) - (f_\nu(x) - f_\nu(x)) \|_\infty
\]
\[+ \| f_\mu'(x) - f_\nu'(x) \|_\infty,
\]
and then from the Mean Value Inequality again for some point $y$ between $x$ and $x + h$
\[
\|\eta_\mu(h) - \eta_\nu(h)\|_\infty \leq n \|f_\mu'(y) - f_\nu'(y)\|_\infty + \|f_\mu'(x) - f_\nu'(x)\|_\infty,
\]
and consequently $\|\eta_\mu(h) - \eta_\nu(h)\|_\infty < (n + 1)\epsilon$ whenever $\mu, \nu > N$, which also holds trivially for $h = 0$. Therefore the functions $\eta_\mu(h)$ are uniformly convergent in an open neighborhood of the origin to a continuous limit function $\eta(h)$, for which $\lim_{h \to 0} \eta(h) = \eta(0) = 0$. For any fixed $h \neq 0$ it follows from (4.30) that
\[
\eta(h) = \frac{1}{\|h\|_\infty} \left( f(x + h) - f(x) - g(x)h \right)
\]
since the functions $f_\nu$ converge to $f$ and the functions $f_\nu'$ converge to $g$; but for $h = 0$ it follows from (4.30) that $\eta(0) = 0$. Therefore
\[
\lim_{h \to 0} \eta(h) = \lim_{h \to 0} \frac{1}{\|h\|_\infty} \left( f(x + h) - f(x) - g(x)h \right) = 0,
\]
which shows that $f'(x) = g(x)$ and thereby concludes the proof.

Since the derivatives $f_\nu'$ determine the functions $f_\nu$ only up to additive constants it is apparent that it is not enough in the preceding theorem just to assume something about the convergence of the derivatives $f_\nu'$ and nothing about the convergence of the functions $f_\nu$ themselves; but by that theorem it is enough just to assume in addition to the uniform convergence of the derivatives that the functions $f_\nu$ converge at just a single point. The preceding theorem does not require any regularity properties of the derivatives $f_\nu'$. If the derivatives are assumed to be continuous there is a considerably simpler proof of the preceding theorem; for functions of a single variable $f_\nu(x) - f_\nu(a) = \int_a^x f_\nu'$ by the fundamental theorem of the calculus, so the convergence of the functions $f_\nu$ follows directly from the convergence of the values $f_\nu(a)$ and of the uniform convergence of the functions $f_\nu'$, and the limit is $f(x) = \lim_{\nu \to \infty} \int_a^x f_\nu' = \int_a^x g$ by Theorem 4.9 therefore $f' = g$. This simple argument can be extended to functions of several variable by using the fundamental theorem of calculus (Stokes’s Theorem) for line integrals.

The integral has been defined so far for integrable functions in a Jordan domain in $\mathbb{R}^m$. It is possible to extend the integral as an improper integral to a wider class of functions over a wider class of regions of integration. The terminology is not to suggest anything really improper, but just to indicate that the extension does not necessarily satisfy the results established for the integrals over Jordan domains. A subset $D \subset \mathbb{R}^n$ is an extended Jordan domain if $D = \bigcup_{i=1}^\infty D_i$ where $D_i$ are open Jordan domains such that $\overline{D_i} \subset D_{i+1}$ for all indices $i$. For any continuous function $f$ in $D$ set
\[
f_+(x) = \max(f(x), 0) \quad \text{and} \quad f_-(x) = \max(-f(x), 0)
\]
for any point $x \in D$. Clearly $f_+(x) \geq 0$ and $f_-(x) \geq 0$ at all points $x \in D$, and $f(x) = f_+(x) - f_-(x)$ is a unique decomposition of the function $f$ as a difference of
two non-negative functions; and both functions \( f_+ \) and \( f_- \) are also continuous in \( D \).

For any integer \( i > 0 \) the functions \( \min(f_+(x),i) \) and \( \min(f_-(x),i) \) also are bounded and continuous in \( D \), and consequently they are integrable over the Jordan domains \( D_i \). Moreover since these are positive functions and \( \min(f_+(x),i) \leq \min(f_+(x),i+1) \) while \( D_i \subset D_{i+1} \) it follows that

\[
\int_{D_i} \min(f_+(x),i) \leq \int_{D_{i+1}} \min(f_+(x),i+1);
\]

consequently the sequence of real numbers \( \int_{D_i} \min(f_+(x),i) \) for \( i = 1, 2, \ldots \) has a well defined limit as \( i \to \infty \), either a real number or \( +\infty \), so it is possible to define

\[
(4.32) \quad \int_D f_+ = \lim_{i \to \infty} \int_{D_i} \max(f_+(x),i)
\]

and correspondingly of course

\[
(4.33) \quad \int_D f_- = \lim_{i \to \infty} \int_{D_i} \max(f_-(x),i).
\]

If at least one of these two integrals is finite the improper integral of the function \( f \) is defined to be

\[
(4.34) \quad \int_D f = \int_D f_+ - \int_D f_-
\]

which is either a real number or \( +\infty \) or \( -\infty \). This appears to depend on the particular choice of the sequence of Jordan domains \( D_i \), but the value of the improper integral is actually independent of that choice. Indeed if the set \( D \) also is the union \( D = \bigcup_{j=1}^\infty E_j \)

for some Jordan domains \( E_j \subset \mathbb{R}^n \) such that \( E_j \subset E_{j+1} \) then the same construction leads to other integrals, which for convenience will be denoted by \( \int_{E_j} f_+ \) and \( \int_{E_j} f_- \); of course \( E = D \) as sets, but that common set is represented as an increasing union of Jordan domains in two distinct ways. Since each set \( D_i \) is compact and \( D_i \subset \bigcup_{j=i}^\infty E_j \)

for the open sets \( E_j \) then \( D_i \) is actually contained in a finite union of the sets \( E_{j_i} \), and consequently in the largest of the sets \( E_{j_i} \) in this finite union, say the set \( E_{j_i} \); and since \( D_i \) is contained in \( E_{j_i} \) whenever \( j > j_i \) it is possible to assume that \( j_i \geq i \). Therefore

\[
\int_{D_i} \min(f_+,i) \leq \int_{E_{j_i}} \min(f_+,j_i) \leq \int_E f_+,
\]

and since that is the case for any \( i \) it follows that \( \int_D f_+ \leq \int_E f_+ \), and of course the same holds for the integrals of the function \( f_- \). On the other hand the argument can be reversed, so that the opposite inequality \( \int_E f_+ \leq \int_D f_+ \) also holds, and consequently \( \int_E f_+ = \int_D f_+ \); and again the same holds for integrals of the function \( f_- \), so altogether

\[
\int_E f = \int_D f.
\]

**Theorem 4.11** Any open subset \( U \subset \mathbb{R}^n \) is an extended Jordan domain, for which the approximating Jordan domains are unions of cells in a partition of \( \mathbb{R}^n \).

**Proof**: Let \( \mathcal{P} \) be a partition of the entire space \( \mathbb{R}^n \) into cells having as their sides intervals of length \( 2^{-r} \). For a sufficiently large index \( r_1 \) there will be some closed cells in this partition that are contained in the open set \( U \); the union of their interiors \( D_i \) then is an open Jordan domain contained in \( U \). For a sufficiently large index \( r_2 > r_1 \),
each point on the compact boundary \( \partial D_1 \) will be contained in some open cells of the partition \( P_{r_2} \), and the closures of these cells also will be contained within \( U \); the union \( D_2 \) of all of the open cells of this partition that are contained within \( U \) is another Jordan domain such that \( \overline{D_1} \subset D_2 \). The process continues, yielding a sequence of open Jordan domains \( D_i \) such that \( \overline{D_i} \subset D_{i+1} \); and since the sizes of the cells in this partition tend to 0 any point in \( U \) will eventually lie in one of the Jordan domains \( D_i \), which suffices to conclude the proof.

Although any bounded open subset of \( \mathbb{R}^n \) is an extended Jordan domain, bounded open subsets of \( \mathbb{R}^n \) are not necessarily Jordan domains themselves. For an example even in \( \mathbb{R}^1 \), let \( r_j \) be a list of the rational numbers in the open unit interval \((0,1)\) where \( j \geq 1 \), let \( \Delta_j \) be an open interval of length at most \( |\Delta_j| = \frac{\epsilon}{2^j} \) contained in \((0,1)\) and containing the point \( r_j \) where \( \epsilon < \frac{1}{2} \), and let \( U = \bigcup \Delta_j \). Clearly \( \overline{U} = [0,1] \), and it follows that \( \partial U = [0,1] \sim U \). If \( U \) were a Jordan domain its boundary would be a compact set of content 0, which could be covered by finitely many open intervals of total length less than \( \epsilon \); thus there would be finitely many open intervals \( I_j \subset \mathbb{R}^1 \) such that \( ([0,1] \sim U) \subset \bigcup_{j=1}^N I_j \) and \( \sum_{j=1}^N |I_j| < \epsilon \), and consequently

\[
[0,1] \subset \left( \bigcup_{j=1}^N I_j \cup \bigcup_{i=1}^\infty \Delta_i \right).
\]

However the compact set \([0,1]\) would be covered by finitely many of these open intervals, so that actually

\[
[0,1] \subset \left( \bigcup_{j=1}^N I_j \cup \bigcup_{i=1}^M \Delta_i \right),
\]

which is impossible since \( \sum_{j=1}^n |I_j| + \sum_{i=1}^M |\Delta_i| < 2\epsilon < 1 \); consequently \( U \) cannot be a Jordan domain.
4.4 Change of Variables

In discussing the effects of a change of variables in integration it is convenient to invoke the notion of an oriented integral. That notion is familiar in the study of integrals of functions of a single variable, but it is customarily merely mentioned in passing just as a matter of notation if it is noted explicitly at all. A cell in $\mathbb{R}^1$ is an interval, customarily denoted by $[a, b]$ where $a \leq b$ and oriented in the direction from $a$ to $b$, the customary orientation of the real line by increasing values of the real numbers. The interval bounded by two real numbers $a$ and $b$ can also be denoted by $\Delta(a, b)$, viewed as an unoriented cell and consequently as a cell that is defined independently of which of the two boundary points $a$ or $b$ is the larger number, so that $\Delta(a, b) = \Delta(b, a)$. The content of this cell is $|\Delta(a, b)| = |b - a|$, and the Riemann integral $\int_{\Delta(a, b)} f$ is defined as in (4.8). The oriented integral is defined in terms of the Riemann integral by

\begin{equation}
\int_a^b f(x) \, dx = \begin{cases} 
\int_{\Delta(a, b)} f & \text{if } a \leq b, \\
-\int_{\Delta(a, b)} f & \text{if } a > b.
\end{cases}
\end{equation}

This is the integral over the cell $\Delta(a, b) = \Delta(b, a)$ when that cell is oriented in the direction of the change of the variable $x$ as it moves from the boundary point $a$ to the boundary point $b$, so in the positive direction of the natural orientation of the real line of the variable $x$ if $a < b$ but in the negative direction, the inverse of the natural orientation of the real line of the variable $x$, if $a > b$. Consequently

\begin{equation}
\int_a^b f(x) \, dx = -\int_b^a f(x) \, dx
\end{equation}

for any order of the real numbers $a$ and $b$. For example, the Riemann integral of a constant function $f(x) = k$ is $\int_{\Delta(a, b)} k = k |b - a|$, a result that is independent of which of the two real numbers $a$ or $b$ is the larger; on the other hand it follows readily from (4.35) that the oriented integral is $\int_a^b k \, dx = k(b - a)$, also a result that is independent of which of the two real numbers is larger. In particular the integrals over $\Delta(a, x)$ for any fixed point $a$ can be viewed as defining functions of the real variable $x \in \mathbb{R}^1$, and it follows from (4.35) that

\begin{equation}
\int_{\Delta(a, x)} k = k|x - a| \quad \text{and} \quad \int_a^x k \, dx = k(x - a);
\end{equation}

both of these integrals are continuous functions of the variable $x \in \mathbb{R}^1$, but the oriented integral is differentiable at the point $a$ while the Riemann integral is not, one way in which the oriented integral is more convenient. In general, if $f : \mathbb{R} \to \mathbb{R}$ is an integrable function on the real line then for the Riemann integral it follows readily that

\begin{equation}
\int_{\Delta(a, b)} + \int_{\Delta(b, c)} = \begin{cases} 
\int_{\Delta(a, c)} f & \text{if } a < b < c, \\
\int_{\Delta(a, c)} f + 2\int_{\Delta(b, c)} f & \text{if } a < c < b,
\end{cases}
\end{equation}

while on the other hand for the oriented integral it follows equally readily from (4.35) that

\begin{equation}
\int_a^b f(x) \, dx + \int_b^c f(x) \, dx = \int_a^c f(x) \, dx \quad \text{for any order of } a, b, c.
\end{equation}
4.4. CHANGE OF VARIABLES

Theorem 4.12 (Fundamental Theorem of Calculus) If $f$ is a continuous function on an interval $[a, b] \subset \mathbb{R}^1$ then $F_0(x) = \int_a^x f(t) \, dt$ is a continuously differentiable function of the variable $x \in (a, b)$ such that $F_0'(x) = f(x)$; and conversely if $F(x)$ is a continuously differentiable function in $(a, b)$ such that $F'(x) = f(x)$ then $\int_a^x f(t) \, dt = F(x) - F(a)$ for $x \in (a, b)$.

Proof: If $F_0(x) = \int_a^x f(t) \, dt$ it follows from (4.38) that $F_0(x + h) - F_0(x) = \int_x^{x+h} f(t) \, dt$ whenever $x, x + h \in [a, b]$, where $h$ can be either positive or negative. Since $f$ is continuous then for any fixed point $x \in [a, b]$ and any $\epsilon > 0$ if $h$ is sufficiently small

$$f(x) - \epsilon \leq f(t) \leq f(x) + \epsilon \quad \text{for all } t \in \Delta(x, x + h);$$

the oriented integral of this inequality over the interval $\Delta(x, x + h)$ yields the inequality

$$(f(x) - \epsilon)h \leq F_0(x + h) - F_0(x) \leq (f(x) + \epsilon)h,$$

and dividing by $h$ and taking the limit as $h$ tends to 0 shows that

$$f(x) - \epsilon \leq F_0'(x) \leq f(x) + \epsilon;$$

since that is the case for all $\epsilon > 0$ it follows that $F_0'(x) = f(x)$. If $F(x)$ is a differentiable function in $[a, b]$ such that $F'(x) = f(x)$ then $F(x) = F_0(x) + c$ for some constant $c$, and consequently $\int_a^x f(t) \, dt = F_0(b) - F_0(a) = F(b) - F(a)$, which suffices for the proof.

A $C^1$ homeomorphism $\phi : [a, b] \to \Delta(\phi(a), \phi(b))$ either is either orientation preserving, so that $\phi'(a) \leq 0$ and $\phi'(x) \geq 0$ for all points $x \in (a, b)$, or is orientation reversing, so that $\phi'(a) \geq 0$ and $\phi'(x) \leq 0$ for all points $x \in (a, b)$. In either case, if $f$ is a continuous function in $[c, d]$ then for any differentiable function $F$ in $[a, b]$ such that $F'(x) = f(x)$ it follows from Theorem 4.12 that

$$\int_{\phi(a)}^{\phi(b)} f(x) \, dx = F(\phi(b)) - F(\phi(a)).$$

By the chain rule $(F \circ \phi)' = (f \circ \phi) \phi'$, so by Theorem 4.12 again

$$\int_a^b f(\phi(x)) \phi'(x) \, dx = F(\phi(b)) - F(\phi(a)).$$

Comparing the two preceding observations shows that

$$(4.40) \quad \int_{\phi(a)}^{\phi(b)} f(x) \, dx = \int_a^b f(\phi(x)) \phi'(x) \, dx.$$

This is the familiar formula for the change of variables in an oriented integral; actually it is customary to set $y = \phi(x)$ and rewrite the preceding equation in the form

$$(4.41) \quad \int_{\phi(a)}^{\phi(b)} f(x) \, dx = \int_a^b f(y) \frac{dy}{dx} \, dx.$$

On the other hand for the Riemann integral it follows from (4.35) that

$$(4.42) \quad \int_{\Delta(a,b)} f = \int_a^b f(x) \, dx \quad \text{since } a \leq b.$$
Indeed if \( f \) separately \( \psi \) variable. In order to show that (4.13) holds for the integral translations, permutations of the variables, or mappings that change only a single point \( \phi \) the restriction of the mapping.

To begin with a useful general observation, if the theorem holds for mappings

\[
\int_{\Delta(a,b)} f(x) \, dx = \begin{cases} 
\int_{\phi(a)}^{\phi(b)} f(x) \, dx & \text{if } \phi(a) \leq \phi(b), \\
-\int_{\phi(a)}^{\phi(b)} f(x) \, dx & \text{if } \phi(a) \geq \phi(b), 
\end{cases}
\]

and consequently in terms of the unoriented integral (4.40) takes the form

\[
\int_{\Delta(c,d)} f = \begin{cases} 
\int_{\Delta(a,b)} (f \circ \phi) \phi' & \text{if } \phi(a) \leq \phi(b), \\
-\int_{\Delta(a,b)} (f \circ \phi) \phi' & \text{if } \phi(a) \geq \phi(b); 
\end{cases}
\]

in the first case in (4.44) the mapping \( \phi \) is orientation-preserving and \( \phi'(x) \geq 0 \), while in the second case in (4.44) the mapping \( \phi \) is orientation-reversing and \( \phi'(x) \leq 0 \), so both cases can be subsumed under the general formula

\[
\int_{\phi(\Delta)} f = \int_{\Delta} (f \circ \phi)|\phi'| 
\]

for any cell \( \Delta \subset \mathbb{R}^1 \). This is the form in which the result extends to mappings in higher dimensions.

**Theorem 4.13** If \( \phi : D \rightarrow E \) is \( C^1 \) homeomorphism between bounded open subsets \( D \subset \mathbb{R}^n \) and \( E \subset \mathbb{R}^n \) and \( \det \phi'(x) \neq 0 \) at each point \( x \in D \) then

\[
\int_{E} f = \int_{D} (f \circ \phi)|\det \phi'| 
\]

for any continuous function \( f \) in \( E \).

**Proof:** To begin with a useful general observation, if the theorem holds for mappings \( \phi : D \rightarrow E \) and \( \psi : E \rightarrow F \) then it also holds for the composition \( \psi \circ \phi : D \rightarrow F \). Indeed if \( f \) is a continuous function in \( F \) then by (4.46) for the two mappings \( \phi \) and \( \psi \) separately

\[
\int_{F} f = \int_{E} (f \circ \psi)|\det \psi'| = \int_{D} ((f \circ \psi) \circ \phi)|\det (\psi'(\phi))||\det \phi'| \\
= \int_{D} (f \circ (\psi \circ \phi))|\det (\psi \circ \phi)'|
\]

since \( (\psi \circ \phi)' = (\psi'(\phi)) \phi' \) by the chain rule for differentiation.

The theorem will be proved first purely locally. As a consequence of Theorem 3.2, the restriction of the mapping \( \phi : D \rightarrow E \) to a sufficiently small cell \( \Delta \) about any point \( a \in D \) can be written locally as a composition of mappings that are either translations, permutations of the variables, or mappings that change only a single variable. In order to show that (4.13) holds for the integral \( \int_{\Delta} f \), in the light of the

\[\text{The hypothesis that } \det \phi'(x) \neq 0 \text{ at each point } x \in D \text{ can be eliminated by applying a special case of Sard's Theorem, showing that the image under a } C^1 \text{ mapping } \phi \text{ of the set of points } x \text{ at which } \det \phi'(x) = 0 \text{ is a set of measure } 0. \text{ Sard's Theorem in general is discussed and proved in S. Sternberg's book } \textit{Lectures on Differential Geometry}, \text{ Prentice-Hall 1964.} \text{ In practice though the theorem as stated generally suffices, just by splitting the domain of integration suitably.} \]
4.4. CHANGE OF VARIABLES

Figure 4.4: Mappings that change a single variable

observation in the preceding paragraph it suffices just to show that it holds in the special cases that the mapping $\phi : \Delta \rightarrow E$ is either a translation, a permutation of the variables, or a mapping that changes only a single variable. That is quite obvious in the first two cases, since the definition of the integral clearly is independent of either change of variables and in both cases $\det \phi'(x) = 1$ for all $x \in \Delta$. If the mapping $\phi$ changes only the coordinate $x_1$ then when points $x \in \mathbb{R}^n$ are written

$$x = \begin{pmatrix} x_1 \\ x_{II} \end{pmatrix}$$

where $x_1 \in \mathbb{R}^1$ and $x_{II} \in \mathbb{R}^{n-1}$, the mapping $\phi$ has the form

$$\phi(x) = \begin{pmatrix} \phi_1(x) \\ x_{II} \end{pmatrix}$$

where $\phi_1(x) \in \mathbb{R}^1$ and $x_{II} \in \mathbb{R}^{n-1}$, as indicated in the accompanying Figure 4.4; so

$$\phi'(x) = \begin{pmatrix} \partial_1 \phi_1(x) & \partial_{II} \phi_1(x) \\ 0 & I_{n-1} \end{pmatrix}$$

where $I_{n-1}$ is the $(n-1) \times (n-1)$ identity matrix, and consequently $\det \phi'(x) = \partial_1 \phi_1(x)$. For any fixed point $x_{II} \in \mathbb{R}^{n-1}$ the subset

$$\Delta(x_{II}) = \left\{ x = \begin{pmatrix} x_1 \\ x_{II} \end{pmatrix} \in \Delta \right\}$$

for some $x_1$, is mapped to the set

$$E(x_{II}) = \left\{ y = \begin{pmatrix} y_1 \\ x_{II} \end{pmatrix} \in E \right\};$$

the set $\Delta$ is the union of the subsets $\Delta(x_{II})$ as the parameter $x_{II}$ ranges over a subset $\Delta_{II} \subset \mathbb{R}^{n-1}$, and $E$ is the union of the corresponding subsets $E(y_{II})$ as the parameter $y_{II} = x_{II}$ ranges over the subset $\Delta_{II} \subset \mathbb{R}^{n-1}$. For a continuous function $f$ in $\Delta$ it follows from Fubini’s Theorem and (4.45) applied to the integrals along the
one-dimensional subsets $\Delta(x_{II})$ and $E(y_{II})$ that

$$
\int_E f = \int_{y_{II} \in E} \int_{x_{II} \in E(y_{II})} f(y_{II}, x_{II}) \\
= \int_{y_{II} \in E} \int_{x_{II} \in \Delta(x_{II})} f(x_{II}, y_{II}) \ |\partial_1 \phi'(x)| \\
= \int_{x_{II} \in \Delta} \int_{x_{II} \in \Delta(y_{II})} f(x_{II}, y_{II}) \ |\det \phi'(x)| \\
= \int_{\Delta} (f \circ \phi) \ |\det \phi'(x)|.
$$

Thus (4.46) does hold locally.

By Theorem 4.11 the open set $D$ is an extended Jordan domain, so it can be written as the union of an increasing sequence of open Jordan domains $D_i$, each of which is a union $D_i = \bigcup_j \Delta_{ij}$ of cells $\Delta_{ij}$; and by passing to refinements of these cells it can be assumed that (4.46) holds for each of the cells $\Delta_{ij}$. The images of the boundaries of these cells under the mapping $\phi$ are submanifolds of $\mathbb{R}^n$, by Corollary 3.2 to the Rank Theorem, so the images of these cells are also Jordan domains. Since these cells are disjoint aside from subsets of their boundaries, which have content 0, it follows from Theorem 4.6 that $\int_{D_i} f = \sum_j \int_{D_{ij}} f$; and since (4.46) holds for the integral over each cell $D_{ij}$ it holds for the integrals over the Jordan domains $D_i$. The integral over $D$ is the limit of integrals over the Jordan domains $D_i$ as in (4.32) and (4.33), and that suffices for the proof.

For an explicit example, consider the integral $\int_E \log(x_1^2 + x_2^2)$ where

$$
E = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid a \leq x_1^2 + x_2^2 \leq b, \ x_1 \geq 0, \ x_2 \geq 0 \right\}.
$$

This is most conveniently handled by introducing polar coordinates through the change of coordinates

$$
\phi : \mathbb{R}^2(x_1, x_2) \longrightarrow \mathbb{R}^2(r, \theta)
$$

where $x_1 = r \cos \theta, \ x_2 = r \sin \theta$. The mapping $\phi$ establishes a one-to-one mapping $\phi : D \longrightarrow E$ where

$$
D = \left\{ (r, \theta) \in \mathbb{R}^2 \mid a \leq r \leq b, \ 0 \leq \theta \leq \frac{\pi}{2} \right\};
$$

and

$$
\phi'(r, \theta) = \frac{\partial(x_1, x_2)}{\partial(r, \theta)} = \begin{pmatrix}
\frac{\partial x_1}{\partial r} & \frac{\partial x_1}{\partial \theta} \\
\frac{\partial x_2}{\partial r} & \frac{\partial x_2}{\partial \theta}
\end{pmatrix} = \begin{pmatrix}
\cos \theta & -r \sin \theta \\
\sin \theta & r \cos \theta
\end{pmatrix}
$$

so $|\det \phi'(r, \theta)| = r$. The formula for the change of variables and an application of Fubini’s Theorem show that

$$
\int_E \log(x_1^2 + x_2^2) = \int_D \log(r^2) r \\
= \int_a^b \int_0^{\pi/2} 2r \log r \ d\theta \ dr \\
= \frac{\pi}{2} \int_a^b 2r \log r = \frac{\pi}{2} r^2 \log r - \frac{1}{2} \bigg|_a^b \\
= \frac{\pi}{2} (b^2 \log b - a^2 \log a) - \frac{\pi}{4} (b^2 - a^2).
$$
Chapter 5

Differential Forms

5.1 Line Integrals

Integrals over submanifolds of the vector spaces \( \mathbb{R}^n \), and over suitable generalizations of submanifolds, play a significant role in applications. The first examples are integrals over curves \( \gamma \subset U \) in open subsets \( U \subset \mathbb{R}^n \), where a curve in \( U \) is defined to be the image of a continuous mapping \( \phi : [0,1] \rightarrow U \). Curves are oriented by the natural orientation of the parameter interval \([0,1]\), so are viewed as traversed in the direction from the initial point \( \phi(0) \) to the terminal point \( \phi(1) \) in \( U \). Much of the subsequent discussion will focus on \( C^1 \) curves, those for which the mappings \( \phi : [0,1] \rightarrow U \) are \( C^1 \) mappings in an open neighborhood of the closed interval \([0,1] \subset \mathbb{R}^n \); but it is common also to consider \( C^k \) curves for any integer \( k > 0 \) or even \( C^\infty \) curves, those curves for which the mappings have continuous partial derivatives of all orders. Unless explicitly assumed otherwise, however, curves are defined by merely continuous mappings \( \phi : [0,1] \rightarrow U \). Also unless explicitly assumed otherwise, curves are not required to be defined by one-to-one mappings \( \phi : [0,1] \rightarrow U \), and \( C^1 \) curves are not required to be defined by mappings for which the derivative \( \phi' \) is everywhere nonvanishing; thus \( C^1 \) curves are not necessarily submanifolds of \( U \) even locally, although of course a submanifold of \( U \) is a special case of a curve in \( U \). Two \( C^1 \) parametrizations \( \phi : [a,b] \rightarrow \gamma \) and \( \psi : [c,d] \rightarrow \gamma \) are said to be \( C^1 \) equivalent parametrizations if there is a \( C^1 \) function \( h \) in an open neighborhood of the closed interval \([a,b] \subset \mathbb{R} \) such that the mapping \( h : [a,b] \rightarrow [c,d] \) is a \( C^1 \) homeomorphism for which \( \phi(t) = \psi(h(t)) \) for all \( t \in [a,b] \); this is a natural equivalence relation between the parametrizations of \( C^1 \) curves, and of course can be defined correspondingly for merely continuous curves or \( C^k \) curves for any \( k \).

If \( \gamma \subset U \) is a curve in an open subset \( U \subset \mathbb{R}^n \) described by a parametrization \( \phi : [0,1] \rightarrow \gamma \) and if \( f : U \rightarrow \mathbb{R} \) is a function in \( U \) the composition \( f \circ \phi \), also denoted by \( \phi^*(f) \), is a well defined function on the parameter interval \([0,1]\) called the induced function under the mapping \( \phi \). If the parametrization \( \phi \) and the function \( f \) are both \( C^k \) mappings then the induced function \( \phi^*(f) \) is also a \( C^k \) function. If the vector space of \( C^k \) functions in a set \( S \) is denoted by \( \Lambda^0_k(S) \) it is clear that

\[
(5.1) \quad \phi^* : \Lambda^0_k(U) \rightarrow \Lambda^0_k([0,1]),
\]

is a linear mapping between these two vector spaces. The integral of a continuous
function \( f \in \Lambda^0 \) on a curve \( \gamma \) described by a mapping \( \phi : [0, 1] \longrightarrow U \) is defined by

\[
(5.2) \quad \int_{\gamma, \phi} f = \int_{[0, 1]} \phi^*(f);
\]

as the notation indicates, this integral depends not just on the subset \( \gamma \subset U \) but also on the parametrization \( \phi \). For example, the curve \( \gamma \subset \mathbb{R}^2 \) from the point \((0,0)\) to the point \((1,1)\) described by the equation \( x_2 = x_1^2 \) can be parametrized by the mapping \( \phi_1(x_1) = (x_1, x_1^2) \) in terms of the parameter \( 0 \leq x_1 \leq 1 \), or alternatively can be parametrized by the mapping \( \phi_2(x_2) = (\sqrt{x_2}, x_2) \) in terms of the parameter \( 0 \leq x_2 \leq 1 \); and

\[
\int_{\gamma, \phi_1} (x_1 + x_2) = \int_0^1 (x_1 + x_1^2) dx_1 = \frac{5}{6},
\]

while

\[
\int_{\gamma, \phi_2} (x_1 + x_2) = \int_0^1 (\sqrt{x_2} + x_2) dx_2 = \frac{7}{6}.
\]

The way to obtain a more intrinsic integral over a curve \( \gamma \) is to use an intrinsic parametrization of the curve, such as parametrization by arc length. If \( \gamma \subset U \) is a curve in an open subset \( U \subset \mathbb{R}^n \) and is parametrized by a continuous mapping \( \phi : [0, 1] \longrightarrow U \) then for any partition \( P \) of the interval \([0, 1]\) by points of subdivision \( 0 = t_0 < t_1 < \cdots < t_m = 1 \) the curve \( \gamma \) can be approximated by the polygon joining the successive points \( \phi(t_i) \) for \( 0 \leq i \leq m \); and the length of this approximating polygon is

\[
(5.3) \quad L(\gamma, P) = \sum_{i=1}^m \|\phi(t_i) - \phi(t_{i-1})\|_2,
\]

which can be considered as an approximation of the length of the curve. When a point of subdivision \( t' \) is added between \( t_{i-1} \) and \( t_i \) the length \( \|\phi(t_i) - \phi(t_{i-1})\|_2 \) of that piece of the polygonal approximation of the curve is replaced by the sum

\[
\|\phi(t_i) - \phi(t'_i)\|_2 + \|\phi(t'_i) - \phi(t_{i-1})\|_2
\]

of the lengths of the two pieces of the polygonal approximation in the refined partition, which by the triangle inequality is larger; consequently adding further points of subdivision increases the length of the approximating polygon. The **arc length** of the curve then is defined by

\[
(5.4) \quad L(\gamma) = \sup_{P} L(\gamma, P),
\]

where the supremum is extended over all partitions \( P \) of the parameter interval \([0, 1]\). Of course it may be the case that this supremum is infinite; those curves for which the supremum is finite are called **rectifiable** curves, and for such curves the arc length \( (5.4) \) has a uniquely determined finite value. If \( L(a, b) \) denotes the arc length of the segment \( \phi([a, b]) \) of a rectifiable curve \( \gamma \) parametrized by a mapping \( \phi : [0, 1] \longrightarrow \gamma \) then

\[
(5.5) \quad L(a, b) = L(a, c) + L(c, b) \quad \text{if} \quad 0 \leq a \leq c \leq b \leq l;
\]

for the length of a polygonal approximation to the segment of the curve for parameter values in the interval \([a, b]\) is only increased by adjoining the point \( c \) to the other
5.1. LINE INTEGRALS

points of subdivision, so when examining the supremum of the lengths of the polygonal approximations it always can be assumed that $c$ is one of the points of subdivision and consequently the partition consists of a partition of the segment of curve parametrized by the interval $[a, c]$ and a partition of the segment of curve parametrized by the interval $[c, b]$, and the supremum consequently is the sum of the supreme of the partitions of these two separate sections of the curve. If $L(0, t_1) = L(0, t_2)$ for parameter values $t_1 < t_2$ it follows from the definition of arc length that $\| \phi(t_2) - \phi(t_1) \|_2 \leq L(t_1, t_2) = L(0, t_2) - L(0, t_1) = 0$ and consequently that $\phi(t_1) = \phi(t_2)$; therefore associating to any value $s \in [0, l]$, where $l = L(0, 1)$ is the arc length of the curve, the point $\phi_0(s) = \phi(t) \in \gamma$, where $t \in [0, 1]$ is a choice of the parameter value such that $L(0, t) = s$, defines a mapping $\phi_0 : [0, l] \rightarrow \gamma$. This mapping is an alternative parametrization of the curve $\gamma$, called the parametrization by arc length, and is actually a continuous parametrization. Indeed if $0 \leq s_1 \leq s_2 \leq l$ and if $t_1, t_2 \in [0, 1]$ are any parameter values such that $L(0, t_2) = s_1$ and $L(0, t_2) = s_2$ it follows from the definitions of arc length and of the mapping $\phi_0$ that

$$\| \phi_0(s_1) - \phi_0(s_2) \|_2 = \| \phi(t_1) - \phi(t_2) \|_2 \leq L(t_1, t_2) = L(0, t_2) - L(0, t_1) = s_2 - s_1,$$

so the mapping $\phi_0$ is continuous. For consistency arc length as the parameter always will be denoted by $s$. The integral of a continuous function $f$ in $U$ along a rectifiable curve $\gamma \subset U \subset \mathbb{R}^n$ of length $l$ then can be defined intrinsically as the integral

$$\int_\gamma f = \int_0^l f(s) ds$$

with respect to arc length on the curve $\gamma$.

**Theorem 5.1** A $C^1$ curve $\gamma \subset U \subset \mathbb{R}^n$ with a $C^1$ parametrization $\phi : [0, 1] \rightarrow U$ is rectifiable and its arc length is

$$L(\gamma) = \int_0^1 \| \phi' \|_2.$$

**Proof:** For any partition $P$ of the interval $[0, 1]$ it follows from the fundamental theorem of calculus in one variable applied to each coordinate function of the mapping $\phi$ and the inequality (4.24) that

$$L(P) = \sum_{i=1}^m \| \phi(t_i) - \phi(t_{i-1}) \|_2 \leq \sum_{i=1}^m \int_{t_{i-1}}^{t_i} \| \phi' \|_2 \leq \sum_{i=1}^m \int_{t_{i-1}}^{t_i} \| \phi' \|_2 = \int_0^1 \| \phi' \|_2;$$

and since $\| \phi' \|_2$ is continuous this integral is finite so the curve $\gamma$ is rectifiable. If $L(s, t)$ denotes the length of the segment of the curve parametrized by the interval $[s, t] \subset [0, 1]$ then for any point $t \in [0, 1]$ and any $h > 0$ sufficiently small that $t + h \leq 1$ it follows from (5.8) applied to the restriction of the mapping $\phi$ to the interval $[t, t + h]$ that

$$L(t, t + h) \leq \int_t^{t+h} \| \phi' \|_2.$$

On the other hand for any unit vector $u \in \mathbb{R}^n$ it follows from the definition of arc length, the Cauchy-Schwarz inequality and the mean value theorem for the real-valued function $u \cdot \phi(t)$ that

$$L(t, t + h) \geq \| \phi(t + h) - \phi(t) \|_2 \geq | u \cdot ( \phi(t + h) - \phi(t) ) | = | u \cdot \phi'(t) | h,$$
for some point $\tau$ in the interval $t < \tau < t + h$. In particular if $u$ is the unit vector in the direction of the vector $\phi'(\tau)$ then $|u \cdot \phi'(\tau)| = \|\phi'(\tau)\|_2$, so for this unit vector the preceding inequality reduces to

$$L(t, t + h) \geq \|\phi'(\tau)\|_2 h.$$  

(5.10)

Since $L(t, t + h) = L(0, t + h) - L(0, t)$ by (5.5) it follows from the inequalities (5.9) and (5.10) that

$$\|\phi'(\tau)\|_2 \leq \frac{1}{h} \left( L(0, t + h) - L(0, t) \right) \leq \frac{1}{h} \int_t^{t+h} \|\phi'\|_2,$$

which in the limit as $h$ tends to 0, and consequently $\tau$ tends to $t$, becomes the inequality

$$\|\phi'(t)\|_2 \leq \lim_{h \to 0} \frac{1}{h} \left( L(0, t + h) - L(0, t) \right) \leq \|\phi'(t)\|_2;$$

therefore the function $L(0, t)$ of the variable $t \in (0, 1)$ is differentiable and

$$\frac{d}{dt}L(0, t) = \|\phi'(t)\|_2.$$

It finally follows from the fundamental theorem of calculus that

$$L(\gamma) = L(0, 1) = \int_0^1 \frac{d}{dt}L(0, t) = \int_0^1 \|\phi'\|_2,$$

and that concludes the proof.

**Corollary 5.1** If $\gamma \in U \subset \mathbb{R}^n$ is a $C^1$ curve with a $C^1$ parametrization $\phi : [0, 1] \to U$ such that $\phi'(t) \neq 0$ for all $t \in [0, 1]$ then the parametrization of this curve by arc length is a $C^1$ parametrization that is $C^1$ equivalent to the parametrization $\phi$.

**Proof:** The function $h(t) = L(0, t) = \int_0^t \|\phi'\|_2$ is a continuously differentiable function of the parameter $t \in [0, 1]$ for which $h'(t) = \|\phi'(t)\|_2 > 0$, so it is a strictly increasing function describing a one-to-one $C^1$ mapping $h : [0, 1] \to [0, l]$ between the two parameter intervals, where $l$ is the length of the curve; and by the inverse mapping theorem the inverse of the mapping $h$ is also $C^1$, so the mapping $h$ is a $C^1$ homeomorphism between the two parameter intervals. The composition $\psi(s) = \phi(h^{-1}(s))$ then is the parametrization of $\gamma$ by arc length, exhibited as being $C^1$ equivalent to the parametrization $\phi$, thereby concluding the proof.

**Corollary 5.2** A $C^1$ parametrization $\phi : [0, l] \to \mathbb{R}^n$ of a curve $\gamma \in \mathbb{R}^n$ is the parametrization by arc length if and only if $\|\phi'(s)\|_2 = 1$ for all points $s \in [0, l]$.

**Proof:** If $\phi : [0, l] \to \mathbb{R}^n$ is the parametrization by arc length then by Theorem 5.1 the length of the segment $\phi([0, s])$ of the curve is $s = \int_0^s \|\phi'\|_2$, so by the fundamental theorem of calculus

$$1 = \frac{d}{ds}s = \frac{d}{ds} \int_0^s \|\phi'\|_2 = \|\phi'(s)\|_2$$

for all $s \in [0, l]$. Conversely if $\phi : [0, l] \to \mathbb{R}^n$ is a parametrization of the curve $\gamma$ such that $\|\phi'(s)\|_2 = 1$ then by Theorem 5.1 the arc length of the segment $\phi([0, s])$ of that curve is $\int_0^s \|\phi'\|_2 = \int_0^s 1 = s$, and that suffices for the proof.
5.1. LINE INTEGRALS

If \( \gamma \subset U \) is a \( C^1 \) curve in an open subset \( U \subset \mathbb{R}^n \) parametrized by arc length under the mapping \( \phi : [0, l] \rightarrow \gamma \), it follows from Corollary 3.5 to the Rank Theorem that the image of an open neighborhood of any parameter value \( s_0 \in [0, l] \) is a \( C^1 \) submanifold of an open neighborhood of the point \( \phi(s_0) \in U \); the vector \( \phi'(s_0) \) is a tangent vector to that submanifold, as in the discussion on page 47, and has length \( ||\phi'(s_0)|| = 1 \) by the preceding Corollary 5.2, so it is the unit tangent vector to the local piece of the curve \( \gamma \) at the point \( \phi(s_0) \), often denoted by \( \tau(s_0) = \phi'(s_0) \).

The curve \( \gamma \) may intersect itself at various points, and at a point of intersection there are unit tangent vectors of various different directions, the unit tangent vector to the various local components of the curve \( \gamma \). If \( f : U \rightarrow \mathbb{R}^n \) is a continuous vector field in the open subset \( U \), the dot product \( f(\phi(s)) \cdot \tau(\phi(s)) = f(\phi(s)) \cdot \phi'(s) \) is a well defined continuous function of the variable \( s \in [0, l] \) so it has a well defined integral

\[
(5.11) \quad \int_{\gamma} f \cdot \tau \, ds = \int_{0}^{l} f(\phi(s)) \cdot \phi'(s) \, ds.
\]
called the line integral of the vector field \( f \) along the curve \( \gamma \). The integral of this vector field is not only intrinsically defined, but is actually independent of the parametrization of the curve.

**Theorem 5.2** If \( \psi : [a, b] \rightarrow \gamma \) is a \( C^1 \) parametrization of a \( C^1 \) curve \( \gamma \subset U \subset \mathbb{R}^n \) that is \( C^1 \) equivalent to the parametrization of \( \gamma \) by arc length then

\[
(5.12) \quad \int_{\gamma} f \cdot \tau \, ds = \int_{a}^{b} f(\psi(t)) \cdot \psi'(t) \, dt
\]
for any continuous vector field \( f \) in \( U \).

**Proof:** If \( h : [a, b] \rightarrow [0, l] \) is a \( C^1 \) homeomorphism such that \( \psi(t) = \phi(h(t)) \) then by the chain rule \( \psi'(t) = \phi'(h(t))h'(t) \), so by the formula for the change of variables in integration in Theorem 4.13 it follows from (5.11) that

\[
\int_{\gamma} f \cdot \tau \, ds = \int_{0}^{1} f(\phi(s)) \cdot \phi'(s) \, ds = \int_{a}^{b} f(\phi(h(t))) \cdot \phi'(h(t))h'(dt) \, dt
\]
\[
= \int_{a}^{b} f(\psi(t)) \cdot \psi'(t) \, dt,
\]
which suffices for the proof.

As an example, consider the integral of the vector field \( f = (x_1^2, x_1 x_2) \in \mathbb{R}^2 \) over the parabola described parametrically by the mapping \( \phi : [0, 1] \rightarrow \mathbb{R}^2 \) for which \( \phi(t) = (t, t^2) \). Since \( \phi'(t) = (1, 2t) \) then \( f(\phi(t)) = (t^2, t^3) \) so \( f(\phi(t)) \cdot \phi'(t) = t^2 + 2t^3 \) and consequently

\[
\int_{\gamma} f \cdot d\tau \, ds = \int_{0}^{1} f(\phi(t)) \cdot \phi'(t) \, dt
\]
\[
= \int_{0}^{1} (t^2 + 2t^3) \, dt = \left[ \frac{1}{3} t^3 + \frac{2}{5} t^5 \right]_{0}^{1} = \frac{11}{15}.
\]
A particularly interesting example is the line integral of a vector field that is the gradient \( f(x) = \nabla h(x) \) of a function \( h \) in an open set \( U \subset \mathbb{R}^n \), a vector field called a conservative vector field with the potential \( h \).
Theorem 5.3 If $\mathbf{f}(x) = \nabla h(x)$ is a conservative vector field with a $C^1$ potential $h$ in a connected open subset $U \subseteq \mathbb{R}^n$ and if $\gamma$ is a $C^1$ curve in $U$ from a point $a \in U$ to a point $b \in U$ then

\begin{equation}
\int_{\gamma} \mathbf{f} \cdot \mathbf{\tau} \, ds = h(b) - h(a).
\end{equation}

**Proof:** Let $\psi : [0, l] \rightarrow U$ be the parametrization of the curve $\gamma$ by arc length, where $\psi(s) = \{\psi_j(s)\}$ and $\psi(0) = a$, $\psi(l) = b$. By the chain rule

\[
\frac{d}{ds} h(\psi(s)) = h'(\psi(s)) \psi'(s) = \sum_{j=1}^{n} \partial_j h(\psi(s)) \psi'_j(s) = \sum_{j=1}^{n} f_j(\psi(s)) \psi'_j(s)
\]

and consequently

\[
\int_{\gamma} \mathbf{f} \cdot \mathbf{\tau} \, ds = \int_{0}^{1} \mathbf{f}(\psi(s)) \cdot \psi'(s) = \int_{0}^{1} dh(\psi(s)) = h(\psi(l)) - h(\psi(0)) = h(b) - h(a),
\]

and that suffices for the proof.

Thus the integral of a conservative vector field along a curve really depends only on the beginning and end points of the curve, and is quite independent of the particular path taken between these two points. There are vector fields that are not conservative, so for which the line integral between two points depends on the path chosen between those two points. For example for the vector field $\mathbf{f}(x) = \{x_2, -x_1\}$ in $\mathbb{R}^2$, and the curves $\gamma_1$ and $\gamma_2$ from the origin to the point $\{1, -1\}$ described by the mappings $\psi_1(t) = \{t, -t\}$ and $\psi_2(t) = \{t, -t^2\}$ for $t \in [0, 1]$

\[
\begin{align*}
\int_{0}^{1} f(\psi_1(t)) \psi_1'(t) \, dt &= \int_{0}^{1} (-t, -t) \cdot (1, -1) \, dt = \int_{0}^{1} 0 \, dt = 0 \\
\int_{0}^{1} f(\psi_2(t)) \psi_2'(t) \, dt &= \int_{0}^{1} (-t^2, -t) \cdot (1, -2t) \, dt = \int_{0}^{1} t^2 \, dt = \frac{1}{3}.
\end{align*}
\]

Actually the condition that the line integral of a vector field is independent of the path characterizes conservative vector fields.

Theorem 5.4 A $C^1$ vector field $\mathbf{f} : U \rightarrow \mathbb{R}^n$ in a connected open subset $U \subseteq \mathbb{R}^n$ is conservative if and only if the line integral $\int_{\gamma} \mathbf{f} \cdot \mathbf{\tau} \, ds$ depends only on the end points of the path $\gamma$ for any $C^1$ curve $\gamma \subset U$.

**Proof:** The preceding theorem demonstrated that if $\mathbf{f}$ is a conservative vector field then the line integral $\int_{\gamma} \mathbf{f} \cdot \mathbf{\tau} \, ds$ depends just on the beginning and end points of the path of integration $\gamma$ for any $C^1$ curve $\gamma \subset U$. Conversely suppose that $\mathbf{f}$ is a $C^1$ vector field in a connected open subset $U \subseteq \mathbb{R}^n$ such that the line integral $\int_{\gamma} \mathbf{f} \cdot \mathbf{\tau} \, ds$ depends only on the beginning and end points of the path of integration $\gamma$ for any $C^1$ curve $\gamma \subset U$. If $a \in U$ the integral $h(x) = \int_{a}^{x} \mathbf{f} \cdot \mathbf{\tau} \, ds$ then is a well defined function of the point $x \in U$. For any point $x = (x_1, \ldots, x_n) \in U$ and any index $1 \leq i \leq n$ set

\[
\psi_i(t) = (x_1, \ldots, x_{i-1}, t, x_{i+1}, \ldots, x_n),
\]
so that $\psi_i(x_i) = x$. The function $h(x)$ can be written as the sum of the integral along any $C^1$ curve $\gamma_1$ from $a$ to $\psi_i(b_i)$ followed by the curve $\gamma_2$ that is the image under the mapping $\psi_i$ of the interval $b_i \leq t \leq x_i$, provided that $b_i$ is sufficiently near $x_i$; thus $\gamma_2$ is a path that approaches the point $x$ along a line consisting of points that differ only in the $i$-th coordinate, and

$$h(x) = \int_{\gamma_1} f \cdot \tau \, ds + \int_0^{x_i} f_i(x_1, \ldots, x_{i-1}, t, x_{i+1} \ldots x_n) \, dt.$$ 

Then by the fundamental theorem of calculus it follows that $\partial h(x) = f_i(x)$; and since that is the case for $1 \leq i \leq n$ it follows that $f = \nabla h$ so $f$ is a conservative vector field with the potential $h$, which concludes the proof.

### 5.2 Differential forms

The differentials $dx_j$ of the coordinate functions $x_j$ in $\mathbb{R}^n$ are a basis for the vector space $\mathbb{R}^n$, as discussed in Section 2.4. Vector fields over an open subset $U \subset \mathbb{R}^n$, mappings $f : U \rightarrow \mathbb{R}^n$ with the coordinate functions $f(x) = \{f_j(x)\}$, when written as the sums

$$\omega(x) = \sum_{j=1}^n f_j(x) \, dx_j$$

of multiples of the basis vectors $dx_j$, are called **differential forms** of degree 1 in $U$, or for short just differential 1-forms in $U$. Mappings $f : U \rightarrow \mathbb{R}^k$ for indices $2 \leq r \leq n$ with the coordinate functions $f = \{f_{j_1 \ldots j_r}(x)\}$ for $1 \leq j_1 < j_2 < \cdots < j_r \leq n$, when written as the sums

$$\omega(x) = \sum_{j_1 < \cdots < j_r} f_{j_1 \ldots j_r}(x) dx_{j_1} \wedge \cdots \wedge dx_{j_r}$$

of multiples of the basis vectors $dx_{j_1} \wedge \cdots \wedge dx_{j_r}$ that span the exterior algebra $\Lambda^r$ of $r$-vectors over the vector space $\mathbb{R}^n$ as discussed in Appendix A, are called **differential forms of degree** $r$ in $U$, or for short just differential $r$-forms in $U$. These are continuous differential forms if the coefficients $f_{j}(x)$ or $f_{j_1 \ldots j_r}(x)$ are continuous functions in $U$, and are $C^k$ differential forms if these coefficients are $C^k$ functions in $U$. For differential forms of degree $r > 1$ it is sometimes convenient to introduce the additional vectors $dx_{j_1} \wedge \cdots \wedge dx_{j_r}$ for any order of the indices $i_1, \ldots, i_r$, where these satisfy the skew-symmetry conditions (A.2) and (A.3); thus each of these additional vectors is equal to either one of the basic vectors of $\Lambda^r$ or the negative of one of these basic vectors. Differential $r$-forms in $U$ can be written alternatively in terms of these vectors as sums

$$\omega(x) = \sum_{i_1, \ldots, i_r=1}^n g_{i_1 \ldots i_r}(x) dx_{i_1} \wedge \cdots \wedge dx_{i_r}.$$ 

The coefficients $g_{i_1 \ldots i_r}(x)$ in (5.15) are of course uniquely determined, while the coefficients $g_{i_1 \ldots i_r}(x)$ in (5.16) are not uniquely determined. For example a differential form of degree 2 in $\mathbb{R}^3$ can be written

$$\omega(x) = g_{12}(x)dx_1 \wedge dx_2 + g_{13}(x)dx_1 \wedge dx_3 + g_{21}(x)dx_2 \wedge dx_1 + g_{23}(x)dx_2 \wedge dx_3 + g_{31}(x)dx_3 \wedge dx_1 + g_{32}(x)dx_3 \wedge dx_2,$$
 CHAPTER 5. DIFFERENTIAL FORMS

or since \( dx_{i_1} \wedge dx_{i_2} = -dx_{i_2} \wedge dx_{i_1} \) it can be written alternatively

\[
\omega(x) = (g_{12}(x) - g_{21}(x))dx_{i_1} \wedge dx_{i_2} + (g_{13}(x) - g_{31}(x))dx_{i_1} \wedge dx_{i_3} + (g_{23}(x) - g_{12}(x))dx_{i_2} \wedge dx_{i_3}
\]

thus as the differential form

\[
\omega(x) = f_{12}(x)dx_{i_1} \wedge dx_{i_2} + f_{13}(x)dx_{i_1} \wedge dx_{i_3} + f_{23}(x)dx_{i_2} \wedge dx_{i_3}
\]

where

\[
f_{12}(x) = g_{12}(x) - g_{21}(x), \quad f_{13}(x) = g_{13}(x) - g_{31}(x), \quad f_{23}(x) = g_{23}(x) - g_{32}(x).
\]

The functions \( f_{i_1i_2} \) are uniquely determined, while it is possible for instance to replace \( g_{12}(x) \) and \( g_{21}(x) \) by \( g_{12}(x) + g(x) \) and \( g_{21}(x) + g(x) \) for any function \( g(x) \) and still have the same differential 2-form. Of course since \( dx_{i_1} \wedge dx_{i_2} = dx_{i_2} \wedge dx_{i_1} \) the coefficients \( g_{ij}(x) \) are totally irrelevant. A differential form written as (5.16) is said to be in the **reduced** form, while a differential form written as (5.15) is said to be in the **unreduced** form. It is often convenient to use **multi-index notation**, writing (5.16) in the form

\[
\omega(x) = \sum_I g_I(x) dx_I \quad \text{for} \quad I = (i_1, i_2, \ldots, i_r),
\]

where \( g_I(x) \) is an abbreviation for \( g_{i_1 \ldots i_r}(x) \) and correspondingly \( dx_I \) is an abbreviation for \( dx_{i_1} \wedge \cdots \wedge dx_{i_r} \); if not specified otherwise the assumption is that in multi-index notation the differential forms are in the unreduced form, so the indices \( i_1, i_2, \ldots, i_k \) range independently over all integer values in the range \( 1 \leq i_j \leq n \) rather than just over ordered sets of integers. The set of \( \mathcal{C}^k \) differential forms of degree \( r \) in the set \( U \) is denoted by \( \Lambda^r(U) \); however it is quite common to drop the subscript \( k \) when the regularity of the differential forms is either clear from the context or irrelevant. It is customary to identify the set \( \Lambda^r(U) \) of differential forms of degree \( 0 \) with the set of \( \mathcal{C}^k \) of functions in \( U \), as on page 77; and differential forms \( \omega(x) \in \Lambda^r(U) \) for open subsets \( U \subset \mathbb{R}^n \) have the simple reduced form

\[
\omega(x) = f_{i_1 \ldots i_n}(x) dx_{i_1} \wedge \cdots \wedge dx_{i_n}
\]

so they can also be identified with the set of \( \mathcal{C}^k \) functions \( f_{i_1 \ldots i_n}(x) \) in \( U \). Of course \( \Lambda_r(U) = 0 \) whenever \( r > n \), as a consequence of (A,2).

Differential forms can be combined algebraically, either by linear combinations or by exterior multiplication. The **linear combination** \( a\omega(x) + b\sigma(x) \) of two differential \( r \)-forms \( \omega(x) = \sum_I f_I(x) dx_I \) and \( \sigma(x) = \sum_I g_I(x) dx_I \) for \( a, b \in \mathbb{R} \) is defined to be the differential \( r \)-form

\[
(a\omega + b\sigma)(x) = \sum_I (a f_I(x) + b g_I(x)) \ dx_I;
\]

under this operation the set \( \Lambda^r(U) \) has the natural structure of a real vector space, although an infinite-dimensional vector space. The **exterior product** \( \omega(x) \wedge \sigma(x) \) of an \( r \)-form \( \omega(x) = \sum_I f_I(x) dx_I \in \Lambda^r(U) \) and an \( s \)-form \( \sigma(x) = \sum_I g_I(x) dx_I \in \Lambda^s(U) \), which is defined as the \( r + s \)-form

\[
(\omega \wedge \sigma)(x) = \sum_{I,J} f_I(x) g_J(x) dx_I \wedge dx_J \in \Lambda^{r+s}(U).
\]
5.2. DIFFERENTIAL FORMS

This operation is clearly associative, in the sense that \( \omega_1 \wedge (\omega_2 \wedge \omega_3) = (\omega_1 \wedge \omega_2) \wedge \omega_3 \)
for any differential forms \( \omega_i \in \Lambda^i(U) \); however it is not commutative but rather satisfies the condition that

\[
(\omega \wedge \sigma)(\mathbf{x}) = (-1)^{rs}(\sigma \wedge \omega)(\mathbf{x}) \quad \text{if} \quad \omega \in \Lambda^s(U) \quad \text{and} \quad \sigma \in \Lambda^r(U);
\]

for \( dx_i \wedge dx_j = (-1)^{rs}dx_j \wedge dx_i \) since to rearrange the left-hand side each differential \( dx_j \) must be permuted with each of the \( r \) differentials \( dx_i \), a change of sign \((-1)^r\) for each of the \( s \) differentials \( dx_i \). In particular the exterior product of a 0-form, a function \( f(\mathbf{x}) \), and an \( r \)-form \( \omega(\mathbf{x}) = \sum_j g_j(\mathbf{x})dx_j \) is the \( r \)-form \( (f \omega)(\mathbf{x}) = \sum_j f(\mathbf{x})g_j(\mathbf{x})dx_j \).

Exterior products and linear operations are related by

\[
(\Lambda \wedge I)(\omega) = (\Lambda \wedge I)\omega;
\]

\( I \omega = \omega \) for any differential forms \( \omega \).

This operation is clearly associative, in the sense that \( \omega_1 \wedge (\omega_2 \wedge \omega_3) = (\omega_1 \wedge \omega_2) \wedge \omega_3 \);

for \( dx_i \wedge dx_j = (-1)^{rs}dx_j \wedge dx_i \) since to rearrange the left-hand side each differential \( dx_j \) must be permuted with each of the \( r \) differentials \( dx_i \), a change of sign \((-1)^r\) for each of the \( s \) differentials \( dx_i \). In particular the exterior product of a 0-form, a function \( f(\mathbf{x}) \), and an \( r \)-form \( \omega(\mathbf{x}) = \sum_j g_j(\mathbf{x})dx_j \) is the \( r \)-form \( (f \omega)(\mathbf{x}) = \sum_j f(\mathbf{x})g_j(\mathbf{x})dx_j \).

Exterior products and linear operations are related by

\[
(a_1 \omega_1 + a_2 \omega_2) \wedge \sigma = a_1 \omega_1 \wedge \sigma + a_2 \omega_2 \wedge \sigma
\]

for differential forms \( \omega_i \in \Lambda^i(U) \) and \( \sigma \in \Lambda^r(U) \), as is evident from their definition. Under these two operations the direct sum

\[
\Lambda(U) = \bigoplus_{i=1}^{n} \Lambda^i(U)
\]

of all the nontrivial vector spaces \( \Lambda^i(U) \) has the structure of an anticommutative associative algebra, which is merely a summary of the preceding observations.

These algebraic operations on differential forms are quite familiar in the special case of three dimensional vector spaces. A vector field in an open set \( U \subset \mathbb{R}^3 \) is naturally associated with a differential 1-form, as already discussed; but since \( \dim \Lambda^2(\mathbb{R}^3) = 3 \) a vector field also can be identified with a differential 2-form. Thus to a vector field \( \mathbf{f}(\mathbf{x}) = \{f_i(\mathbf{x})\} \) in \( U \subset \mathbb{R}^3 \) there are naturally associated the two differential forms

\[
\omega_1(\mathbf{x}) = f_1(\mathbf{x})dx_1 + f_2(\mathbf{x})dx_2 + f_3(\mathbf{x})dx_3 \in \Lambda^1(U)
\]

and

\[
\Omega_2(\mathbf{x}) = f_1(\mathbf{x})dx_2 \wedge dx_3 + f_2(\mathbf{x})dx_3 \wedge dx_1 + f_3(\mathbf{x})dx_1 \wedge dx_2 \in \Lambda^2(U),
\]

where the notation for the 2-form \( \Omega_2(\mathbf{x}) \), involving cyclic permutations of the indices, is chosen to fit into the classical form for the algebraic operations. The exterior product of the 1-form associated to a vector field \( \mathbf{f} : U \rightarrow \mathbb{R}^3 \) and the 2-form associated to a vector field \( \mathbf{g} : U \rightarrow \mathbb{R}^3 \) is easily seen to be the 3-form

\[
\omega_3(\mathbf{x}) \wedge \Omega_2(\mathbf{x}) = \left( f_1(\mathbf{x})g_1(\mathbf{x}) + f_2(\mathbf{x})g_2(\mathbf{x}) + f_3(\mathbf{x})g_3(\mathbf{x}) \right) dx_1 \wedge dx_2 \wedge dx_3;
\]

this is the 3-form associated to the dot product or inner product \( \mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x}) \) of the two vector fields \( \mathbf{f}(\mathbf{x}) \) and \( \mathbf{g}(\mathbf{x}) \), so that

\[
\omega_3(\mathbf{x}) \wedge \Omega_2(\mathbf{x}) = \mathbf{f}(\mathbf{x}) \cdot \mathbf{g}(\mathbf{x}) dx_1 \wedge dx_2 \wedge dx_3,
\]

and in this case the exterior product is commutative. On the other hand a straightforward calculation shows that the exterior product of the 1-forms associated to two vector fields \( \mathbf{f} : U \rightarrow \mathbb{R}^3 \) and \( \mathbf{g} : U \rightarrow \mathbb{R}^3 \) is the 2 form

\[
\omega_2(\mathbf{x}) \wedge \omega_2(\mathbf{x}) = \det \begin{pmatrix} f_2(\mathbf{x}) & f_3(\mathbf{x}) \\ g_2(\mathbf{x}) & g_3(\mathbf{x}) \end{pmatrix} dx_2 \wedge dx_3 + \det \begin{pmatrix} f_3(\mathbf{x}) & f_1(\mathbf{x}) \\ g_3(\mathbf{x}) & g_1(\mathbf{x}) \end{pmatrix} dx_3 \wedge dx_1
\]

\[
+ \det \begin{pmatrix} f_1(\mathbf{x}) & f_2(\mathbf{x}) \\ g_1(\mathbf{x}) & g_2(\mathbf{x}) \end{pmatrix} dx_1 \wedge dx_2;
\]
which is the 2 form associated to the vector field \( \mathbf{f}(x) \times \mathbf{g}(x) \), traditionally called **cross-product** of the two vector fields \( \mathbf{f}(x) \) and \( \mathbf{g}(x) \) and defined by

\[
(5.31) \quad \mathbf{f}(x) \times \mathbf{g}(x) = \left\{ \det \begin{pmatrix} f_2(x) & f_3(x) \\ g_2(x) & g_3(x) \end{pmatrix}, \det \begin{pmatrix} f_3(x) & f_1(x) \\ g_3(x) & g_1(x) \end{pmatrix}, \det \begin{pmatrix} f_1(x) & f_2(x) \\ g_1(x) & g_2(x) \end{pmatrix} \right\}.
\]

Thus in these terms

\[
(5.32) \quad \omega_f(x) \wedge \omega_g(x) = \Omega_{f \times g}(x),
\]

and in this case the exterior product is anticommutative in the sense that

\[
(5.33) \quad \mathbf{f}(x) \times \mathbf{g}(x) = -\mathbf{g}(x) \times \mathbf{f}(x).
\]

It is clear from (5.31) that the cross-product \( \mathbf{f}(x) \times \mathbf{g}(x) \) is trivial at a point \( x \) precisely when the two vectors \( \mathbf{f}(x) \) and \( \mathbf{g}(x) \) are linearly dependent. It is also clear from (5.31) that for any other vector field \( \mathbf{h}(x) \) in \( U \)

\[
(5.34) \quad \mathbf{h}(x) \cdot (\mathbf{f}(x) \times \mathbf{g}(x)) = \det \begin{pmatrix} h_1(x) & h_2(x) & h_3(x) \\ f_1(x) & f_2(x) & f_3(x) \\ g_1(x) & g_2(x) & g_3(x) \end{pmatrix},
\]

which provides an alternative characterization of the cross-product. An evident consequence of (5.34) is that \( \mathbf{f}(x) \cdot (\mathbf{f}(x) \times \mathbf{g}(x)) = \mathbf{g}(x) \cdot (\mathbf{f}(x) \times \mathbf{g}(x)) = 0 \); hence the cross-product \( \mathbf{f}(x) \times \mathbf{g}(x) \) is a vector in \( \mathbb{R}^3 \) that is perpendicular to both \( \mathbf{f}(x) \) and \( \mathbf{g}(x) \) if these vectors are linearly independent, so its is determined uniquely up to sign. To determine the sign, it follows from (5.31) that if \( \mathbf{f}(x_0) = \{1 \quad 0 \quad 0\} \) and \( \mathbf{g}(x_0) = \{0 \quad 1 \quad 0\} \) at a point \( x_0 \in U \) then \( \mathbf{f}(x_0) \times \mathbf{g}(x_0) = \{1 \quad 0 \quad 0\} \); consequently the three vectors \( \mathbf{f}(x_0), \mathbf{g}(x_0), \mathbf{f}(x_0) \times \mathbf{g}(x_0) \) have the same orientation as the three vectors \( dx_1, dx_2, dx_3 \in \Lambda^1 \mathbb{R}^3 \), the natural orientation associated of the vector space \( \mathbb{R}^3 \) with this choice of coordinates.

Differential forms are transformed under mappings between various vector spaces \( \mathbb{R}^n \) in quite natural ways. If \( \phi : D \rightarrow E \) is a mapping from an open subset \( D \subset \mathbb{R}^m \) into an open subset \( E \subset \mathbb{R}^n \) then to any function \( f \) in the set \( E \) there can be associated the **induced function** \( \phi^*(f) \) in the set \( D \), the function defined by

\[
(5.35) \quad \phi^*(f)(x) = f(\phi(x)) \quad \text{for all points} \ x \in D.
\]

It is clear that \( \phi^*(c_1 f_1 + c_2 f_2) = c_1 \phi^*(f_1) + c_2 \phi^*(f_2) \) for any functions \( f_1, f_2 \) defined in \( D \) and any constants \( c_1, c_2 \in \mathbb{R} \). If the mapping \( \phi \) and the function \( f \) are \( C^k \) then so is the induced function \( \phi^*(f) \), and in that case associating to a function \( f \in \Lambda^k_0(E) \) the induced function \( \phi^*(f) \in \Lambda^k_0(D) \) is a linear mapping

\[
(5.36) \quad \phi^* : \Lambda^k_0(E) \rightarrow \Lambda^k_0(D).
\]

More generally to any mapping \( f : E \rightarrow \mathbb{R}^N \) there can be associated the **induced mapping** \( \phi^*(f) \) in the set \( D \), the mapping defined by

\[
(5.37) \quad \phi^*(f)(x) = f(\phi(x)) \quad \text{for all points} \ x \in D;
\]

and again if the mappings \( \phi \) and \( f \) are \( C^k \) then so is the induced mapping \( \phi^*(f) \). However the differential forms associated to mappings transform somewhat differently, since the bases for the spaces \( \mathbb{R}^m \) containing \( D \) and \( \mathbb{R}^n \) containing \( E \) are quite different.
5.2. DIFFERENTIAL FORMS

Indeed if \(x_1, \ldots, x_n\) are the coordinates in \(\mathbb{R}^n\) the vectors \(dx_1, \ldots, dx_n\) are the associated basis for that vector space, and if \(y_1, \ldots, y_n\) are the coordinates in \(\mathbb{R}^n\) then \(dy_1, \ldots, dy_n\) are the associated basis for \(\mathbb{R}^n\). If the mapping \(\phi : D \rightarrow E\) is \(C^k\) then the coordinates \(y_i\) of the point \(\phi(x)\) are continuously differentiable functions, and by the chain rule the derivatives of these functions are

\[
y'_i(x) = \left( \frac{\partial y_i}{\partial x_1}, \frac{\partial y_i}{\partial x_2}, \ldots, \frac{\partial y_i}{\partial x_n} \right),
\]

which can be written equivalently in terms of the bases \(dx_i\) of the space \(\mathbb{R}^n\) and \(dy_i\) of the space \(\mathbb{R}^n\) as the linear relation

\[
dy_i = \sum_{j=1}^{m} \frac{\partial y_i}{\partial x_j} dx_j.
\]

Consequently if \(\phi : D \rightarrow E\) is a \(C^k\) mapping then there can be associated to any \(C^k\) differential \(r\)-form \(\omega = \sum_i f_i(y)dy_i \in \Lambda^r_k(E)\) the induced differential form, \(\phi^* \omega \in \Lambda^r_{k-1}(D)\) on \(D\) defined by

\[
\phi^* \left( \sum_i f_i(y)dy_i \right) = \sum_{i,j} f_i(\phi(x)) \frac{\partial y_i}{\partial x_j} dx_j \wedge \cdots \wedge dx_r,
\]

This is sometimes called the pullback of the differential form \(\omega\) from \(E\) to \(D\). It is clear that \(\phi^*(c_1\omega_1 + c_2\omega_2) = c_1\phi^*(\omega_1) + c_2\phi^*(\omega_2)\) for any differential \(r\)-forms \(\omega_1, \omega_2\) in \(D\) and any constants \(c_1, c_2 \in \mathbb{R}\), so the mapping \(\phi^*\) is a linear mapping

\[
\phi^*: \Lambda^r_{k}(E) \longrightarrow \Lambda^r_{k-1};
\]

and it is also clear that

\[
\phi^*(\omega \wedge \sigma) = \phi^*(\omega) \wedge \phi^*(\sigma)
\]

for any differential \(r\)-form \(\omega\) and any differential \(s\)-form \(\sigma\) on \(E\). For example if \(D, E \subset \mathbb{R}^n\) are both subsets of vector spaces of the same dimension and \(\omega \in \Lambda^r_{k-1}(E)\) is a differential \(n\)-form in \(E\) then \(\omega(y) = f(y)dy_1 \wedge \cdots \wedge dy_n\) for some function \(f\) in \(E\) and (5.39) reduces to

\[
\phi^*( f(y)dy_1 \wedge \cdots \wedge dy_n) = \sum_j f(\phi(x)) \frac{\partial y_1}{\partial x_j} \cdots \frac{\partial y_n}{\partial x_j} dx_j \wedge \cdots \wedge dx_n = f(\phi(x)) \det \phi'(x) dx_1 \wedge \cdots \wedge dx_n.
\]

This illustrates nicely the role of exterior algebras as convenient tools in handling determinants, and in a sense underlies all the applications of differential forms in analysis and geometry.

**Theorem 5.5** If \(\phi : D \rightarrow E\) is a \(C^k\) mapping between open subsets \(D \subset \mathbb{R}^m\) and \(E \subset \mathbb{R}^n\) while \(\psi : E \rightarrow F\) is a \(C^{k-1}\) mapping between open subsets \(E \subset \mathbb{R}^n\) and \(F \subset \mathbb{R}^p\) then \((\psi \circ \phi)^*: \Lambda^r_{(k)}(F) \rightarrow \Lambda^r_{(k-2)}(D)\) is the composition \((\psi \circ \phi)^* = \phi^* \circ \psi^*\) of the mappings \(\phi^*: \Lambda^r_{(k)}(E) \rightarrow \Lambda^r_{(k)}(D)\) and \(\psi^*: \Lambda^r(F)_{(k)} \rightarrow \Lambda^r(E)_{(k)}\).
Consequently, \( \psi^*(\omega)(x) = \sum_{i,j} f_i(\psi(x)) \frac{\partial y_j}{\partial x_i}(x) dx_j \)
and
\[ \phi^* (\psi^*(\omega))(t) = \sum_{i,j,k} f_i(\phi(t)) \frac{\partial y_j}{\partial x_i}(\phi(t)) \frac{\partial x_k}{\partial t}(t). \]

On the other hand for the composite mapping \( \psi \circ \phi \)
\[ (\psi \circ \phi)^*(\omega)(t) = \sum_{i,k} f_i (\psi \circ \phi)(t) \frac{\partial y_i}{\partial t}(t) dt, \]
and it follows from the chain rule that
\[ \frac{\partial y_i}{\partial t} = \sum_j \frac{\partial y_i}{\partial x_j}(\phi(t)) \frac{\partial x_j}{\partial t}(t). \]

Consequently \( (\psi \circ \phi)^*(\omega) = \phi^* (\psi^*(\omega)) \), which suffices for the proof.

In addition to the algebraic operations and change of variables formula for differential forms discussed so far, there are the analytic operations of differentiation and integration of differential forms. The exterior derivative of a differential form \( \omega(x) \in \Lambda^r_k(U) \) for \( k > 0 \) is a differential form \( d\omega(x) \in \Lambda^{r+1}_{k-1}(U) \). Explicitly, for a differential form \( f(x) \in \Lambda^0_k \) of degree 0, a function defined in the subset \( U \subset \mathbb{R}^n \), the exterior derivative is defined to be the ordinary derivative of that function expressed as a differential form, so

\[
(5.43) \quad df(x) = \sum_{j=1}^n \partial_j f(x) dx_j \quad \text{if} \ f \in \Lambda^0_k(U).
\]

For a differential form \( \omega(x) = \sum_I f_I(x) dx_I \in \Lambda^r_k \) of degree \( r > 0 \) the exterior derivative is defined by

\[
(5.44) \quad d\omega(x) = \sum_I df_I(x) \wedge dx_I,
\]
or in more detail

\[
(5.45) \quad d \left( \sum_{i_1, \ldots, i_r=1}^{n} f_{i_1 \ldots i_r}(x) \, dx_{i_1} \wedge \cdots \wedge dx_{i_r} \right) = \sum_{j, i_1, \ldots, i_r=1}^{n} \partial_j f_{i_1 \ldots i_r}(x) \, dx_j \wedge dx_{i_1} \wedge \cdots \wedge dx_{i_r}.
\]

A clear consequence of this is that \( d\omega = 0 \) for any \( n \)-form \( \omega \) in an open subset \( U \subset \mathbb{R}^n \).

**Theorem 5.6** Exterior differentiation is a linear mapping

\[
(5.46) \quad d : \Lambda^r_k(U) \rightarrow \Lambda^{r+1}_{k-1}(U)
\]
of \( C^k \) differential \( r \)-forms in an open subset \( U \subset \mathbb{R}^n \) for \( k \geq 1 \), such that

(i) \( d (\omega(x) \wedge \sigma(x)) = d\omega(x) \wedge \sigma(x) + (-1)^r \omega(x) \wedge d\sigma(x) \)
if \( \omega(x) \in \Lambda^r_k(D) \) and \( \sigma(x) \in \Lambda^1_k(D) \) for \( k \geq 1 \), while

(ii) \( d d\omega(x) = 0 \) for any differential form \( \omega \in \Lambda^r(D) \).
5.2. DIFFERENTIAL FORMS

**Proof:** That exterior differentiation is a linear mapping on $r$-forms is an immediate consequence of the definitions (5.44) and (5.45).

(i) For functions it follows from the usual rules for differentiation of a product that

$$d(f(x)g(x)) = f(x)dg(x) + g(x)df(x).$$

In general for differential forms $\omega = \sum f_i(x)dx_i \in \Lambda^r(D)$ and $\sigma = \sum g_j(x)dx_j \in \Lambda^s(D)$ it follows from the special case of differentiation of functions that

$$d(\omega \wedge \sigma) = \sum_{i,j} \left( f_i(x)g_j(x)dx_i \wedge dx_j \right) = \sum_{i,j} \left( \frac{\partial f_i(x)}{\partial x_j} \right) dx_i \wedge dx_j,$$

since the 1-form $dg_j(x)$ must be moved across the $r$-differentials in $dx_i$. (ii) If $\omega(x) = \sum f_i(x)dx_i \in \Lambda^r_{(i)}$ for $k \geq 2$ then

$$d(\omega(x)) = \sum_{i} \left( \frac{\partial f_i(x)}{\partial x_j} \right) dx_i \wedge dx_j,$$

which is zero since $\partial_{jk}f_i(x)$ is symmetric in the indices $j,k$ while $dx_k \wedge dx_j$ is skew-symmetric in these indices. That suffices for the proof.

A differential form $\omega(x)$ is said to be **closed** if $d\omega(x) = 0$, and it is said to be **exact** if $\omega(x) = d\sigma(x)$ for some differential form $\sigma(x)$. A consequence of (ii) in the preceding theorem is that any exact differential form is necessarily a closed differential form; but the converse is not true in general. For example, in the complement of the origin in the plane $\mathbb{R}^2$ of the variables $x_1, x_2$ the function $\theta(x_1, x_2) = \tan^{-1}(x_2/x_1)$, the angle in polar coordinates, is well defined in any half-plane bounded by a line through the origin, such as the half-planes $x_1 > 0$ and $x_1 < 0$, but it cannot be defined in the entire complement of the origin as a single-valued function. However

$$\omega(x) = d\theta(x_1, x_2) = -\frac{x_2}{x_1^2 + x_2^2} dx_1 + \frac{x_1}{x_1^2 + x_2^2} dx_2$$

is a well defined closed differential 1-form in the complement of the origin in $\mathbb{R}^2$. If $\omega(x)$ were an exact differential form in the full complement of the origin it would be the exterior derivative $\omega(x) = df(x)$ of a well defined function $f(x)$ in the complement of the origin; but then $df(x) = d\theta(x)$ so that $d(f(x) - \theta(x)) = 0$ hence $f(x) - \theta(x) = c$ is a constant in the complement of the origin, which would imply that $\theta(x) = f(x) - c$ is a well defined function in the complement of the origin, a contradiction. However it is at least true locally that any closed form is exact.

**Theorem 5.7 (Poincaré’s Lemma)** If $\omega(x) \in \Lambda^r_{(1)}(\Delta)$ is a closed $C^1$ differential form of degree $r > 0$ in a cell $\Delta \subset \mathbb{R}^n$ then there is a $C^2$ differential form $\sigma(x) \in \Lambda^{r-1}_{(1)}(\Delta)$ of degree $r - 1$ in $\Delta$ such that $\omega(x) = d\sigma(x)$.
For some differential form \( \omega \) which involves just the differentials \( dx \) then are linearly independent since \( \partial_i \omega = \partial_i f(x)dx_1 \wedge dx_2 \) for some index \( k \geq 1 \), and consider a differential \( r \)-form \( \omega(x) \) that involves only the differentials \( dx_1, \ldots, dx_k \).

This differential form can be written as the sum

\[
\omega(x) = \omega_1(x) \wedge dx_{k+1} + \omega_2(x)
\]

where the differential forms \( \omega_1(x) \) and \( \omega_2(x) \) involve only the differentials \( dx_1, \ldots, dx_k \), so that

\[
\omega_1(x) = \sum_I f_I(x)dx_I \quad \text{and} \quad \omega_2(x) = \sum_J g_J(x)dx_J \quad \text{where} \quad I, J \subset \{1, \ldots, k\};
\]

assume that these differential forms are the reduced forms and hence that \( 1 \leq i_1 < i_2 < \ldots < i_{r-1} \leq k \) and \( 1 \leq j_1 < j_2 < \cdots < j_r \leq k \). By assumption

\[
0 = \partial_i \omega(x) = \partial_i (\omega(x) \wedge dx_{k+1} + \omega_2(x))
\]

\[
= \sum_{l,i} \partial_i f_l(x)dx_i \wedge dx_l \wedge dx_{k+1} + \sum_{j,l} \partial_j g_j(x)dx_j \wedge dx_l \quad \text{where} \quad I, J \subset \{1, \ldots, k\}.
\]

The only terms in this expression that involve \( dx_{k+1} \) and \( dx_l \) for \( l > k + 1 \) are just in the first sum and just for the index \( i = l \); and the basis vectors \( dx_1 \wedge dx_l \wedge dx_{k+1} \) are linearly independent since \( I \subset \{1, \ldots, k\} \) where \( i_1 < i_2 < \cdots < i_{r-1} \) and \( k < k+1 < l \); consequently \( \partial_i f_l(x) = 0 \) for \( l > k + 1 \), so \( f_l(x) = f_l(x_1, \ldots, x_{k+1}) \) is a function just of the variables \( x_1, \ldots, x_{k+1} \) for all \( l > k + 1 \). The integrals

\[
h_l(x_1, \ldots, x_{k+1}) = \int_a^{x_{k+1}} f_l(x_1, \ldots, x_k, t)dt \quad \text{where} \quad I \subset \{1, \ldots, k\}
\]

then are \( C^2 \) functions in the cell \( \Delta \) for which \( \partial_{k+1} h_l(x) = f_l(x) \); and consequently the differential \((r - 1)\)-form

\[
\omega_3(x) = \sum_I g_I(x)dx_I \quad \text{where} \quad I \subset \{1, \ldots, k\},
\]

which involves just the differentials \( dx_1, \ldots, dx_k \), satisfies

\[
d\omega_3(x) = \sum_I \left( \sum_{i=1}^k \partial_i h_i(x)dx_i \wedge dx_I + \partial_{i+1} h_i(x)dx_{k+1} \wedge dx_I \right)
\]

\[
= \sum_I \left( \sum_{i=1}^k \partial_i h_i(x)dx_i \wedge dx_I + f_i(x)dx_{k+1} \wedge dx_I \right)
\]

\[
= \omega_3(x) + (-1)^k \omega_1(x) \wedge dx_{k+1}
\]

for some differential form \( \omega_3(x) \) that also involves only the differentials \( dx_1, \ldots, dx_k \). Consequently \( \omega(x) - (-1)^k d\omega_3(x) = \omega_2(x) - (-1)^k \omega_4(x) \) involves only the differentials \( dx_1, \ldots, dx_k \) and \( d(\omega(x) - (-1)^k d\omega_3(x)) = 0 \); it then follows from the induction hypothesis there is a differential form \( \omega_5(x) \) in the cell \( \Delta \) such that

\[
\omega(x) - (-1)^k d\omega_3(x) = d\omega_5(x),
\]
hence such that $\omega(x) = d\sigma(x)$ where $\sigma(x) = (-1)^k \omega_3(x) + \omega_5(x)$. That establishes the induction step and thereby concludes the proof.

The preceding theorem is an example of an integrability theorem for a linear system of partial differential equations. For example, the existence of a function $h$ such that $dh = \sigma$ for a 1-form $\sigma = \sum_{j=1}^n f_j(x)dx_j$ in $\mathbb{R}^n$ amounts to the solution of the system of partial differential equations $\partial_j h(x) = f_j(x)$. It is clear that if there is a $C^2$ solution $h(x)$ then $\partial_j f_i(x) = \partial_i \partial_j h(x) = \partial_i \partial_j h(x) = \partial_i f_j(x)$; that is just the condition that $d\omega(x) = 0$, which by Poincaré’s Lemma is also a sufficient condition for the existence of the solution $h(x)$. There are corresponding interpretations for Poincaré’s Lemma for differential forms of higher degree, although they are a bit more complicated to state without using the machinery of differential forms. For example, the existence of a function $h$ for differential forms of higher degree, although they are a bit more complicated to state without using the machinery of differential forms. The exterior derivatives of differential forms in $\mathbb{R}^3$ also have classical forms, which are probably quite familiar from physics. For the 2-form $\Omega_f(x)$ associated to a vector field $f(x)$ in an open set $U \subset \mathbb{R}^3$ as in (5.27) it follows quite directly from the definition of exterior derivative that

$$d\Omega_f(x) = \left( \partial_1 f_3(x) + \partial_2 f_1(x) + \partial_3 f_2(x) \right) dx_1 \wedge dx_2 \wedge dx_3;$$

the coefficient of this differential 3-form is a function known as the divergence of the vector field $f(x)$ and often denoted by $\text{div} f(x)$, so that

$$\text{div} f(x) = \frac{\partial f_1(x)}{\partial x_1} + \frac{\partial f_2(x)}{\partial x_2} + \frac{\partial f_3(x)}{\partial x_3}.$$ 

A traditional alternative notation is expressed in terms of the vector differential operator defined by

$$\nabla = \left\{ \frac{\partial}{\partial x_1}, \frac{\partial}{\partial x_2}, \frac{\partial}{\partial x_3} \right\}.$$ 

When applied to functions $\nabla f(x)$ is the gradient of the function $f(x)$, as discussed on page 23; and the dot product of this operator with a vector field is the function

$$\nabla \cdot f(x) = \frac{\partial}{\partial x_1} f_1(x) + \frac{\partial}{\partial x_2} f_2(x) + \frac{\partial}{\partial x_3} f_3(x) = \text{div} f(x).$$

In terms of this operator (5.47) can be rewritten

$$d\Omega_f(x) = \left\{ \begin{array}{l} \text{div} f(x) \ dx_1 \wedge dx_2 \wedge dx_3 \\
\n\n= \nabla \cdot f(x) \ dx_1 \wedge dx_2 \wedge dx_3. \end{array} \right.$$ 

For the 1-form $\omega_f(x)$ associated to a vector field $f(x)$ in an open set $U \subset \mathbb{R}^3$ as in (5.27) it follows by a straightforward calculation that

$$d\omega_f(x) = (\partial_1 f_3(x) - \partial_3 f_1(x)) dx_2 \wedge dx_3 + (\partial_1 f_1(x) - \partial_1 f_3(x)) dx_3 \wedge dx_1 + (\partial_2 f_1(x) - \partial_1 f_2(x)) dx_1 \wedge dx_2.$$ 

On the other hand it follows from (5.31) that

$$\nabla \times f(x) = \left( \partial_2 f_3(x) - \partial_3 f_2(x) \right) \partial_3 f_1(x) - \partial_1 f_3(x) \partial_1 f_2(x) - \partial_2 f_1(x)$$

and consequently

$$d\omega_f(x) = \Omega_{\nabla \times f}(x).$$
Theorem 5.8 If \( \phi : D \rightarrow E \) is a \( C^k \) mapping from an open subset \( D \subset \mathbb{R}^m \) into an open subset \( E \subset \mathbb{R}^n \) then for any differential form \( \omega \in \Lambda^*(E) \)

\[
(5.56) \quad d \phi^*(\omega) = \phi^*(d\omega).
\]

Proof: For a differential form \( \omega \in \Lambda^*(E) \) written explicitly as \( \omega = \sum I f_I(y) dy_I \) it follows from the definition (5.39) that

\[
\phi^*(\omega) = \sum I f_I(\phi(x)) \frac{\partial y_{i_1}}{\partial x_{j_1}} \cdots \frac{\partial y_{i_n}}{\partial x_{j_r}} dx_{j_1} \wedge \cdots \wedge dx_{j_r}.
\]

The exterior derivative of \( \omega \) is

\[
d \omega = \sum I, k \frac{\partial f_I}{\partial y_k} dy_k \wedge dy_I
\]

and it follows from the definition (5.39) that

\[
\phi^*(d\omega) = \sum I, j, k, l \frac{\partial f_I}{\partial y_k} \left( \frac{\partial y_{i_1}}{\partial x_{j_1}} \cdots \frac{\partial y_{i_r}}{\partial x_{j_r}} \right) dx_{j_1} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_r}.
\]

On the other hand

\[
d \phi^*(\omega) = \sum I, j, l \frac{\partial}{\partial x_l} \left( f_I(\phi(x)) \frac{\partial y_{i_1}}{\partial x_{j_1}} \cdots \frac{\partial y_{i_r}}{\partial x_{j_r}} \right) dx_{j_1} \wedge dx_{j_1} \wedge \cdots \wedge dx_{j_r} + \cdots
\]

where the remaining terms involve successively the second derivatives of the the partial derivatives \( \frac{\partial y_{i_2}}{\partial x_{j_1}} \frac{\partial y_{i_3}}{\partial x_{j_1}} \), and so on. These terms all vanish, since for example \( \frac{\partial^2 y_{i_1}}{\partial x_{j_1} \partial x_{j_1}} \) is symmetric in the indices \( l \) and \( j_1 \) while \( dx_l \wedge dx_{j_1} \) is skew-symmetric in these indices. That suffices for the proof.

Integrals of differential \( n \)-forms in open subsets \( E \subset \mathbb{R}^n \) can be defined, but only in terms of the orientation of the vector space \( \mathbb{R}^n \); so these integrals can be interpreted as oriented integrals, thus extending that notion from spaces of dimension 1 as defined in (4.35) to spaces of dimensions \( n > 1 \). The integral of a continuous \( n \)-form \( \omega(y) = f(y) dy_1 \wedge \cdots \wedge dy_n \) over a subset \( E \subset \mathbb{R}^n \) of an oriented vector space \( \mathbb{R}^n \) is defined in terms of the integral of the function \( f \) over \( E \) by

\[
(5.57) \quad \int_E f(y) dy_1 \wedge \cdots \wedge dy_n = \int_E f \quad \text{if } dy_1 \wedge \cdots \wedge dy_n \text{ is the orientation of } E.
\]

It follows from the properties of \( n \)-forms that

\[
(5.58) \quad \int_E f(y) dy_{i_1} \wedge \cdots \wedge dy_{i_n} = \text{sgn} \left( \begin{array}{cccc} i_2 & i_2 & \cdots & i_n \\ 1 & 2 & \cdots & n \end{array} \right) \int_E f(y) dy_1 \wedge \cdots \wedge dy_n
\]
where $\text{sgn} \left( \begin{array}{cccc} i_1 & i_2 & \cdots & i_n \\ 1 & 2 & \cdots & n \end{array} \right)$ is the sign of the permutation $(i_1, i_2, \cdots, i_n)$ of the integers $(1, 2, \cdots, n)$; thus under a permutation of the variables the integral of a differential form changes by the sign of the permutation.

**Theorem 5.9** If $\phi : D \to E$ is a $C^1$ homeomorphism between two connected open subsets $D, E \subset \mathbb{R}^n$ such that $\phi'(x) \neq 0$ at all points $x \in D$, then for any continuous differential $n$-form $\omega$ in $E$

\[(5.59) \quad \int_E \omega = \pm \int_D \phi^*(\omega),\]

where the sign is $+$ if the mapping $\phi$ preserves the orientation and $-$ if the mapping $\phi$ reverses the orientation.

**Proof:** Suppose that the vector space $\mathbb{R}^n$ containing $D$ is oriented by the differential form $dx_1 \wedge \cdots \wedge dx_n$ in terms of coordinates $x_1, \ldots, x_n$, and that the vector space $\mathbb{R}^n$ containing $E$ is oriented by the differential form $dy_1 \wedge \cdots \wedge dy_n$ in terms of coordinates $y_1, \ldots, y_n$. The integral of a continuous differential $n$-form $\omega(y) = f(y)dy_1 \wedge \cdots \wedge dy_n$ in $E$ is defined in terms of the integral of the coefficient $f$ by

\[(5.60) \quad \int_E \omega = \int_E f,\]

as in (5.57); and the integral of the induced differential form $\phi^*(\omega)$ in $D$, which by (5.42) has the explicit form $\phi^*(\omega)(\mathbf{x}) = f(\phi(\mathbf{x})) \det \phi'(\mathbf{x})dx_1 \wedge \cdots \wedge dx_n$, is defined correspondingly by

\[(5.61) \quad \int_D \phi^*(\omega) = \int_D (f \circ \phi) \det \phi'.\]

If the mapping $\phi$ preserves orientation then $\phi'(x) > 0$ at all points $x \in D$ and consequently $\det \phi'(x) = |\det \phi'(x)|$, while if the mapping $\phi$ reverses orientation then $\phi'(x) < 0$ at all points $x \in D$ and consequently $\det \phi'(x) = -|\det \phi'(x)|$; therefore (5.61) can be rewritten

\[(5.62) \quad \int_D \phi^*(\omega) = \pm \int_D (f \circ \phi)| \det \phi'|\]

where the sign is $+$ if the mapping $\phi$ preserves orientation and $-$ if the mapping $\phi$ reverses orientation. By the change of variables formula (4.46) in Theorem 4.13

\[(5.63) \quad \int_E f = \int_D (f \circ \phi)| \det \phi'|,\]

and (5.59) follows immediately from the combination of (5.60), (5.62) and (5.63), to conclude the proof.

### 5.3 Stokes’s Theorem

There are oriented integrals analogous to those introduced in (5.57) over suitable lower dimensional subsets of the spaces $\mathbb{R}^n$; and these involve the integrals of differential forms of lower degrees. An $m$-dimensional **singular cell** in an open subset
$U \subset \mathbb{R}^n$ is defined to be the image $\Gamma \subset U$ of a continuous mapping $\phi : \Delta \rightarrow U$ from an oriented closed cell $\Delta \subset \mathbb{R}^n$ into $U$. Much of the subsequent discussion will focus on $C^1$ singular cells, those for which the mappings $\phi$ are $C^1$ mappings, more precisely, those for which the mappings $\phi$ are $C^1$ mappings from an open neighborhood of the closed cell $\Delta$ into $U$. It is also common to consider $C^k$ singular cells for any integer $k > 0$ or even $C^\infty$ singular cells, those for which the mappings $\phi$ are of class $C^k$ for all integers $k > 0$. Unless explicitly assumed otherwise, though, singular cells are just the images of continuous mappings. In particular a 1-dimensional singular cell is a curve, as defined on page 77, since an interval $[0, 1]$ is a 1-dimensional cell that is naturally oriented. As in the case of curves, the mappings describing singular cells are not required to be one-to-one mappings, nor for $C^k$ singular cells is it required that the differential $\phi$ of the mapping have maximal rank; so singular cells are not necessarily submanifolds of $U$ even locally, and that is the reason for the adjective singular.

Although curves are naturally oriented in the direction of increasing values of the parameter, it is generally necessary to describe the orientation of the cell $\Delta \subset \mathbb{R}^n$ specifically. As usual, if the parameters in $\mathbb{R}^m$ are $t_1, \ldots, t_m$ then the natural orientation is $dt_1 \wedge \cdots \wedge dt_m$; but that may not be the appropriate orientation for some purposes, so even in that case it is necessary to specify the orientation.

If $\Gamma \subset U$ is an $m$-dimensional $C^1$ singular cell that is the image of a $C^1$ mapping $\phi : \Delta \rightarrow \Gamma$ then to a differential $m$-form $\omega \in \Lambda^m(U)$ in the open set $U$ there can be associated the induced differential form $\phi^*(\omega) \in \Lambda^m(\Delta)$, which is an $m$-form in the oriented cell $\Delta \subset \mathbb{R}^m$; and the integral of the differential form $\omega$ on the singular cell $\Gamma$ is defined to be the oriented integral

$$\int_{\Gamma, \phi} \omega = \int_{\Delta} \phi^*(\omega) = \int_{\Delta} f(t) \ dt_1 \wedge \cdots \wedge dt_m.$$  

As indicated by the notation, by definition this integral depends on the choice of the parametrization $\phi : \Delta \rightarrow \Gamma$. Two $C^1$ parametrizations $\phi_1 : \Delta_1 \rightarrow \Gamma$ and $\phi_2 : \Delta_2 \rightarrow \Gamma$ are called $C^1$ equivalent parametrizations of a singular cell $\Gamma$ if there is a $C^1$ orientation preserving homeomorphism $\psi : \Delta_1 \rightarrow \Delta_2$ such that $\phi_1 = \phi_2 \circ \psi$.

In that case it follows from Theorem 4.13 for the change of variables in integration and Theorem 5.5 for the composition of induced mappings of differential forms that

$$\int_{\Delta} \phi^*(\omega) = \int_{\Delta'} \psi^*(\phi^*(\omega)) = \int_{\Delta'} (\psi \circ \phi)^*(\omega);$$

thus the parametrizations $\phi$ and $\psi \circ \phi$ lead to the same value of the integral of the differential form $\omega$ over the singular cell $\Gamma$, so the dependence on the parametrization in a natural sense is independent of the mapping and the integral (5.65) generally will denoted merely by $\int_{\Gamma} \omega$. Note though that if the mapping $\psi$ is orientation reversing the integrals defined by the two mappings $\phi_1$ and $\phi_2$ are of opposite signs.

Line integrals in $\mathbb{R}^n$ are examples of integrals over singular cells, since a curve is a special case of a singular cell as already noted. If the curve $\gamma \subset U$ in an open subset $\mathbb{R}^n$ is the image of a $C^1$ mapping $\phi : [0, 1] \rightarrow U$ from an interval $[0, 1] \subset \mathbb{R}^1$, where the parameter in $[0, 1]$ is $t$ and the parameters in $\mathbb{R}^n$ are $x_1, \ldots, x_n$ with the orientation $dx_1 \wedge \cdots \wedge dx_n$ then to a vector field $f$ in $U$ there is associated the 1-form

$$\omega_f(x) = \sum_{j=1}^n f_j(x) dx_j.$$
5.3. **STOKES’S THEOREM**

The induced differential form under the mapping \( \phi \) is

\[
\phi^*(\omega_t)(t) = \sum_{j=1}^{n} f_j(\phi(t)) \frac{dx_j}{dt} dt,
\]

which can be written in terms of the vector field \( f(x) \) as the dot product

\[
(5.66) \quad \phi^*(\omega_t)(t) = f(\phi(t)) \cdot \phi'(t) dt;
\]

thus as in (5.11) the integral of the differential form \( \omega_t(x) \) over the curve \( \gamma \) is just the line integral

\[
(5.67) \quad \int_{\gamma} \phi^*(\omega_t) = \int_{\gamma} f \cdot \tau ds,
\]

which can be interpreted as the integral of the dot product of the vector field \( f \) and the unit tangent vector \( \tau \) to the curve \( \gamma \), integrated with respect to the arc length of the curve, to which all the results of Section 5.1 apply.

Another particularly useful special case is that of integration of differential 2-forms over 2-dimensional singular cells \( \mathbb{R}^3 \). If the singular 2-cell \( \Gamma \subset U \) of an open subset \( U \subset \mathbb{R}^3 \) is the image of a \( C^1 \) mapping \( \phi : \mathbb{S} \to U \) from a closed cell \( \mathbb{S} \subset \mathbb{R}^2 \) into \( U \), where the parameters in \( \mathbb{R}^2 \) are \( t_1, t_2 \) with the natural orientation \( dt_1 \wedge dt_2 \) and the parameters in \( \mathbb{R}^3 \) are \( x_1, x_2, x_3 \) with the natural orientation \( dx_1 \wedge dx_2 \wedge dx_3 \), then to any vector field \( f \) in \( U \) there can be associated the 2-form

\[
\Omega_f(x) = f_1(x)dx_2 \wedge dx_3 + f_2(x)dx_3 \wedge dx_1 + f_3(x)dx_1 \wedge dx_2
\]

with the classical cyclic notation. The induced differential form under the mapping \( \phi \) is the differential 2-form in \( \Delta \)

\[
\phi^*(\Omega_f)(t) = \sum_{j_1,j_2=1}^2 \left( f_1(\phi(t)) \frac{\partial x_2}{\partial t_{j_1}} \frac{\partial x_3}{\partial t_{j_2}} + f_2(\phi(t)) \frac{\partial x_3}{\partial t_{j_1}} \frac{\partial x_1}{\partial t_{j_2}} \right) dt_{j_1} \wedge dt_{j_2}
\]

\[
+ f_3(\phi(t)) \left( \frac{\partial x_1}{\partial t_{j_1}} \frac{\partial x_2}{\partial t_{j_2}} - \frac{\partial x_2}{\partial t_{j_1}} \frac{\partial x_3}{\partial t_{j_2}} + \frac{\partial x_3}{\partial t_{j_1}} \frac{\partial x_1}{\partial t_{j_2}} \right) dt_{j_1} \wedge dt_{j_2}
\]

\[
= f_1(\phi(t)) \left( \frac{\partial x_2}{\partial t_{j_1}} \frac{\partial x_3}{\partial t_{j_2}} - \frac{\partial x_3}{\partial t_{j_1}} \frac{\partial x_2}{\partial t_{j_2}} \right) dt_{j_1} \wedge dt_{j_2}.
\]

This can be written more succinctly as

\[
(5.68) \quad \phi^*(\Omega_f)(t) = \det \left( \begin{array}{ccc} f_1(\phi(t)) & f_2(\phi(t)) & f_3(\phi(t)) \\ \frac{\partial x_1}{\partial t_{j_1}} & \frac{\partial x_2}{\partial t_{j_1}} & \frac{\partial x_3}{\partial t_{j_1}} \\ \frac{\partial x_1}{\partial t_{j_2}} & \frac{\partial x_2}{\partial t_{j_2}} & \frac{\partial x_3}{\partial t_{j_2}} \end{array} \right) dt_{j_1} \wedge dt_{j_2},
\]

another example of the role of differential forms in handling determinants; and with the notation introduced in (5.31) this can in turn be written

\[
(5.69) \quad \phi^*(\Omega_f)(t) = f(\phi(t)) \cdot \left( \frac{\partial x}{\partial t_{j_1}} \wedge \frac{\partial x}{\partial t_{j_2}} \right) dt_{j_1} \wedge dt_{j_2},
\]
where the coefficient of this differential form thus is the dot product of the vector field \( f(\phi(t)) \) and the cross-product vector field \( \frac{\partial x}{\partial t_1} \times \frac{\partial x}{\partial t_2} \) where

\[
\frac{\partial x}{\partial t_j} = \begin{pmatrix} \frac{\partial x_1}{\partial t_j} \\ \frac{\partial x_2}{\partial t_j} \\ \frac{\partial x_3}{\partial t_j} \end{pmatrix} \quad \text{for} \quad j = 1, 2.
\]

In these terms then

\[
(5.70) \quad \int_{\Gamma} \Omega_t = \int_{\Delta} f(\phi(t)) \cdot \left( \frac{\partial x}{\partial t_1} \times \frac{\partial x}{\partial t_2} \right) dt_1 \wedge dt_2.
\]

The value of this integral depends on both the orientation of the space \( \mathbb{R}^3 \), which determines the order of the columns in (5.68), and the orientation of the space \( \mathbb{R}^2 \), which determines the order of the rows in (5.68). If the two vectors \( \frac{\partial x}{\partial t_1} \) and \( \frac{\partial x}{\partial t_2} \) are linearly independent at a point \( t \in \Delta \) then rank \( \phi'(t) = 2 \) and it follows from the rank theorem that the image of an open neighborhood of the point \( t \in \Delta \) is a 2-dimensional submanifold of \( U \), which of course may not be all of the image \( \Gamma \) near the point \( \phi(t) \in U \); the order of these two vectors describes an orientation of that submanifold, and the cross-product of the two vectors is a vector perpendicular to that submanifold oriented so that the three vectors

\[
\frac{\partial x}{\partial t_1}, \frac{\partial x}{\partial t_2}, \frac{\partial x}{\partial t_1} \times \frac{\partial x}{\partial t_2}
\]

have the chosen orientation of \( \mathbb{R}^3 \), the same orientation as the three vectors \( dx_1, dx_2, dx_3 \).

The unit vector in the direction of the cross-product \( \frac{\partial x}{\partial t_1} \times \frac{\partial x}{\partial t_2} \) is the **unit normal vector** to that submanifold, which is denoted by \( \nu(t) \); thus

\[
(5.71) \quad \frac{\partial x}{\partial t_1} \times \frac{\partial x}{\partial t_2} = \left\| \frac{\partial x}{\partial t_1} \times \frac{\partial x}{\partial t_2} \right\| \nu.
\]

In parallel with the treatment of line integrals in (5.67) the integral (5.70) can be written

\[
(5.72) \quad \int_{\Gamma} \Omega_t = \int_{\Gamma} f \cdot \nu \, dS
\]

and can be viewed as the integral over the surface \( \Gamma \) of the dot product of the vector field \( f \) and the unit normal vector \( \nu \) to \( \Gamma \), integrated with respect to the surface area of \( \Gamma \); thus the **surface area**\(^1\) of \( \Gamma \), which in parallel to the treatment of the arc length

\[
1\text{The direct definition of surface area is an extremely complicated matter, as discussed in detail for instance in the classical book by T. Radó \textit{Length and Area}, American Mathematical Society Colloquium Publications, New York, 1948. Those difficulties are finessed by using the pattern of the calculation of arc length through line integrals to define surface area, reversing the order of the discussion in Section 5.1. It is customary to attempt to justify this definition of surface area by observing that the length } \left\| \frac{\partial x}{\partial t_1} \times \frac{\partial x}{\partial t_2} \right\| \text{ of the cross product of the two vectors } \frac{\partial x}{\partial t_1} \text{ and } \frac{\partial x}{\partial t_2} \text{ is the area of the parallelogram spanned by these two vectors, which can be viewed as a planar approximation to the surface since the vectors are tangent vectors approximating the lengths of the two curves that are the image of the } t_1 \text{ and } t_2 \text{ coordinate axes under the mapping } \phi : \mathbb{R}^2 \rightarrow \mathbb{R}^3. \text{ What is perhaps more convincing though is that the integral } \int_{\Gamma} dS \text{ for some standard surfaces yields the usual values for surface areas.}
of curves is denoted by $|\Gamma|$, can be defined implicitly by

\begin{equation}
|\Gamma| = \int_{\Delta} \left\| \frac{\partial \mathbf{x}}{\partial t_1} \times \frac{\partial \mathbf{x}}{\partial t_2} \right\|.
\end{equation}

For example, if $\gamma \subset \mathbb{R}^3$ is the curve from the origin $\mathbf{a} = \{0, 0, 0\}$ to the point $\mathbf{b} = \{1, 1, 1\}$ that is the image of the mapping $\phi : [0, 1] \rightarrow \mathbb{R}^3$ where $\phi(t) = \{t, t^2, t^3\}$ and $\mathbf{f}$ is the vector field $\mathbf{f}(\mathbf{x}) = \{x_1^2, x_2, x_1x_2x_3\}$ then

\[
\int_{\gamma} \mathbf{f} \cdot \mathbf{\tau} \, ds = \int_{\gamma} (x_1^2 dx_1 + x_2 dx_2 + x_1x_2x_3 dx_3)
\]

\[
= \int_{0}^{1} (t^2 dt + t^2 \cdot 2tdt + t^6 \cdot 3t^2 dt)
\]

\[
= \int_{0}^{1} (3t^2 + 3t^4) dt = \frac{4}{3}.
\]

For another example, if $\Gamma \subset \mathbb{R}^3$ is the cylindrical surface $x_1^2 + x_2^2 = 1$, $0 \leq x_3 \leq 1$ oriented so that the unit normal vector to the surface points outwards, and $\mathbf{f}$ is the vector field $\mathbf{f}(\mathbf{x}) = \{x_1, x_2, x_3\}$, consider the integral $\int_{\Gamma} \mathbf{f} \cdot \mathbf{\nu} \, dS$. The singular cell $\Gamma$ can be described as the image of the mapping $\phi : \Delta \rightarrow \Gamma$ from the cell

\[
\Delta = \{ \mathbf{x} \in \mathbb{R}^3 \mid 0 \leq t_1 \leq 2\pi, 0 \leq t_1, t_2 \leq 1 \},
\]

with the orientation described by the differential form $dt_1 \wedge dt_2$, where

$\phi(t) = \{\cos t_1, \sin t_1, t_2\}$.

At the point $0 = \{0, 0, 1\} = \phi(0, 0)$ on the boundary of the surface $\Gamma$

\[
\frac{\partial \mathbf{x}}{\partial t_1} = \begin{bmatrix} -\sin t_1 & \cos t_1 & 0 \end{bmatrix} \bigg|_{t_1=t_2=0} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}
\]

and

\[
\frac{\partial \mathbf{x}}{\partial t_2} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}
\]

and consequently

\[
\frac{\partial \mathbf{x}}{\partial t_1} \times \frac{\partial \mathbf{x}}{\partial t_2} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix},
\]

which does point outward from the interior of the cylinder so the orientations chosen are the ones desired. (If the orientation had been the reverse of that desired, the appropriate orientation could be obtained by reversing the order of the parameters $t_1$ and $t_2$.) The 2-form associated to the vector field $\mathbf{f}$ is

\[
\Omega_\Gamma = x_2 \, dx_2 \wedge dx_3 + x_1 dx_3 \wedge dx_1 + x_3 dx_1 \wedge dx_2
\]

and consequently with the parametrization chosen

\[
\int_{\Gamma} \mathbf{f} \cdot \mathbf{u} \, dS = \int_{\Gamma} \Omega_\Gamma = \int_{\Delta} x_2 \, dx_2 \wedge dx_3 + x_1 dx_3 \wedge dx_1 + x_3 dx_1 \wedge dx_2
\]

\[
= \det \begin{bmatrix} \cos t_1 & \sin t_1 & t_2 \\ -\sin t_1 & \cos t_1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \, dt_1 \wedge dt_2
\]

\[
= \int_{t_1=0}^{2\pi} \int_{t_2=0}^{1} dt_2 dt_1 = 2\pi i.
\]
Furthermore since
\[ \frac{\partial \mathbf{x}}{\partial t_1} \times \frac{\partial \mathbf{x}}{\partial t_2} = \det \begin{pmatrix} dx_1 & dx_2 & dx_3 \\ -\sin t_1 & \cos t_1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \cos t_1 dx_1 + \sin t_1 dx_2 + 0 dx_3 \]
as a vector expressed in terms of the basis vectors \( dx_1, \ dx_2, \ dx_3 \) it follows that
\[ \left\| \frac{\partial \mathbf{x}}{\partial t_1} \times \frac{\partial \mathbf{x}}{\partial t_2} \right\| = 1. \]

Then the area of the cylinder \( \Gamma \) is
\[ \|\Gamma\| = \int_{\Gamma} 1dS = \int_{\Delta} 1 = 2\pi. \]
That is of course the known value, and it should be somewhat comforting that the general formula does work in this case.

To extend the notion of a singular cell, a singular \( m \)-chain in an open subset \( U \subset \mathbb{R}^n \) is defined to be a finite formal sum
\[ C = \sum_{j=1}^{r} \nu_j \cdot \Gamma_j \quad \text{for } \nu_j \in \mathbb{Z} \text{ and singular } m \text{-cells } \Gamma_j; \]
for some purposes it is also useful to consider a real singular \( m \)-chain, a formal sum (5.74) where \( \nu_j \in \mathbb{R} \). The integral of a differential \( m \)-form \( \omega \) over the singular chain (5.74) is defined by
\[ \int_{C} \omega = \sum_{j=1}^{r} \nu_j \int_{\Gamma_j} \omega = \sum_{j=1}^{r} \nu_j \int_{\Gamma_j} \omega \]
whether the coefficients \( \nu_j \) are integers or real numbers. For example, a 1-chain is just a finite formal sum of ordered curves, such as arises when a curve \( \gamma \) parametrized by a mapping \( \phi : [0, 1] \rightarrow \mathbb{R}^n \) is split into a collection of pieces by taking a partition of the interval \([0, 1]\); the restriction of the mapping \( \phi \) to each interval \( I_i \) in the partition is a parametrized curve \( \gamma_i \), and the initial curve \( \gamma \) can be viewed as the chain \( \gamma = \sum_i \gamma_i \). The integral over this chain is just the sum of the integrals over the individual curves forming it. The curve \( \gamma \) covered twice is the chain \( 2 \cdot \gamma \), so the integral over this chain is twice the integral over \( \gamma \); and the curve \( \gamma \) but with the reversed orientation is the chain \( -1 \cdot \gamma \); so the integral over this chain is the negative of the integral over \( \gamma \). A curve with finitely many singular points is often viewed as a chain consisting of the smooth individual pieces. Chains also appear naturally as boundaries of cells. The boundary of a cell
\[ \overline{\Delta}(x_1, x_2, \ldots, x_m) = \left\{ \mathbf{x} \in \mathbb{R}^n \mid 0 \leq x_i \leq 1 \right\} \]
in the vector space \( \mathbb{R}^m \), with coordinates \( x_1, \ldots, x_m \) and with the orientation described by the differential form \( dx_1 \wedge dx_2 \wedge \cdots \wedge dx_m \), when considered just as a point set consists of the \( 2m \) sets
\[ \Gamma(x_1, \ldots, x_{i-1}, \epsilon, x_{i+1}, \ldots, x_m) = \left\{ \mathbf{x} \in \overline{\Delta} \mid x_i = \epsilon \right\} \quad \text{for } 1 \leq i \leq m, \ \epsilon = \pm 1. \]
5.3. STOKES’S THEOREM

Each of these sets can be viewed as a singular \((m-1)\) cell defined by the obvious mapping

\[
\phi_i(\epsilon) : \Delta(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m) \rightarrow \Gamma_i(x_1, \ldots, x_{i-1}, \epsilon, x_{i+1}, \ldots, x_m)
\]

where the cell \(\Delta(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_m)\) has the orientation described by the differential \((m-1)\) form \(dx_1 \wedge \cdots \wedge dx_{i-1} \wedge dx_{i+1} \wedge \cdots \wedge dx_m\). The boundary chain of the \(m\)-cell \(\Delta\) is the singular chain formed by these boundary cells with the orientations given by

\[
\partial \Delta(x_1, x_2, \ldots, x_m) = \sum_{1 \leq i \leq m} \sum_{\epsilon = \pm 1} (-1)^{i+\epsilon} \Gamma(x_1, \ldots, x_{i-1}, \epsilon, x_{i+1}, \ldots, x_m).
\]

For example if \(m = 2\) the four sets forming the point set boundary of \(\Delta(x_1, x_2)\) are the singular 1-cells \(\Gamma(0, x_2), \Gamma(1, x_2), \Gamma(x_1, 0), \Gamma(x_1, 1)\), which are the images of the 1-cells \(\Delta(x_i)\) for \(i = 1, 2\) oriented as usual in the direction of increasing parameter values \(x_i\); the boundary chain is

\[
\partial \Delta(x_1, x_2) = \Gamma(x_1, 0) + \Gamma(1, x_2) - \Gamma(x_1, 1) - \Gamma(0, x_2),
\]

which is the customary counterclockwise orientation of the boundary, as is evident from the accompanying Figure 5.1. The corresponding example for \(m = 3\) is the

\[\text{Figure 5.1: The singular cells forming the boundary of a cell in } \mathbb{R}^2\]

singular 2-chain

\[
\partial \Delta(x_1, x_2, x_3) = -\Gamma(0, x_2, x_3) + \Gamma(1, x_2, x_3) + \Gamma(x_1, 0, x_3) - \Gamma(x_1, 1, x_3) - \Gamma(x_1, x_2, 0) + \Gamma(x_1, x_2, 1),
\]

which with a bit of sketching should appear to be the natural definition in that case as well. The boundary chain of a singular cell \(\Gamma\) represented as the image of a mapping \(\phi : \Delta(x_1, x_2, \ldots, x_m) \rightarrow \Gamma\) is defined to be the image of the boundary cell of \(\Delta(x_1, x_2, \ldots, x_m)\) under the mapping \(\phi\), so

\[
\partial \Gamma = \phi \left( \partial \Delta(x_1, x_2, \ldots, x_m) \right) = \sum_{1 \leq i \leq m} \sum_{\epsilon = \pm 1} (-1)^{i+\epsilon} \phi \left( \Gamma(x_1, \ldots, x_{i-1}, \epsilon, x_{i+1}, \ldots, x_m) \right),
\]

and the boundary cell of a singular chain \(C = \sum_i \nu_i \Gamma_i\) is defined by

\[
\partial C = \sum_i \nu_i \partial \Gamma_i.
\]
An $m$-chain $C$ is said to be a **cycle** if $\partial C = 0$, and it is said to be a **boundary** if $C = \partial \zeta_0$ for some $(m-1)$ chain $\zeta_0$.

**Theorem 5.10** The boundary of a cell is a cycle, or equivalently $\partial \partial \Delta = 0$ for any cell $\Delta \subset \mathbb{R}^m$.

**Proof:** For a cell $\Delta = \Delta(x_1, \ldots, x_m) \in \mathbb{R}^m$ with the usual orientation it follows from the definition (5.79) that

$$\partial \Delta = \sum_{i, \epsilon} (-1)^{i+\epsilon} \Gamma(x_1, \ldots, x_{i-1}, \epsilon, x_{i+1}, \ldots, x_m),$$

and consequently

$$\partial (\partial \Delta(x_1, \ldots, x_m)) = \sum_{i, \epsilon} (-1)^{i+\epsilon} \left( \sum_{j < i, \delta} (-1)^{j+\delta} \Delta(x_1, \ldots, x_{j-1}, \delta, \ldots, x_{i-1}, \epsilon, \ldots, x_m) 
+ \sum_{j > i, \delta} (-1)^{j+\delta-1} \Delta(x_1, \ldots, x_{i-1}, \epsilon, \ldots, x_{j-1}, \delta, \ldots, x_m) \right)$$

$$= \sum_{j < i, \delta, \epsilon} (-1)^{i+j+\epsilon+\delta} \Delta(x_1, \ldots, x_{j-1}, \delta, \ldots, x_{i-1}, \epsilon, \ldots, x_m)$$

$$- \sum_{i < j, \delta, \epsilon} (-1)^{i+j+\epsilon+\delta} \Delta(x_1, \ldots, x_{i-1}, \epsilon, \ldots, x_{j-1}, \delta, \ldots, x_m)$$

$$= 0$$

since the two terms of opposite sign are equal when the indices $i$ and $j$ and the indices $\epsilon$ and $\delta$ are interchanged. Then for a singular cell $\Gamma$ described by a mapping $\phi : \Delta(x_1, x_2, \ldots, x_m) \longrightarrow \Gamma$ it follows that

$$\partial \partial \Gamma = \phi (\partial \partial \Delta(x_1, x_2, \ldots, x_m))$$

$$= \phi (0) = 0,$$

which suffices for the proof.

The set of real singular $m$-chains in an open subset $U \subset \mathbb{R}^n$ form a real vector space $C_m(U, \mathbb{R})$, where the sum of two chains $C' = \sum_i \nu_i \Gamma_i'$ and $C'' = \sum_j \nu_j \Gamma_j''$ is the chain $C' + C'' = \sum_i \nu_i \Gamma_i' + \sum_j \nu_j \Gamma_j''$. It is clear from the definition of the boundary chain that $\partial (C' + C'') = \partial C' + \partial C''$, and consequently that the set of real singular $m$-cycles form a vector subspace $Z_m(U, \mathbb{R}) \subset C_m(U, \mathbb{R})$; and it follows from the preceding theorem that the set of real singular $m$-boundaries form a vector subspace $B_m(U, \mathbb{R}) \subset Z_m(U, \mathbb{R})$. The quotient vector space

$$(5.82) \quad H_m(U, \mathbb{R}) = \frac{Z_m(U, \mathbb{R})}{B_m(U, \mathbb{R})}$$

is called the **real singular homology group** of dimension $m$ of the set $U$. For ordinary singular chains the same construction can be carried out, but for abelian groups rather than for vector spaces and the corresponding quotient (5.82) is called the **singular homology group** of the subset $U$. These notions can be extended to
manifolds rather than just open subsets $U \subset \mathbb{R}^n$, and provide one of the basic algebraic
descriptions of the topological properties of manifolds.²

The fundamental theorem of calculus for functions of a single variable extends to
functions of several variables in the form of Stokes's Theorem; the extension appears
more complicated than the classical case because of the more complicated geometry
in higher dimensions, but the proof is a straightforward application of the classical
fundamental theorem in each variable separately. It really suffices to prove the result
in the following special case.

**Theorem 5.11** If $\omega$ is a $C^1$ differential form of degree $m - 1$ defined in an open
neighborhood of a cell $\Delta \subset \mathbb{R}^m$ then

$$\int_{\partial \Delta} \omega = \int_{\Delta} d\omega. \quad (5.83)$$

**Proof:** Consider in particular a differential form

$$\omega = f(x) dx_1 \wedge \cdots \wedge dx_{r-1} \wedge dx_{r+1} \wedge \cdots \wedge dx_m,$$

for which

$$d\omega = \partial_r f(x) dx_r \wedge dx_1 \wedge \cdots \wedge dx_{r-1} \wedge dx_{r+1} \wedge \cdots \wedge dx_m$$

$$= \partial_r f(x) (-1)^{r-1} dx_1 \wedge \cdots \wedge dx_m. \quad (5.84)$$

The differential form $\omega$ when restricted to the cells forming the boundary of $\Delta$ is
nontrivial only on the singular cells $\Gamma(x_1, \ldots, x_{r-1}, \epsilon, x_{r+1}, \ldots, x_m)$, for $\epsilon = 0$ or 1, since
for any other segment of the boundary one of the variables $x_i$ for $i \neq r$ is constant so
in the integration of $\omega$ the differential $dx_i$ vanishes identically; therefore

$$\int_{\partial \Delta} \omega = (-1)^r \int_{\Gamma(x_1, \ldots, x_{r-1}, 0, x_{r+1}, \ldots, x_m)} \omega' + (-1)^{r+1} \int_{\Gamma(x_1, \ldots, x_{r-1}, 1, x_{r+1}, \ldots, x_m)} \omega.$$

Both of the singular cells for this integration are parametrized by the cell $\Delta' =$
$\Delta(x_1, \ldots, x_{r-1}, x_{r+1}, \ldots, x_m)$, with the natural orientation, so the integrals of the
differential form $\omega$ over these cells by definition are just the integrals of the function
$f(x)$ over the parameter cell $\Delta'$, and consequently

$$\int_{\partial \Delta} \omega = \int_{\Delta'} \left( (-1)^r f(x_1, \ldots, x_{r-1}, 0, x_{r+1}, \ldots x_m) + (-1)^{r+1} f(x_1, \ldots, x_{r-1}, 1, x_{r+1}, \ldots x_m) \right).$$

By the fundamental theorem of calculus in a single variable the integrand in the
preceding integral is just $(-1)^{r-1} \int_0^1 \partial_r f(x)$; substituting this into the preceding integral
and using Fubini's Theorem shows that

$$\int_{\partial \Delta} \omega = (-1)^{r-1} \int_{[0,1]} \partial_r f(x) \left( \int_{\Delta'} \omega \right)$$

$$= (-1)^{r-1} \int_{\Delta} \partial_r f(x).$$

²Homology groups are discussed in detail in most books on algebraic topology, as for
instance in the book by Allen Hatcher *Algebraic Topology*, (Cambridge University Press,
2002).
However in view of (5.84)
\[ (-1)^{r-1} \int_{\Delta} \partial_r f(x) = \int_{\Delta} d\omega, \]
and that suffices for the proof.

**Corollary 5.3 (Stokes’s Theorem)** If \( \Gamma \subset U \) is a \( C^1 \) \( m \)-dimensional singular cell in an open subset \( U \subset \mathbb{R}^n \) and \( \omega \in \Lambda^m(U) \) is a differential \((m-1)\)-form in \( U \) then

\[ \int_{\partial \Gamma} \omega = \int_{\Gamma} d\omega. \]

**Proof:** The singular cell \( \Gamma \) is the image of a \( C^1 \) mapping \( \phi : \Delta \rightarrow \Gamma \) for some cell \( \Delta \subset \mathbb{R}^m \), and \( \phi \partial \Delta = \partial \Gamma \). Then for any differential \((m-1)\)-form \( \omega \) in \( U \) it follows from the preceding theorem, Theorem 5.8 (iii) and the definition of the integral over a singular cell that

\[ \int_{\partial \Gamma} \omega = \int_{\partial \Delta} \phi^* (\omega) = \int_{\Delta} d\phi^* (\omega) = \int_{\Gamma} \omega, \]

which suffices for the proof.

Examples of the application of Stokes’s Theorem abound, many expressed in standard classical forms in spaces of dimensions two and three. If \( f : U \rightarrow \mathbb{R}^2 \) is a \( C^1 \) vector field in an open subset \( U \subset \mathbb{R}^2 \) and \( \Gamma \subset \mathbb{R}^2 \) is a singular cell described by a \( C^1 \) parametrization \( \phi : \Delta \rightarrow \Gamma \) then Stokes’s Theorem asserts that

\[ \int_{\partial \Gamma} \omega = \int_{\Gamma} d\omega \]

for the 1-form \( \omega(x) = f_1(x) dx_1 + f_2(x) dx_2 \) associated to the vector field \( f \). The integral over the boundary of \( \Gamma \) is a line integral over that curve in \( \mathbb{R}^2 \); and the integral over \( \Gamma \) is the integral of the 2-form \( d\omega(x) = \left( \partial_1 f_2(x) - \partial_2 f_2(x) \right) dx_1 \wedge dx_2 \), which for the orientation of \( \mathbb{R}^2 \) determined by \( dx_1 \wedge dx_2 \) is just the integral of the coefficient of this 2-form, so Stokes’s Theorem takes the form

\[ \int_{\partial \Gamma} f_1(x) dx_1 + f_2(x) dx_2 = \int_{\Gamma} \left( \frac{\partial f_2(x)}{\partial x_1} - \frac{\partial f_1(x)}{\partial x_2} \right); \]

this is classically known as Green’s Theorem. This result can be used to calculate the area of a region from its boundary alone\(^3\); for if \( \gamma \) is the boundary chain of a singular cell \( D \subset \mathbb{R}^2 \) then for the vector field \( f(x) = \frac{1}{2} \{-x_2, x_1\} \)

\[ \int_{\gamma} f(x) \cdot \tau = \frac{1}{2} \int_{\gamma} (-x_2 dx_1 + x_2 dx_2) \]

\[ = \frac{1}{2} \int_D 2dx_1 \wedge dx_2 = |D|, \]

---

\(^3\)This method for calculating the area of a region from its boundary is embodied in mechanical and electronic machines, called planimeters, that have long been used in engineering and surveying; these machines are discussed for example in the book by John Bryant and Christopher J. Sangwin *How round is your circle? Where engineering and mathematics meet.* Princeton University Press, 2008.
5.3. STOKES’S THEOREM

the area of the region \( D \). In many applications Green’s formula is applied to singular chains rather than just to singular cells. For example the annular region

\[
\left\{ (x_1, x_2) \in \mathbb{R}^2 \mid 1 \leq x_1^2 + x_2^2 \leq 2 \right\}
\]

can be written as the singular chain \( \Gamma_1 + \Gamma_2 \) where \( \Gamma_1 \) is the left half of the annular region and \( \Gamma \) is the right half of the annular region, as in the accompanying Figure 5.2; the common boundary cells are the segments \( x_1 = 0, 1 \leq x_2 \leq 2 \) and \( x_1 = 0, -2 \leq x_2 \leq -1 \), which have different orientations for the two singular cells \( \Gamma_i \), so an integral over \( \partial(\Gamma_1 + \Gamma_2) \) reduces to the sum of an integral over the outer singular 1-cell

\[
\gamma_0 = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 2 \right\}
\]

with the counterclockwise orientation and an integral over the inner singular 1-cell

\[
\gamma_i = \left\{ (x_1, x_2) \in \mathbb{R}^2 \mid x_1^2 + x_2^2 = 1 \right\}
\]

with the clockwise orientation.

Figure 5.2: Boundary of an annular region

If \( f : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) is a \( C^1 \) vector field in an open subset \( U \subset \mathbb{R}^3 \) and \( \Gamma \subset \mathbb{R}^3 \) is a singular 3-cell described by a \( C^1 \) parametrization \( \phi : \Delta \rightarrow \Gamma \) then Stokes’s Theorem asserts that

\[
\int_{\partial\Gamma} \Omega_f = \int_{\Gamma} d\Omega_f
\]

for the 2-form \( \Omega_f(x) = f_1(x)dx_2 \wedge dx_3 + f_2(x)dx_3 \wedge dx_1 + f_3(x)dx_1 \wedge dx_2 \) associated to the vector field \( f \). The integral over the singular 2-chain \( \partial\Gamma \) is the classical integral \( \int_{\partial\Gamma} \mathbf{f} \cdot \nu \, dS \) where \( \nu \) is the unit normal to the surfaces forming the chain \( \partial\Gamma \) and \( dS \) is the element of surface area of \( \partial\Gamma \). The exterior derivative of the differential form \( \Omega_f \) is the differential form associated to the divergence \( \text{div} \mathbf{f} = \nabla \cdot \mathbf{f} \) of the vector field \( \mathbf{f} \) so its integral over \( \Gamma \) is just the integral of the divergence of \( \mathbf{f} \), hence Stokes’s Theorem takes the form

\[
\int_{\partial\Gamma} \mathbf{f}(x) \cdot \nu \, dS = \int_{\Gamma} \nabla \cdot \mathbf{f}(x);
\]

this is classically known as Gauss’s Theorem. The analogue of (5.87) in 3-dimensions is a calculation of the volume of a singular cell \( \Gamma \subset \mathbb{R}^3 \) by an integral over its boundary
alone; if
\[ f(x) = \frac{1}{3} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \]
then \( \nabla \cdot f(x) = 1 \) and it follows from Gauss’s Theorem that
\begin{align}
\int_{\partial \Gamma} f(x) \cdot \nu \ dS &= \int_{\Gamma} \nabla \cdot f(x) \\
&= \text{int}, 1 = |\Gamma|,
\end{align}
the volume of the cell \( \Gamma \). Again for many applications this is applied to singular chains rather than just to singular cells.

For yet another classical example, if \( f : U \rightarrow \mathbb{R}^3 \) is a \( C^1 \) vector field in an open subset \( U \subset \mathbb{R}^3 \) and \( \Gamma \subset \mathbb{R}^3 \) is a singular 2-cell described by a \( C^1 \) parametrization \( \phi : \Delta \rightarrow \Gamma \) then Stokes’s Theorem asserts that
\[ \int_{\partial \Gamma} \omega_f = \int_{\Gamma} d\omega_f \]
for the 1-form \( \omega_f(x) = f_1(x)dx_1 + f_2(x)dx_2 + f_3(x)dx_3 \) associated to the vector field \( f \). The integral of the 1-form over the curve \( \partial \Gamma \) is the standard line integral, and the integral of the differential 2-form \( d\omega_f \) over the surface \( \Gamma \) has the classical form
\[ \int_{\Gamma} f(x) \cdot \nu \ dS, \]
so Stokes’s Theorem takes the form
\[ \int_{\partial \Gamma} f(x) \cdot \nu \ dS = \int_{\Gamma} f(x) \cdot \nu \ dS; \]
this special case is classically known as Stokes’s Theorem, the title that in more recent times also has been attached to the generalization of this result to integrals over singular cells of arbitrary dimensions. Yet again for many applications this is applied to singular chains rather than just to singular cells.
5.3. STOKES’S THEOREM

Exterior Algebras
Appendix A

Exterior Algebra

Algebra is the study of a variety of structures: groups, rings, algebras, fields, and so on. A structure of considerable use in the calculus of several variables is exterior algebra, which is really a topic in linear algebra, or in multilinear algebra to be more precise, but which is not covered in most standard courses in linear algebra; so a short survey of the relevant material in an elementary form sufficient for purposes here is included in this appendix.

If $V$ is a real vector space of dimension $n$ with the basis $\xi_1, \ldots, \xi_n$, then to any integer $k$ in the range $1 \leq k \leq n$ there is associated another vector space $\Lambda^k(V)$, the basis of which consists of a set of vectors called basic $k$-vectors and denoted by

$$\xi_{i_1} \wedge \xi_{i_2} \wedge \cdots \wedge \xi_{i_k}$$

where $1 \leq i_1 < i_2 < \cdots < i_k \leq n$.

For $k = 1$ the basic 1-vectors thus are just the basis vectors $\xi_i$ themselves, so that $\Lambda^1(V) = V$ and consequently $\dim \Lambda^1(V) = 1$. For $k = n$ on the other hand there is just a single basic $n$-vector $\xi_1 \wedge \xi_2 \wedge \cdots \wedge \xi_n$, so $\dim \Lambda^n(V) = 1$. In general for $1 \leq k \leq n$ the basic $k$-vectors are described by the sets of $k$-tuples of integers from 1 to $n$ in ascending order, so by the number of distinct subsets of $k$ integers from the set of the first $n$ integers, and consequently $\dim \Lambda^k(V) = \binom{n}{k}$. It is customary to extend this set of vector spaces by setting $\Lambda^0(V) = \mathbb{R}$. It is also customary, and very convenient, to introduce as well the formal $k$-vectors $\xi_{i_1} \wedge \xi_{i_2} \wedge \cdots \wedge \xi_{i_k}$ for arbitrary sets of $k$ indices $i_j$ in the range $1 \leq i_j \leq n$, with the understanding that

$$\xi_{i_1} \wedge \xi_{i_2} \wedge \cdots \wedge \xi_{i_k} = 0$$

if the indices $i_1, i_2, \ldots, i_k$ are not distinct,

$$\xi_{i_1} \wedge \cdots \wedge \xi_{i_k} = (\text{sgn } \pi)\xi_{j_1} \wedge \cdots \wedge \xi_{j_k}$$

if $j_1, \ldots, j_k$ is a permutation $\pi$ of $i_1, \ldots, i_k$,

where $\text{sgn } \pi = \pm 1$ is the sign of the permutation $\pi$; if the permutation is described by a permutation matrix $M_\pi$, with a single entry of 1 in each row and column and entries of 0 elsewhere, then $\text{sgn } \pi = \det M_\pi$, or equivalently $\text{sgn } \pi$ is the number of transpositions of adjacent integers that transform the set $j_1, \ldots, j_k$ to the set $i_1, \ldots, i_k$. Any formal $k$-vector thus is equal to either a unique basic $k$-vector or the negative of a unique basic $k$-vector. A vector $\omega \in \Lambda^k(V)$ can be written as a unique linear combination

$$\omega = \sum_{1 \leq i_1 < i_2 < \cdots < i_k \leq n} a_{i_1i_2 \cdots i_k} \xi_{i_1} \wedge \xi_{i_2} \wedge \cdots \wedge \xi_{i_k}$$
for some real numbers \( a_{i_1i_2\ldots i_k} \), called the reduced form for the vector \( \omega \); alternatively the vector \( \omega \) can be written in various ways as a linear combination

\[
\omega = \sum_{i_1, i_2, \ldots, i_k = 1}^{n} b_{i_1i_2\ldots i_k} \xi_{i_1} \wedge \xi_{i_2} \wedge \cdots \wedge \xi_{i_k}
\]

for some real numbers \( a_{i_1i_2\ldots i_k} \), called unreduced forms for that vector. By using (A.2) and (A.3) an unreduced form for a vector \( \omega \) can be shown to be equal to a unique reduced form for that vector. For example if \( \dim V = 3 \) a basis for the 3-dimensional vector space \( \Lambda^2(V) \) consists of the basic 2-vectors \( \xi_1 \wedge \xi_2, \xi_1 \wedge \xi_3 \) and \( \xi_2 \wedge \xi_3 \); a vector \( \omega \) in the unreduced form

\[
\omega = b_{12} \xi_1 \wedge \xi_2 + b_{13} \xi_1 \wedge \xi_3 + b_{23} \xi_2 \wedge \xi_3
\]

\[
+ b_{21} \xi_2 \wedge \xi_1 + b_{31} \xi_3 \wedge \xi_1 + b_{32} \xi_3 \wedge \xi_2
\]

is equal to the reduced form

\[
\omega = (b_{12} - b_{23}) \xi_1 \wedge \xi_2 + (b_{13} - b_{32}) \xi_1 \wedge \xi_3 + (b_{23} - b_{32}) \xi_2 \wedge \xi_3.
\]

The addition of vectors can be carried out in either the reduced or unreduced form, whichever is more convenient.

There is an additional algebraic operation on vectors in the spaces \( \Lambda^k(V) \) that associates to an ordered pair of vectors \( \omega \in \Lambda^k(V) \) and \( \sigma \in \Lambda^l(V) \) a vector \( \omega \wedge \sigma \in \Lambda^{k+l}(V) \), the exterior product of these two vectors; explicitly if

\[
\omega = \sum_{i_1, \ldots, i_k = 1}^{n} a_{i_1, \ldots, i_k} \xi_{i_1} \wedge \cdots \wedge \xi_{i_k} \in \Lambda^k(V)
\]

and

\[
\sigma = \sum_{j_1, \ldots, j_l = 1}^{n} b_{j_1, \ldots, j_l} \xi_{j_1} \wedge \cdots \wedge \xi_{j_l} \in \Lambda^l(V)
\]

then

\[
\omega \wedge \sigma = \sum_{i_1, \ldots, i_k, j_1, \ldots, j_l = 1}^{n} a_{i_1, \ldots, i_k} b_{j_1, \ldots, j_l} \xi_{i_1} \wedge \cdots \wedge \xi_{i_k} \wedge \xi_{j_1} \wedge \cdots \wedge \xi_{j_l}.
\]

The unreduced form of vectors in \( \Lambda^k(V) \) is clearly very convenient for handling wedge products. The order of the vectors \( \omega \) and \( \sigma \) is important, since

\[
\omega \wedge \sigma = (-1)^{kl} \sigma \wedge \omega.
\]

To demonstrate that, when the vectors \( \omega \) and \( \sigma \) in (A.8) are reversed of course

\[
b_{j_1, \ldots, j_l} a_{i_1, \ldots, i_k} = a_{i_1, \ldots, i_k} b_{j_1, \ldots, j_l}
\]

since these are merely real numbers; however the permutation that exchanges

\[
\xi_{j_1} \wedge \cdots \wedge \xi_{j_l} \wedge \xi_{i_1} \wedge \cdots \wedge \xi_{i_k}
\]

and \( \xi_{i_1} \wedge \cdots \wedge \xi_{i_k} \wedge \xi_{j_1} \wedge \cdots \wedge \xi_{j_l} \)

can be written as the composition first of the permutations that transfers \( \xi_{i_1} \) across \( \xi_{j_1} \), a permutation with sign \((-1)^l\), followed by the permutation that transfers
for the matrix
\[ A \]
\[ (A.10) \]
\[ \omega \]
applications. If \( \dim V \) to the integers 1 \( \omega \) with the basis \( \xi \)
then as above \( \omega \) where \( \text{sgn} R \) while \( \dim \Lambda \)
then case in which \( V \) union \( \Lambda(1) \)
and \( \text{exterior algebra} \), or more precisely the exterior algebra over a real vector space of dimension \( n \).

The basic role of the exterior product is to handle a number of calculations involving determinants. For example if \( \omega_1 = \sum_{j=1}^{n} a_{1j} \xi_j \) and \( \omega_2 = \sum_{k=1}^{n} a_{2k} \) then
\[
\omega_1 \wedge \omega_2 = \sum_{j,k=1}^{n} a_{1j} a_{2k} \xi_j \wedge \xi_k = \sum_{1 \leq j < k \leq n} (a_{1j} a_{2k} - a_{1k} a_{2j}) \xi_j \wedge \xi_k
\]
so \( \omega_1 \wedge \omega_2 \) is the vector consisting of all the distinct determinants of 2 \( \times \) 2 submatrices of the 2 \( \times \) \( n \) matrix \( \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \end{pmatrix} \).

For example, if \( \omega_i \) are any \( n \) vectors in \( V \) written explicitly in terms of the basis vectors \( \xi_1, \ldots, \xi_n \) as
\[
\omega_i = \sum_{j=1}^{n} a_{ij} \xi_j \quad \text{for } 1 \leq i \leq n,
\]
then
\[
\omega_1 \wedge \cdots \wedge \omega_n = \sum_{j_1, \ldots, j_n=1}^{n} a_{1j_1} a_{2j_2} \cdots a_{nj_n} \xi_{j_1} \wedge \xi_{j_2} \wedge \cdots \wedge \xi_{j_n}
\]
\[
= \sum_{j_1, \ldots, j_n=1}^{n} a_{1j_1} a_{2j_2} \cdots a_{nj_n} \text{sgn} \begin{pmatrix} j_1 & j_2 & \cdots & j_n \\ 1 & 2 & \cdots & n \end{pmatrix} \xi_{j_1} \wedge \xi_{j_2} \wedge \cdots \wedge \xi_{j_n}
\]
where \( \text{sgn} \begin{pmatrix} j_1 & j_2 & \cdots & j_n \\ 1 & 2 & \cdots & n \end{pmatrix} \) is the sign of the permutation taking the integers \( j_1, j_2, \ldots, j_n \)
to the integers 1, 2, \ldots, \( n \), so as is familiar from linear algebra the preceding equation can be rewritten
\[ (A.10) \]
\[
\omega_1 \wedge \cdots \wedge \omega_n = \det A \cdot \xi_1 \wedge \xi_2 \wedge \cdots \wedge \xi_n
\]
for the matrix \( A = \{a_{ij}\} \).

The special cases of vector spaces \( V \) of dimensions 2 and 3 arise in a great many applications. If \( \dim V = 2 \) then \( \dim \Lambda^0(V) = 1 \) since \( \Lambda^0(V) = \mathbb{R} \) while \( \dim \Lambda^1(V) = 2 \) with the basis \( \xi_1, \xi_2 \) and \( \dim \Lambda^2(V) = 1 \) with the basis \( \xi_1 \wedge \xi_2 \), which is the only nontrivial wedge product. For any two vectors \( \omega = a_1 \xi_1 + a_2 \xi_2 \) and \( \sigma = b_1 \xi_1 + b_2 \xi_2 \) then as above \( \omega \wedge \sigma = \det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix} \). If \( \dim V = 3 \) then \( \dim \Lambda^0(V) = 1 \) since \( \Lambda^0(V) = \mathbb{R} \) while \( \dim \Lambda^1(V) = 3 \) with the basis \( \xi_1, \xi_2, \xi_3 \) and \( \dim \Lambda^2(V) = 3 \) with the basis \( \xi_1 \wedge \xi_2, \xi_1 \wedge \xi_3, \xi_2 \wedge \xi_3 \) and \( \dim \Lambda^3(V) = 1 \) with the basis \( \xi_1 \wedge \xi_2 \wedge \xi_3 \). What is
particularly interesting in this case is that both $\Lambda^1(V) = V$ and $\Lambda^2(V)$ are vector spaces of dimension 3, so an element of $\Lambda^2(V)$ also can be identified with a vector in the space $V$; it is traditional and natural to make this identification by writing a vector in $\Lambda^2(V)$ as

$$\omega = a_1 \xi_2 \wedge \xi_3 + a_2 \xi_3 \wedge \xi_1 + a_3 \xi_1 \wedge \xi_2,$$

(A.11)

the obvious cyclic order. Various wedge products then introduce various algebraic operations between pairs of vectors in $V$. Thus if $\omega = \sum_{j=1}^3 a_j \xi_j$ and $\sigma = \sum_{j=1}^3 b_k \xi_k$ are two vectors in $V$ their wedge product $\omega \wedge \sigma \in \Lambda^2(V)$ also can be viewed as a vector in $V$. Explicitly

$$\omega \wedge \sigma = \sum_{j,k=1}^3 a_j b_k \xi_j \wedge \xi_k$$

$$= (a_2 b_3 - a_3 b_2) \xi_2 \wedge \xi_3 + (a_3 b_1 - a_1 b_3) \xi_3 \wedge \xi_1 + (a_1 b_2 - a_2 b_1) \xi_1 \wedge \xi_2,$$

so with the identification as in (A.11) the element $\omega \wedge \sigma \in \Lambda^2(V)$ can be identified with the vector $\sum_{l=1}^3 e_l \xi_l$ where

$$e_1 = \det \begin{pmatrix} a_2 & a_3 \\ b_2 & b_3 \end{pmatrix}, e_2 = \det \begin{pmatrix} a_3 & a_1 \\ b_3 & b_1 \end{pmatrix}, e_3 = \det \begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix};$$

(A.12)

alternatively $\omega \wedge \sigma$ viewed as a vector in $\Lambda^1(V)$ can be calculated by formally evaluating the determinant

$$\omega \times \sigma = \det \begin{pmatrix} \xi_1 & \xi_2 & \xi_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{pmatrix}.$$  

(A.13)

The classical notation for the vector associated in this way with the wedge product $\omega \wedge \sigma$ is $\omega \times \sigma$, called the cross product of these two vectors. This algebraic operation appears rather mysteriously in many discussions of calculus in several variables, but is quite natural as an operation in exterior algebra for vector spaces of dimension 3. Another classical algebraic operation is another manifestation of the wedge product. If $\omega = \sum_{j=1}^3 a_j \xi_j \in \Lambda^1(V)$ and if $\sigma = b_1 \xi_2 \wedge \xi_3 + b_2 \xi_3 \wedge \xi_1 + b_3 \xi_1 \wedge \xi_2 \in \Lambda^2(V)$ written as in (A.11) then

$$\omega \wedge \sigma = (a_1 b_1 + a_2 b_2 + a_3 b_3) \xi_1 \wedge \xi_2 \wedge \xi_3;$$

(A.14)

so when this exterior product is identified with a real number the result is just the dot product of the two vectors $(a_1, a_2, a_3)$ and $(b_1, b_2, b_3)$. By the way, it is evident from (A.9) that the cross-product is a skew-symmetric pairing while the dot product is a symmetric pairing. It is an amusing exercise to develop the corresponding algebraic operations on 4-dimensional vector spaces.
Index

algebra
  exterior, 108, 109
almost everywhere, 57
angle, between vectors, 4
arc length, 78
  parametrization by, 79

basic \( k \)-vectors, 107
boundary, 5–7, 100
boundary chain, 99
Cartesian norm, 2
Casorati-Weierstrass Theorem, 9
Cauchy-Schwarz inequality, 4
cell, 5
cell, singular, 93
chain
  boundary, 99
  singular, 98
change of local coordinates, 32
characteristic function of a set, 60
closed set, 7
closure
  of a set, 7
compact set, 8
complete, 6
connected set, 7
conservative vector field, 81
content
  of set, 60
content zero, 56
continuous, 10
  uniformly, 57
  uniformly, 12
continuously differentiable, 16
convergent
  uniformly, 67
coordinate functions
  of a mapping, 10
coordinates
  local, change of, 32
coordinats
  local, 32
critical point, 23
cross-product
  of vector fields in \( \mathbb{R}^3 \), 86
curve, 77
  \( C^r \), 77
curve, parametrized, 46
cycle, 100
degree
  of a differential form, 83
dependent
  functionally, 49
derivative, 14
  directional, 23
  exterior, 101
  partial, 14
derivative, exterior, 88
derived set, 7
diffeomorphism, 29
differentiable, 13
  continuously, 16
differential form, 27, 83
  induced under a mapping, 87
  reduced, 83
  reduced and unreduced forms, 84
  unreduced, 83
differential forms
  exterior products, 84
  linear combinations, 84
directional derivative, 23
domain, Jordan, 59
dot product, 2
edgesize
  of cell, 8
equivalent
  as manifolds, 49
norms, 4
equivalent parametrizations, $C^1$, 77
Euclidean norm, 2
extended Jordan domain, 69
exterior algebra, 108, 109
exterior derivative, 88, 101
exterior product, 108
of differential forms, 84
field, vector, 27
formal $k$-vectors, 107
Fubini’s Theorem, 63
function
induced, 77
induced under a mapping, 86
functionally dependent, 49
Gauss’s Theorem, 104
gradient, 23
greatest lower bound, 6
Green’s Theorem, 102
Heine-Borel Theorem, 9
Hessian, 22, 24
homeomorphic submanifolds, 39
homeomorphism, 29
improper integral, 69
induced differential form
under a mapping, 87
induced function, 77
under a mapping, 86
infimum, 6
inner product, 2
integrable, 58, 61
integral
improper, 69
iterated, 63
lower, 55
oriented, 72, 92
Riemann, 55
upper, 55
integral, line, 81
interval, 6
Inverse Mapping Theorem, 31
iterated integral, 63
Jacobian, 29
Jacobian matrix, 29
Jordan domain, 59
extended, 69
$k$-vectors
basic, 107
formal, 107
Lagrange multipliers, 47
least upper bound, 6
length, arc, 78
limit point, 7
line integral, 81
linear combination
of differential forms, 84
local
change of coordinates, 32
coordinate system, 32
locus, zero, 38
lower integral, 55
lower sum, 54
manifold
continuous, 49
mapping, 10
open, 32
matrix
Jacobian, 29
maximum
local, 23
Mean Value Inequality, 25
Mean Value Theorem, 25
measure zero, 56
minimum
local, 23
multi-index notation, 84
norm, 2
equivalent, 4
notation
multi-index, 84
open covering, 8
open mapping, 32
open neighborhood, 6
open set, 6
orientation, 33
oriented integral, 72, 92
orthogonal, 4
oscillation, 55
parametrization by arc length, 79
INDEX

parametrizations, $C^1$ equivalent, 77
parametrized curve, 46
partial derivative, 14
partition, 53
point
  singular, 33
polarization identity, 3
potential, 81
pullback
  of differential form, 87

Rank Theorem, 43
rectifiable curve, 78
reduced
  form of differential form, 84
reduced form, 83
  of vector in $\Lambda^k(V)$, 108
refinement, 54
relative topology, 7
Riemann integral, 55

separated sets, 7
singular cell, 93
singular chain, 98
singular point, 33
singularities
  of submanifolds, 40
Stokes’s Theorem, 101, 104
subcovering, 8
submanifold, 38
  homeomorphic, 39
  weakly homeomorphic, 49
sum
  lower, 54
  upper, 53
supremum, 6

tangent line, 46
tangent space, 46
tangent vector, 46
Taylor expansion, 20, 21
Theorem of Lebesgue, 57
topology
  relative, 7
topology, on a set, 7

uniformly continuous, 12
uniformly convergent, 67
unit tangent vector, 81
unreduced

form of differential form, 84
unreduced form, 83
  of vector in $\Lambda^k(V)$, 108
upper integral, 55
upper sum, 53

vector
  tangent, 46
  unit tangent, 81
vector field, 27
  conservative, 81
weakly homeomorphic submanifolds, 49

zero locus, 38

index of notation
(x, y), inner product, 2
$E'$, derived set or set of limit points of $E$, 7
$S^*(f, P)$, upper sum of function $f$ for partition $P$, 53
$S_r(f, P)$, upper sum of function $f$ for partition $P$, 54
$\Delta f(x)$, change in function $f(x)$, 24
$\Delta$, cell, 5
$\Delta(a, b)$ unoriented interval between $a$ and $b$, 72
$\Lambda(V)$, exterior algebra on vector space $V$, 108
$\Lambda^k(V)$, space spanned by the basic $k$-vectors in $V$, 107
$\Lambda^0_{(k)}(U)$, vector space of $C^k$ functions in $U$, 77
$\tau$, unit tangent vector to a curve, 81
$x \cdot y$, dot product, 2
$\chi_D(x)$, characteristic function of $D$, 60
e-neighborhood, 5
inf $S$, infimum or greatest lower bound of set $S$, 6
$\int_0^1 f(s) \, ds$, integral with respect to arc length, 79
$\int_{\Gamma} f$, Riemann integral, 55
$\nabla f(x)$, gradient of function $f(x)$, 23
$\omega \wedge \sigma$, exterior product of differential forms, 85
$E$, closure of set $E$, 7
$\partial S$, boundary of set $S$, 5, 7
$\partial_a f(a)$, directional derivative of function $f$, 23
sup $S$, supremum or least upper bound of set $S$, 6

$f_n \Rightarrow^U f$, $f_n$ converge uniformly to $f$ in $U$, 67

$f_n \to f$, $f_n$ converge to $f$, 67

$o_f(S)$, oscillation of function $f$ in a set $S$, 55

$o_f(a)$, oscillation of function $f$ at a point $a$, 55

$s$, arc length as parameter of a rectifiable curve, 79

$\text{grad } f(x)$, gradient of function $f(x)$, 23

$C^\infty$ homeomorphism, 29

$C^1$, class of continuously differentiable functions, 16

$N_\epsilon(a)$, $\epsilon$ neighborhood of $a$, 5

$\mathcal{Z}(f)$, induced function under the mapping $\phi$, 86

$\mathcal{Z}(f\omega)$, induced differential $r$-form under the mapping $\phi$, 87