Foundations of Analysis

Joseph L. Taylor

Version 2.3, Spring 2010

ii

Contents

Preface

This text evolved from notes we developed for use in a two semester undergraduate course on foundations of analysis at the University of Utah. The course is designed for students who have completed three semesters of calculus and one semester of linear algebra. For most of them, this is the first mathematics course in which everything is proved rigorously and they are expected to not only understand proofs, but to also create proofs.

The course has two main goals. The first is to develop in students the mathematical maturity and sophistication they will need when they move on to senior or graduate level mathematics courses. The second is to present a rigorous development of the calculus, beginning with a study of the properties of the real number system.

We have tried to present this material in a fashion which is both rigorous and concise, with simple, straightforward explanations. We feel that the modern tendency to expand textbooks with ever more material, excessively verbose explanations, and more and more bells and whistles, simply gets in the way of the student's understanding of the material.

The exercises differ widely in level of abstraction and level of difficulty. They vary from the simple to the quite difficult and from the computational to the theoretical. There are exercises that ask students to prove something or to construct an example with certain properties. There are exercises that ask students to apply theoretical material to help do a computation or to solve a practical problem. Each section contains a number of examples designed to illustrate the material of the section and to teach students how to approach the exercises for that section.

This text, in its various incarnations, has been used by the author and his colleagues for several years at the University of Utah. Each use has led to improvements, additions, and corrections.

The topics covered in the text are quite standard. Chapters 1 through 6 focus on single variable calculus and are normally covered in the first semester of the course. Chapters 7 through 11 are concerned with calculus in several variables and are normally covered in the second semester.

Chapter 1 begins with a section on set theory. This is followed by the introduction of the set of natural numbers as a set which satisfies Peano's axioms. Subsequent sections outline the construction, beginning with the natural numbers, of the integers, the rational numbers, and finally the real numbers. This

is only an outline of the construction of the reals beginning with Peano's axioms and not a fully detailed development. Such a development would would be much too time consuming for a course of this nature. What is important is that, by the end of the chapter: (1) students know that the real number system is a complete, Archimedean, ordered field; (2) they have some practice at using the axioms satisfied by such a system; and (3) they understand that this system may be constructed beginning with Peano's axioms for the counting numbers.

Chapter 2 is devoted to sequences and limits of sequences. We feel sequences provide the best context in which to first carry out a rigorous study of limits. The study of limits of functions is complicated by issues concerning the domain of the function. Furthermore, one has to struggle with the student's tendency to think that the limit of $f(x)$ as x approaches a is just a pedantic way of describing $f(a)$. These complications don't arise in the study of limits of sequences.

Chapter 3 provides a rigorous study of continuity for real valued functions of one variable. This includes proving the existence of minimum and maximum values for a continuous function on a closed bounded interval as well as the Intermediate Value Theorem and the existence of a continuous inverse function for a strictly monotone continuous function. Uniform continuity is discussed as is uniform convergence for a sequence of functions.

The derivative is introduce in Chapter 4 and the main theorems concerning the derivative are proved. These include the Chain Rule, the Mean Value Theorem, existence of the derivative of an inverse function, the monotonicity theorem, and L'Hôpital's Rule.

In Chapter 5 the definite integral is defined using upper and lower Riemann sums. The main properties of the integral are proved here along with the two forms of the Fundamental Theorem of Calculus. The integral is used to define and develop the properties of the natural logarithm. This leads to the the definition of the exponential function and the development of its properties.

Infinite sequences and series are discussed in Chapter 6 along with Taylor's Series and Taylor's Formula.

The second half of the text begins in Chapter 7 with an introduction to d-dimensional Euclidean space, $\mathbb{R}^d$, as the vector space of d-tuples of real numbers. We review the properties of this vector space while reminding the students of the definition and properties of general vector spaces. We study convergence of sequences of vectors and prove the Bolzano-Weierstrass Theorem in this context. We describe open and closed sets and discuss compactness and connectedness of sets in Euclidean spaces. Throughout this chapter and subsequent chapters we follow a certain philosophy concerning abstract verses concrete concepts. We briefly introduce abstract metric spaces, inner product spaces, and normed linear spaces, but only as an aside. We emphasize that Euclidean space is the object of study in this text, but we do point out now and then when a theorem concerning Euclidean space does or does not hold in a general metric space or inner product space or normed vector space. That is, the course is grounded in the concrete world of $\mathbb{R}^d$, but the student is made aware that there are more exotic worlds in which these concepts are important.

Chapter 8 is devoted to the study of continuous functions between Euclidean

spaces. We study the basic properties of continuous functions as they relate to open and closed sets and compact and connected sets. The third section is devoted to sequences and series of function and the concept of uniform convergence. The last two sections comprise a review of the topic of linear functions between Euclidean spaces and the corresponding matrices. This includes the study of rank, dimension of image and kernel and invertible matrices. We also introduce representations of linear or affine subspaces in parametric form as well as solution sets of systems of equations.

The most important topic in the second half of the course is probably the study, in Chapter 9, of the total differential of a function from $\mathbb{R}^p$ to $\mathbb{R}^q$. This is introduced in the context of affine approximation of a function near a point in its domain. The chain rule for the total differential is proved in what we believe is a novel and intuitively satisfying way. This is followed by applications of the total differential and the chain rule, including the multivariable Taylor's formula and the inverse and implicit function theorems.

Chapter 10 is devoted to integration over Jordan regions in $\mathbb{R}^d$. The development, using upper and lower sums, is very similar to the development of the single variable integral in Chapter 5. Where the proofs are virtually identical to those in Chapter 5, they are omitted. The really new and different material here is that on Fubini's Theorem and the change of variables formula. We give rigorous and detailed proofs of both results along with a number of applications.

The chapter on vector calculus, Chapter 11 uses the modern formalism of differential forms. In this formalism, the major theorems of the subject – Green's Theorem, Stoke's Theorem, and Gauss' s Theorem – all have the same form. We do point out the classical forms of each of these theorems, however. Each of the main theorems is proved first on a rectangle or cube and then extended to more complicated domains through the use of transformation laws for differential forms and the change of variables formula for multiple integrals. Most of the chapter focuses on integration over sets in $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$ which can be parameterized by smooth maps from an interval, a square or a cube, or sets which can be partitioned into sets of this form. However, in an optional section at the end, we introduce integrals over p-chains and p-cycles and state the general form of Stoke's Theorem

There are topics which could have been included in this text, but were not. Some of our colleagues suggested that we include an introductory chapter or section on formal logic. We considered this but decided against it. Our feeling is that logic at this simple level is just language used with precision. Students have been using language for most of their lives, perhaps not always with precision, but that doesn't mean that they are incapable of using it with precision if required to do so. Teaching students to be precise in their use of the language tools that they already possess is one of the main objectives of the course. We do not believe that beginning the course with a study of formal logic would be of much help in this regard and, in fact, might just get in the way.

We could also have included a chapter of Fourier Series. However, we fell that the material that has been included makes for a text that is already a challenge to cover in a two semester course. We feel it unrealistic to think that

an additional chapter at the end would often get covered. In any case, the study of Fourier series is most naturally introduced at the undergraduate level in a course in differential equations.

Chapter 1

The Real Numbers

This course has two goals: (1) to develop the foundations that underlie calculus and all of post calculus mathematics, and (2) to develop students' ability to understand definitions and proofs and to create proofs of their own – that is, to develop students' *mathematical sophistication*.

The typical freshman and sophomore calculus courses are designed to teach the techniques needed to solve problems using calculus. They are not primarily concerned with proving that these techniques work or teaching why they work. The key theorems of calculus are not really proved, although sometimes proofs are given which rely on other reasonable, but unproved assumptions. Here we will give rigorous proofs of the main theorems of calculus. To do this requires a solid understanding of the real number system and its properties. This first chapter is devoted to developing such an understanding.

Our study of the real number system will follow the historical development of numbers: We first discuss the natural numbers or counting numbers (the positive integers), then the integers, followed by the rational numbers. Finally, we discuss the real number system and the property that sets it apart from the rational number system – the completeness property. The completeness property is the missing ingredient in most calculus courses. It is seldom discussed, but without it, one cannot prove the main theorems of calculus.

The natural numbers can be defined as a set satisfying a very simple list of axioms – Peano's axioms. All of the properties of the natural numbers can be proved using these axioms. Once this is done, the integers, the rational numbers, and the real numbers can be constructed and their properties proved rigorously. To actually carry this out would make for an interesting, but rather tedious course. Fortunately, that is not the purpose of this course. We will not give a rigorous construction of the real number system beginning with Peano's axioms, although we will give a brief outline of how this is done. However, the main purpose of this chapter is to state the properties that characterize the real number system and develop some facility at using them in proofs. The rest of the course will be devoted to using these properties to develop rigorous proofs of the main theorems of calculus.

1.1   Sets and Functions

We precede our study of the real numbers with a brief introduction to sets and functions and their properties. This will give us the opportunity to introduce the set theory notation and terminology that will be used throughout the text.

Sets

A *set* is a collection of objects. These objects are called the *elements* of the set. If x is an element of the set A, then we will also say that x *belongs* to A or x is *in* A. A shorthand notation for this statement that we will use extensively is

$$x \in A.$$

Two sets A and B are the same set if they have the same elements – that is, if every element of A is also an element of B and every element of B is also an element of A. In this case, we write $A = B$.

One way to define a set is to simply list its elements. For example, the statement

$$A = \{1, 2, 3, 4\}$$

defines a set A which has as elements the integers from 1 to 4.

Another way to define a set is to begin with a known set A and define a new set B to be all elements $x \in A$ that satisfy a certain condition $Q(x)$. The condition $Q(x)$ is a statement about the element x which may be true for some values of x and false for others. We will denote the set defined by this condition as follows:

$$B = \{x \in A : Q(x)\}.$$

This is mathematical shorthand for the statement "B is the set of all x in A such that $Q(x)$". For example, if A is the set of all students in this class, then we might define a set B to be the set of all students in this class who are sophomores. In this case, $Q(x)$ is the statement "x is a sophomore". The set B is then defined by

$$B = \{x \in A : x \text{ is a sophomore}\}.$$

Example 1.1.1. Describe the set $(0, 3)$ of all real numbers greater than 0 and less than 3 using set notation.

Solution: In this case the statement $Q(x)$ is the statement "$0 < x < 3$". Thus,

$$(0, 3) = \{x \in \mathbb{R} : 0 < x < 3\}.$$

A set B is a *subset* of a set A if B consists of some of the elements of A – that is, if each element of B is also an element of A. In this case, we use the shorthand notation

$$B \subset A.$$

Of course, A is a subset of itself. We say B is a *proper* subset of A if $B \subset A$ and $B \neq A$.
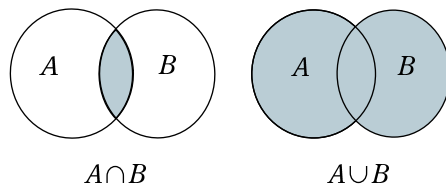
Figure 1.1: Intersection and Union of Two Sets.

For example, the open interval $(0, 3)$ of the preceding example is a proper subset of the set $\mathbb{R}$ of real numbers . It is also a proper subset of the half open interval $(0, 3]$ – that is, $(0, 3) \subset (0, 3]$, but the two are not equal because the second contains 3 and the first does not.

There is one special set that is a subset of every set. This is the empty set $\emptyset$. It is the set with no elements. Since it has no elements, the statement that "each of its elements is also an element of A" is true no matter what the set A is. Thus, by the definition of subset,

$$\emptyset \subset A$$

for every set A.

If A and B are sets, then the *intersection* of A and B, denoted $A \cap B$, is the set of all objects that are elements of A and of B. That is,

$$A \cap B = \{x : x \in A \text{ and } x \in B\}.$$

Similarly, the *union* of A and B, denoted $A \cup B$, is the set of objects which are elements of A or elements of B (possibly elements of both). That is,

$$A \cup B = \{x : x \in A \text{ or } x \in B\}.$$

Example 1.1.2. If A is the closed interval $[-1, 3]$ and B is the open interval $(1, 5)$, describe $A \cap B$ and $A \cup B$.

Solution: $A \cap B = (1, 3]$ and $A \cup B = [-1, 5)$.

If $\mathcal{A}$ is a (possibly infinite) collection of sets, then the intersection and union of the sets in $\mathcal{A}$ are defined to be

$$\bigcap \mathcal{A} = \{x : x \in A \text{ for all } A \in \mathcal{A}\}$$

and

$$\bigcup \mathcal{A} = \{x : x \in A \text{ for some } A \in \mathcal{A}\}.$$

Note how crucial the distinction between "for all' and "for some' is in these definitions.

The intersection $\bigcap \mathcal{A}$ is also often denoted

$$\bigcap_{A \in \mathcal{A}} A \quad \text{or} \quad \bigcap_{s \in S} A_s$$

if the sets in $\mathcal{A}$ are indexed by some index set S. Similar notation is often used for the union.

Example 1.1.3. If $\mathcal{A}$ is the collection of all intervals of the form $[s, 2]$ where $0 < s < 1$, find $\bigcap \mathcal{A}$ and $\bigcup \mathcal{A}$.

   Solution: A number x is in the set

$$\bigcap \mathcal{A} = \bigcap_{s \in (0,1)} [s, 2]$$

if and only if

$$s \leq x \leq 2 \quad \text{for every positive } s < 1. \tag{1.1.1}$$

Clearly every x in the interval $[1, 2]$ satisfies this condition. We will show that no points outside this interval satisfy (1.1.1).

   Certainly an $x > 2$ does not satisfy (1.1.1). If $x < 1$, then $s = x/2 + 1/2$ (the midpoint between x and 1) is a number less than 1 but greater than x, and so such an x also fails to satisfy (1.1.1). This proves that

$$\bigcap \mathcal{A} = [1, 2].$$

   A number x is in the set

$$\bigcup \mathcal{A} = \bigcup_{s \in (0,1)} [s, 2]$$

if and only if

$$s \leq x \leq 2 \quad \text{for some positive } s < 1. \tag{1.1.2}$$

Every such x is in the interval $(0, 2]$. Conversely, we will show that every x in this interval satisfies (1.1.2). In fact, if $x \in [1, 2]$, then x satisfies (1.1.2) for every $s < 1$. If $x \in (0, 1)$, then x satisfies 1.1.2 for $s = x/2$. This proves that

$$\bigcup \mathcal{A} = (0, 2].$$

   If $B \subset A$, then the set of all elements of A which are not elements of B is called the *complement* of B in A. This is denoted $A \setminus B$. Thus,

$$A \setminus B = \{x \in A : x \notin B\}.$$

Here, of course, the notation $x \notin B$ is shorthand for the statement "x is not an element of B".

   If all the sets in a given discussion are understood to be subsets of a given *universal* set X, then we may use the notation B^c for $X \setminus B$ and call it simply the *complement* of B. This will often be the case in this course, with the universal set being the set $\mathbb{R}$ of real numbers or, in later chapters, real n dimensional space $\mathbb{R}^n$ for some n.

Example 1.1.4. If A is the interval $[-2, 2]$ and B is the interval $[0, 1]$, describe $A \setminus B$ and the complement B^c of B in $\mathbb{R}$.

Solution: We have

$$A \setminus B = [-2, 0) \cup (1, 2] = \{x \in \mathbb{R} : -2 \le x < 0 \text{ or } 1 < x \le 2\},$$

while

$$B^c = (-\infty, 0) \cup (1, \infty) = \{x \in \mathbb{R} : x < 0 \text{ or } 1 < x\}.$$

Theorem 1.1.5. *If A and B are subsets of a set X and A^c and B^c are their complements in X, then*

(a) $(A \cup B)^c = A^c \cap B^c$; and

(b) $(A \cap B)^c = A^c \cup B^c$.

Proof. We prove (a) first. To show that two sets are equal, we must show that they have the same elements. An element of X belongs to $(A \cup B)^c$ if and only if it is not in $A \cup B$. This is true if and only if it is not in A and it is not in B. By definition this is true if and only if $x \in A^c \cap B^c$. Thus, $(A \cup B)^c$ and $A^c \cap B^c$ have the same elements and, hence, are the same set.

If we apply part (a) with A and B replaced by A^c and B^c and use the fact that $(A^c)^c = A$ and $(B^c)^c = B$, the result is

$$(A^c \cup B^c)^c = A \cap B.$$

Part (b) then follows if we take the complement of both sides of this identity. $\square$

A statement analogous to Theorem 1.1.5 is true for unions and intersections of collections of sets (Exercise 1.1.7).

Two sets A and B are said to be *disjoint* if $A \cap B = \emptyset$. That is, they are disjoint if they have no elements in common. A collection $\mathcal{A}$ of sets is called a *pairwise disjoint* collection if $A \cap B = \emptyset$ for each pair A, B of distinct sets in $\mathcal{A}$.

Functions

A *function* f from a set A to a set B is a rule which assigns to each element $x \in A$ exactly one element $f(x) \in B$. The element $f(x)$ is called the image of x under f or the value of f at x. We will write

$$f : A \to B$$

to indicate that f is a function from A to B. The set A is called the *domain* of f. If E is any subset of A then we write

$$f(E) = \{f(x) : x \in E\}$$

and call $f(E)$ the *image* of E under f.

We don't assume that every element of B is the image of some element of A. The set of elements of B which are images of elements of A is $f(A)$ and is

called the *range* of f. If every element of B is the image of some element of A (so that the range of f is B), then we say that f is *onto*.

A function $f : A \to B$ is is said to be *one-to-one* if, whenever $x, y \in A$ and $x \neq y$, then $f(x) \neq f(y)$ – that is, if f takes distinct points to distinct points.

If $g : A \to B$ and $f : B \to C$ are functions, then there is a function $f \circ g : A \to C$, called the *composition* of f and g, defined by

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

Since $g(x) \in B$ and the domain of f is B, this definition makes sense.

If $f : A \to B$ is a function and $E \subset B$, then the *inverse image* of E under f is the set

$$f^{-1}(E) = \{x \in A : f(x) \in E\}.$$

That is, $f^{-1}(E)$ is the set of all elements of A whose images under f belong to E.

Inverse image behaves very well with respect to the set theory operations, as the following theorem shows.

Theorem 1.1.6. *If $f : A \to B$ is a function and E and F are subsets of B, then*

(a) $f^{-1}(E \cup F) = f^{-1}(E) \cup f^{-1}(F)$;

(b) $f^{-1}(E \cap F) = f^{-1}(E) \cap f^{-1}(F)$; *and*

(c) $f^{-1}(E \setminus F) = f^{-1}(E) \setminus f^{-1}(F)$ *if $F \subset E$.*

Proof. We will prove (a) and leave the other two parts to the exercises.

To prove (a), we will show that $f^{-1}(E \cup F)$ and $f^{-1}(E) \cup f^{-1}(F)$ have the same elements. If $x \in f^{-1}(E \cup F)$, then $f(x) \in E \cup F$. This means that $f(x)$ is in E or in F. If it is in E, then $x \in f^{-1}(E)$. If it is in F, then $x \in f^{-1}(F)$. In either case, $x \in f^{-1}(E) \cup f^{-1}(F)$. This proves that every element of $f^{-1}(E \cup F)$ is an element of $f^{-1}(E) \cup f^{-1}(F)$.

On the other hand, if $x \in f^{-1}(E) \cup f^{-1}(F)$, then $x \in f^{-1}(E)$, in which case $f(x) \in E$, or $x \in f^{-1}(F)$, in which case $f(x) \in F$. In either case, $f(x) \in E \cup F$, which implies $x \in f^{-1}(E \cup F)$. This proves that every element of $f^{-1}(E) \cup f^{-1}(F)$ is also an element of $f^{-1}(E \cup F)$. Combined with the previous paragraph, this proves that the two sets are equal. □

Image does not behave as well as inverse image with respect the set operations. The best we can say is the following:

Theorem 1.1.7. *If $f : A \to B$ is a function and E and F are subsets of A, then*

(a) $f(E \cup F) = f(E) \cup f(F)$;

(b) $f(E \cap F) \subset f(E) \cap f(F)$;

(c) $f(E) \setminus f(F) \subset f(E \setminus F)$ *if $F \subset E$.*

Proof. We will prove (c) and leave the others to the exercises.

To prove (c), we must show that each element of $f(E) \setminus f(F)$ is also an element of $f(E \setminus F)$. If $y \in f(E) \setminus f(F)$, then $y = f(x)$ for some $x \in E$ and y is not the image of any element of F. In particular, $x \notin F$. This means that $x \in E \setminus F$ and so $y \in f(E \setminus F)$. This completes the proof. $\square$

The above theorem cannot be improved. That is, it is not in general true that $f(E \cap F) = f(E) \cap f(F)$ or that $f(E) \setminus f(F) = f(E \setminus F)$ if $F \subset E$. The first of these facts is shown in the next example. The second is left to the exercises.

Example 1.1.8. Give an example of a function $f : A \to B$ for which there are subsets $E, F \subset A$ with $f(E \cap F) \neq f(E) \cap f(F)$.

Solution: Let A and B both be $\mathbb{R}$ and let $f : A \to B$ be defined by

$$f(x) = x^2.$$

If $E = (0, \infty)$ and $F = (-\infty, 0)$, then $E \cap F = \emptyset$, and so $f(E \cap F)$ is also the empty set. However, $f(E) = f(F) = (0, \infty)$, and so $f(E) \cap f(F) = (0, \infty)$ as well. Clearly $f(E \cap F)$ and $f(E) \cap f(F)$ are not the same in this case.

Cartesian Product

If A and B are sets, then their *Cartesian product* $A \times B$ is the set of all ordered pairs (a, b) with $a \in A$ and $b \in B$. Similarly, the Cartesian product of n sets $A_1, A_2, \cdots, A_n$ is the set $A_1 \times A_2 \times \cdots A_n$ of all ordered n-tuples $(a_1, a_2, \cdots, a_n)$ with $a_i \in A_i$ for $i = 1, \cdots, n$.

If $f : A \to B$ is function from a set A to a set B, then the *graph* of f is the subset of $A \times B$ defined by $\{(a, b) \in A \times B : b = f(a)\}$.

Exercise Set 1.1

1. If $a, b \in \mathbb{R}$ and $a < b$, give a description in set theory notation for each of the intervals $(a, b), [a, b], [a, b)$, and $(a, b]$ (see Example 1.1.1).

2. If A, B, and C are sets, prove that

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C).$$

3. If A and B are two sets, then prove that A is the union of a disjoint pair of sets, one of which is contained in B and one of which is disjoint from B.

4. What is the intersection of all the open intervals containing the closed interval $[0, 1]$? Justify your answer.

5. What is the intersection of all the closed intervals containing the open interval $(0, 1)$? Justify your answer.

6. What is the union of all of the closed intervals contained in the open interval $(0, 1)$? Justify your answer.

7. If $\mathcal{A}$ is a collection of subsets of a set X, formulate and prove a theorem like Theorem 1.1.5 for the intersection and union of $\mathcal{A}$.

8. Which of the following functions $f : \mathbb{R} \to \mathbb{R}$ are one to one and which ones are onto. Justify your answer.

   (a) $f(x) = x^2$;
   (b) $f(x) = x^3$;
   (c) $f(x) = e^x$.

9. Prove Part (b) of Theorem 1.1.6.

10. Prove Part (c) of Theorem 1.1.6.

11. Prove Part (a) of Theorem 1.1.7.

12. Prove Part (b) of Theorem 1.1.7.

13. Give an example of a function $f : A \to B$ and subsets $F \subset E$ of A for which $f(E) \setminus f(F) \neq f(E \setminus F)$.

14. Prove that equality holds in Parts (b) and (c) of Theorem 1.1.7 if the function f is one-to-one.

15. Prove that if $f : A \to B$ is a function which is one-to-one and onto, then f has an *inverse function* – that is, there is a function $g : B \to A$ such that $g(f(x)) = x$ for all $x \in A$ and $f(g(y)) = y$ for all $y \in B$.

16. Prove that a subset G of $A \times B$ is the graph of a function from A to B if and only if the following condition is satisfied: for each $a \in A$ there is exactly one $b \in B$ such that $(a, b) \in G$.

1.2   The Natural Numbers

The natural numbers are the numbers we use for counting, and so, naturally, they are also called the *counting numbers*. They are the positive integers $1, 2, 3, \cdots$.

   The requirements for a system of numbers we can use for counting are very simple. There should be a first number (the number 1), and for each number there must always be a next number (a successor). After all, we don't want to run out of numbers when counting a large set of objects. This line of thought leads to Peano's axioms which characterize the system of natural numbers $\mathbb{N}$:

N1. there is an element $1 \in \mathbb{N}$;

N2. for each $n \in \mathbb{N}$ there is a successor element $s(n) \in \mathbb{N}$;

N3. 1 is not the successor of any element of $\mathbb{N}$;

N4. if two elements of $\mathbb{N}$ have the same successor, then they are equal;

N5. if a subset A of $\mathbb{N}$ contains 1 and is closed under succession (meaning $s(n) \in A$ whenever $n \in A$), then $A = \mathbb{N}$.

Note: at this stage in the development of the natural number system, all we have are Peano's axioms; addition has not yet been defined. When we define addition in $\mathbb{N}$, $S(n)$ will turn out to be $n + 1$.

Everything we need to know about the natural numbers can be deduced from these axioms. That is, using only Peano's axioms, one can define addition and multiplication of natural numbers and prove that they have the usual arithmetic properties. One can also define the order relation on the natural numbers and prove that it has the appropriate properties. To do all of this is not difficult, but it is tedious and time consuming. We will do some of this here in the text and the exercises, but we won't do it all. We will do enough so that students should understand how such a development would proceed. Then we will state and discuss the important properties of the resulting system of natural numbers.

Our main tool in this section will be *mathematical induction*, a powerful technique that is a direct consequence of Axiom **N5**.

Induction

Axiom **N5** above is often called the *induction* axiom, since it is the basis for mathematical induction. Mathematical induction is used in making definitions that involve a sequence of objects to be defined and in proving propositions that involve a sequence of statements to be proved. Here, by a *sequence* we mean a function whose domain is the natural numbers. Thus, a sequence of statements is an assignment of a statement to each $n \in \mathbb{N}$. For example, "n is either 1 or it is the successor of some element of $\mathbb{N}$" is a sequence of statements, one for each $n \in \mathbb{N}$. We will use induction to prove that all of these statements are true once we prove the following theorem.

The following theorem states the mathematical induction principle as it applies to proving propositions.

Theorem 1.2.1. *Suppose $\{P_n\}$ is a sequence of statements, one for each $n \in \mathbb{N}$. These statements are all true provided*

1. *P_1 is true (the base case is true); and*

2. *whenever P_n is true for some $n \in \mathbb{N}$, then $P_{s(n)}$ is also true (the induction step can be carried out).*

Proof. Let A be the subset of $\mathbb{N}$ consisting of those n for which P_n is true. Then hypothesis (1) of the theorem implies that $1 \in A$, while hypothesis (2) implies that $s(n) \in A$ whenever $n \in A$. By Axiom **N5**, $A = \mathbb{N}$, and so P_n is true for every n. $\square$

Example 1.2.2. Prove that each $n \in \mathbb{N}$ is either 1 or is the successor of some element of N.

Solution: If n is 1 then the statement is obviously true. Thus, the base case is true. If the statement is true of n then it is certainly true of $s(n)$, because it is true of any element which is the successor of something in $\mathbb{N}$. Thus, by induction, the statement is true for every $n \in \mathbb{N}$.

Another way to say what was proved in this example is that every natural number except 1 has a predecessor. This statement doesn't seem obvious at this stage of development of $\mathbb{N}$, but its proof was a rather trivial application of induction.

Inductive Definitions

Inductive definitions are used to define sequences. The sequence $\{x_n\}$ to be defined is a sequence of elements of some set X, which may or may not be a set of numbers. We wish to define the sequence in such a way that x_1 is a specified element of X and, for each $n \in \mathbb{N}$, $x_{s(n)}$ is a certain function of x_n. That is, we are given an element $x_1 \in X$ and a sequence of functions $f_n : X \to X$ and we wish to construct a sequence $\{x_n\}$, beginning with x_1, such that

$$x_{s(n)} = f_n(x_n) \quad \text{for all} \quad n \in \mathbb{N}. \tag{1.2.1}$$

This equation, defining $x_{s(n)}$ in terms of x_n, is called a *recursion relation*. Sequences defined in this way occur very often in mathematics. Newton's method from calculus and Euler's method for numerically solving differential equations are two important examples.

Theorem 1.2.3. *Given a set X, an element $x_1 \in X$, and a sequence $\{f_n\}$ of functions from X to X, there is a unique sequence $\{x_n\}$ in X, beginning with x_1, which satisfies $x_{s(n)} = f_n(x_n)$ for all $n \in \mathbb{N}$.*

Proof. Consider the Cartesian product $\mathbb{N} \times X$ – that is, the set of all ordered pairs (n, x) with $n \in \mathbb{N}$ and $x \in X$. We define a function $S : \mathbb{N} \times X \to \mathbb{N} \times X$ by

$$S(n, x) = (s(n), f_n(x)) \tag{1.2.2}$$

We say that a subset E of $\mathbb{N} \times X$ is closed under S if S sends elements of E to elements of E. Clearly the intersection of all subsets of $\mathbb{N} \times X$ that are closed under S and contain $(1, x_1)$ is also closed under S and contains $(1, x_1)$. This is the smallest subset of $\mathbb{N} \times X$, that is closed under S and contains $(1, x_1)$. We will call this set A.

To complete the argument, we will show that the set A is the graph of a function from $\mathbb{N}$ to X – that is, it has the form $\{(n, x_n) : n \in \mathbb{N}\}$ for a certain sequence $\{x_n\}$ in X. This is the sequence we are seeking. To prove A is the graph of a function from $\mathbb{N}$ to X we must show that each $n \in \mathbb{N}$ is the first element of exactly one pair $(n, x) \in A$. We prove this by induction.

The element 1 is the first element of the pair $(1, x_1)$, which is in A by the construction of A. If there were another element $x \in X$ such that $(1, x) \in A$, then we could remove $(1, x)$ from A and have a smaller set containing $(1, x_1)$ and closed under S. This is due to the fact that $(1, x)$ cannot be in the image of S, since 1 is not the successor of any element of $\mathbb{N}$ by **N3**.

Now, for the induction step, suppose for some n we know that there is a unique element $x_n \in X$ such that $(n, x_n) \in A$. Then $S(n, x_n) = (s(n), f_n(x_n))$ is in A. Suppose there is another element $(s(n), x) \in A$ with $x \neq f_n(x_n)$ and suppose this element is in the image of S – that is $(s(n), x) = S(m, y) = (s(m), f_m(y))$ for some $(m, y) \in A$. Then $n = m$ by **N4**, and $y = x_n$ by the induction assumption. Thus if $(s(n), x)$ is really different from $(s(n), f_n(x_n))$, then it cannot be in the image of S. As before this means we can remove it from A and still have a set closed under S and containing $(1, x_1)$. Since A is the smallest such set, we conclude there is no such element $(s(n), x)$. By induction, for each element of $\mathbb{N}$ there is a unique element $x_n \in X$ such that $(n, x_n) \in A$. Thus, A is the graph of a function $n \to x_n$ from $\mathbb{N}$ to X.

This shows the existence of a sequence with the required properties. We leave the proof that this sequence is unique to the exercises. $\square$

Note that the proof of the above theorem used all of Peano's axioms, not just **N5**.

Using Peano's Axioms to Develop Properties of $\mathbb{N}$

In this subsection, we will demonstrate some of the steps involved in developing the arithmetic and order properties of $\mathbb{N}$ using only Peano's axioms. It is not a complete development, but just a taste of what is involved. We begin with the definition of addition.

Definition 1.2.4. We fix $m \in \mathbb{N}$ and define a sequence $\{m+n\}_{n \in \mathbb{N}}$ inductively as follows:

$$
\begin{aligned}
m + 1 &= s(m), \quad \text{and} \\
m + s(n) &= s(m + n).
\end{aligned} \tag{1.2.3}
$$

These two conditions determine a unique sequence $\{m + n\}_{n \in \mathbb{N}}$ by Theorem 1.2.3.

By the above definition, the successor $s(n)$ of n is our newly defined $n+1$. At this point we will begin using $n + 1$ in place of $s(n)$ in our inductive arguments and definitions.

Example 1.2.5. Using the above definition and Peano's axioms, prove the associative law for addition in $\mathbb{N}$. That is, prove

$$m + (n + k) = (m + n) + k \quad \text{for all} \quad k, n, m \in \mathbb{N}.$$

Solution: We fix m and n and, for each $k \in \mathbb{N}$, let P_k be the proposition $m + (n + k) = (m + n) + k$. We prove that P_k is true for all $k \in \mathbb{N}$ by induction on k.

The base case P_1 is just

$$m + (n + 1) = (m + n) + 1. \qquad (1.2.4)$$

which is the recursion relation (1.2.3) used in the definition of addition once we replace $s(n)$ with $n + 1$. Thus, P_1 is true by definition.

For the induction step, we assume P_k is true for some k – that is, we assume

$$m + (n + k) = (m + n) + k.$$

We then take the successor of both sides of this equation to obtain

$$(m + (n + k)) + 1 = ((m + n) + k) + 1.$$

If we use (1.2.4) on both sides of this equation, the result is

$$m + ((n + k) + 1) = (m + n) + (k + 1).$$

Using (1.2.4) again, this time on the left side of the equation, leads to

$$m + (n + (k + 1)) = (m + n) + (k + 1).$$

Since this is proposition P_{k+1}, the induction is complete.

Example 1.2.6. Using Definition 1.2.4 and Peano's axioms, prove that $1 + n = n + 1$ for every $n \in \mathbb{N}$.

Solution: Let P_n be the statement $1 + n = n + 1$. We prove by induction that P_n is true for every n. It is trivially true in the base case $n = 1$, since P_1 just says $1 + 1 = 1 + 1$.

For the induction step, we assume that P_n is true for some n – that is we assume $1 + n = n + 1$. If we add 1 to both sides of this equation (i.e. take the successor of both sides), we have

$$(1 + n) + 1 = (n + 1) + 1.$$

By Definition 1.2.4, the left side of this equation is equal to $1 + (n + 1)$. Thus,

$$1 + (n + 1) = (n + 1) + 1.$$

Thus, P_{n+1} is true if P_n is true and the induction is complete.

A similar induction, this time on m, with n fixed can be used to prove the commutative law of addition – that is, $m + n = n + m$ for all $n, m \in \mathbb{N}$. The base case for this induction is the statement proved above. The associative law proved in Example 1.2.5 is needed in the proof of the induction step. We leave the details to the exercises.

We leave the definition of multiplication in $\mathbb{N}$ to the exercises. Its definition and the fact that it also satisfies the associative and commutative laws follows a pattern similar to the one above for addition. Once multiplication is defined, we can define factors and prime numbers:

Definition 1.2.7. If a number $n \in \mathbb{N}$ can be written as $n = mk$ with both $m \in \mathbb{N}$ and $k \in \mathbb{N}$, then k and m are called *factors* of n and are said to *divide* n. If $n \neq 1$ and the only factors of n are 1 and n, then n is said to be *prime*.

The order relation in $\mathbb{N}$ can be defined as follows:

Definition 1.2.8. If $n, m \in \mathbb{N}$, we will say that n is less than m, denoted $n < m$, if there is a $k \in \mathbb{N}$ such that $m = n + k$. We say n is less than or equal to m and write $n \leq m$ if $n < m$ or $n = m$.

Some of the properties of this order relation are worked out in the exercises. One of these is that each factor of n is necessarily less than or equal to n (Exercise 1.2.7).

Example 1.2.9. Prove that each natural number $n > 1$ is a product of primes.
  Solution: Here we understand that a prime number itself is a product of primes – a product with only one factor. Note that if k and m are two numbers which are products of primes, then their product km is also a product of primes.
  Let the proposition P_n be that every $m \in \mathbb{N}$, with $1 < m \leq n$, is a product of primes.
  Base case: P_1 is true because there is no $m \in \mathbb{N}$ with $1 < m \leq 1$.
  Induction step: suppose n is a natural number for which P_n is true. Then each m with $1 < m \leq n$ is a product of primes . Now $n + 1 > 1$ and so it is either a prime, or it factors as a product km with k and m not equal to 1 or $n + 1$. In the first case, P_{n+1} is true. In the second case, both k and m are less than $n + 1$ and, hence, less than or equal to n. Since P_n is true, k and m are products of primes. This implies that $n + 1 = km$ is also a product of primes and, in turn, this implies that P_{n+1} is true.
  By induction, P_n is true for all $n \in \mathbb{N}$ and this means that every natural number $n > 1$ is a product of primes.

Additional Examples of the Use of Induction

At this point we leave the discussion of Peano's axioms and the development of the properties of the natural numbers. The remainder of the section is devoted to further examples of inductive proofs and inductive definitions. Some of these involve the real number system, which won't be discussed until Section 1.4. Never-the-less we are happy to anticipate its development and use its properties in these examples.

Example 1.2.10. Prove by induction that every number of the form $5^n - 2^n$, with $n \in \mathbb{N}$ is divisible by 3.
  Solution: The proposition P_n is that $5^n - 2^n$ is divisible by 3.
  Base case: Since $5 - 2 = 3$, P_1 is true;
  Induction step: We need to show that P_{n+1} is true whenever P_n is true. We do this by rewriting the expression $5^{n+1} - 2^{n+1}$ as

$$5^{n+1} - 5 \cdot 2^n + 5 \cdot 2^n - 2^{n+1} = 5(5^n - 2^n) + (5 - 2)2^n.$$

If P_n is true then the first term on the right is divisible by 3. The second term on the right is also divisible by 3, since $5 - 2 = 3$. This implies that $5^{n+1} - 2^{n+1}$ is divisible by 3 and, hence, that P_{n+1} is true. This completes the induction step.

By induction (that is, by Theorem 1.2.1), P_n is true for all n.

Example 1.2.11. Define a sequence $\{x_n\}$ of real numbers by setting $x_1 = 1$ and using the recursion relation

$$x_{n+1} = \sqrt{x_n + 1}. \tag{1.2.5}$$

Show that this is an increasing sequence of positive numbers less than 2.

Solution: The function $f(x) = \sqrt{x + 1}$ may be regarded as a function from the set of positive real numbers into itself. We can apply Theorem 1.2.3, with each of the functions f_n equal to f, to conclude that a sequence $\{x_n\}$ is uniquely defined by setting $x_1 = 1$ and imposing the recursion relation (1.2.5).

Let P_n be the proposition that $x_n < x_{n+1} < 2$. We will prove that P_n is true for all n by induction.

Base Case: P_1 is the statement $x_1 < x_2 < 2$. Since $x_1 = 1$ and $x_2 = \sqrt{2}$, this is true.

Induction Step: Suppose P_n is true for some n. Then $x_n < x_{n+1} < 2$. If we add one and take the square root, this becomes

$$\sqrt{x_n + 1} < \sqrt{x_{n+1} + 1} < \sqrt{3}.$$

Using the recursion relation (1.2.5), this yields

$$x_{n+1} < x_{n+2} < \sqrt{3}$$

Since $\sqrt{3} < 2$, P_{n+1} is true. This completes the induction step.

We conclude that P_n is true for all $n \in \mathbb{N}$.

Binomial Formula

The proof of the binomial formula is an excellent example of the use of induction.

We will use the notation

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}.$$

This is the number of ways of choosing k objects from a set of n objects.

Theorem 1.2.12. *If x and y are real numbers and $n \in \mathbb{N}$, then*

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

Proof. We prove this by induction on n.

Base Case: Since $\binom{1}{0}$ and $\binom{1}{1}$ are both 1, the binomial formula is true when $n = 1$.

Induction Step: If we assume the formula is true for a certain n, then multiplying both sides of this formula by $x + y$ yields

$$
\begin{aligned}
(x+y)^{n+1} &= x \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} + y \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} x^{k+1} y^{n-k} + \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k+1}.
\end{aligned}
\tag{1.2.6}
$$

If we change variables in the first sum on the second line of (1.2.6) by replacing k by $k - 1$, then our expression for $(x + y)^{n+1}$ becomes

$$
\begin{aligned}
x^{n+1} &+ \sum_{k=1}^{n} \binom{n}{k-1} x^k y^{n-k+1} + \sum_{k=1}^{n} \binom{n}{k} x^k y^{n-k+1} + y^{n+1} \\
&= x^{n+1} + \sum_{k=1}^{n} \left[\binom{n}{k-1} + \binom{n}{k} \right] x^k y^{n+1-k} + y^{n+1}.
\end{aligned}
\tag{1.2.7}
$$

If we use the identity (to be proved in Exercise 1.4.17)

$$
\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k},
$$

then the right side of equation (1.2.7) becomes

$$
x^{n+1} + \sum_{k=1}^{n} \binom{n+1}{k} x^k y^{n+1-k} + y^{n+1} = \sum_{k=0}^{n+1} \binom{n+1}{k} x^k y^{n+1-k}.
$$

Thus, the binomial formula is true for $n + 1$ if it is true for n. This completes the induction step and the proof of the theorem. $\square$

Exercise Set 1.2

In the first seven exercises use only Peano's axioms and results that were proved in Section 1.2 using only Peano's axioms.

1. Prove the commutative law for addition, $n + m = m + n$, holds in $\mathbb{N}$. Use induction and Examples 1.2.6 and 1.2.5.

2. Prove that if $n, m \in \mathbb{N}$, then $m + n \neq n$. Hint: use induction on n.

3. Use the preceding exercise to prove that if $n, m \in \mathbb{N}$, $n \leq m$, and $m \leq n$ then $n = m$.

4. Prove that the order relation on $\mathbb{N}$ has the transitive property: if $k < n$ and $n < m$, then $k < m$.

5. Use the preceding exercise and Peano's axioms to prove that if $n \in \mathbb{N}$, then for each element $m \in \mathbb{N}$ either $m \le n$ or $n \le m$. Hint: use induction on n.

6. Show how to define the product nm of two natural numbers. Hint: use induction on m.

7. Use the definition of product you gave in the preceding exercise to prove that if $n, m \in \mathbb{N}$ then $n \le nm$.

For the remaining exercises you are no longer restricted to just using Peano's axioms and their immediate consequences.

8. Using induction, prove that $n^2 + 3n + 3$ is odd for every $n \in \mathbb{N}$;

9. Using induction, prove that $7^n - 2^n$ is divisible by 5 for every $n \in \mathbb{N}$.

10. Using induction, prove that $\sum_{k=1}^{n} k = \dfrac{n(n+1)}{2}$ for every $n \in \mathbb{N}$.

11. Using induction, prove that $\sum_{k=1}^{n} (2k - 1) = n^2$ for every $n \in \mathbb{N}$.

12. Finish the prove of Theorem 1.2.3 by showing that there is only one sequence $\{x_n\}$ which satisfies the conditions of the theorem.

13. Let a sequence $\{x_n\}$ of numbers be defined recursively by

$$x_1 = 0 \quad \text{and} \quad x_{n+1} = \frac{x_n + 1}{2}.$$

Prove by induction that $x_n \le x_{n+1}$ for all $n \in \mathbb{N}$. Would this conclusion change if we set $x_1 = 2$?

14. Let a sequence $\{x_n\}$ of numbers be defined recursively by

$$x_1 = 1 \quad \text{and} \quad x_{n+1} = \frac{1}{1 + x_n}.$$

Prove by induction that x_{n+2} is between x_n and x_{n+1} for each $n \in \mathbb{N}$.

15. Mathematical induction also works for a sequence $P_k, P_{k+1}, \cdots$ of propositions, indexed by the integers $n \ge k$ for some $k \in \mathbb{N}$. The statement is: If P_k is true and P_{n+1} true whenever P_n is true and $n \ge k$, then P_n is true for all $n \ge k$. Prove this.

16. Use induction in the form stated in the preceding exercise to prove that $n^2 < 2^n$ for all $n \ge 5$.

17. Prove the identity

$$\binom{n}{k-1} + \binom{n}{k} = \binom{n+1}{k},$$

which was used in the proof of Theorem 1.2.12.

18. Write out the binomial formula in the case $n = 4$.

19. Prove the well ordering principal for the natural numbers: each non-empty subset S of $\mathbb{N}$ contains a smallest element. Hint: apply the induction axiom to the set
$$T = \{n \in \mathbb{N} : n < m \text{ for all } m \in S\}.$$

20. Use the result of Exercise 1.2.19 to prove the division algorithm: If n and m are natural numbers with $m < n$, and if m does not divide n, then there are natural numbers q and r such that $n = qm + r$ and $r < m$. Hint: consider the set S of all natural numbers s such that $(s+1)m > n$.

1.3 Integers and Rational Numbers

The need for larger number systems than the natural numbers became apparent early in mathematical history. We need the number 0 in order to describe the number of elements in the empty set. The negative numbers are needed to describe deficits. Also, the operation of subtraction leads to non-positive integers unless $n - m$ is to be defined only for $m < n$.

Beginning with the system of natural numbers $\mathbb{N}$ and its properties derivable from Peano's axioms, the system of integers $\mathbb{Z}$ can easily be constructed. One simply adjoins to $\mathbb{N}$ a new element called 0 and, for each $n \in \mathbb{N}$ a new element called $-n$. Of course, one then has to define addition and multiplication and an order relation "$\leq$" for this new set $\mathbb{Z}$ in a way that is consistent with the existing definitions of these things for $\mathbb{N}$. When addition and multiplication are defined, we want them to have the properties that $0 + n = n$, and $n + (-n) = 0$. It turns out that these requirements and the commutative, associative and distributive laws (described below) are enough to uniquely determine how addition and multiplication are defined in $\mathbb{Z}$.

When all of this has been carried out, the new set of numbers $\mathbb{Z}$ can be shown to be a *commutative ring*, meaning that it satisfies the axioms listed below.

The Commutative Ring of Integers

A binary operation on a set A is rule which assigns to each ordered pair (a, b) of elements of A a third element of A.

Definition 1.3.1. A *commutative ring* is set R with two binary operations, addition $((a, b) \to a + b)$ and multiplication $((a, b) \to ab)$, that satisfy the following axioms:

A1. (Commutative Law of Addition) $x + y = y + x$ for all $x, y \in R$;

A2. (Associative Law of Addition) $x + (y + z) = (x + y) + z$ for all $x, y, z \in R$;

A3. (Additive Identity) there is an element $0 \in R$ such that $0 + x = x$ for all $x \in R$;

A4. (Additive Inverses) for each $x \in R$, there is an element $-x$ such that $x + (-x) = 0$;

M1. (Commutative Law of Multiplication) $xy = yx$ for all $x, y \in R$;

M2. (Associative Law of Multiplication) $x(yz) = (xy)z$ for all $x, y, z \in R$;

M3. (Multiplicative Identity) there is an element $1 \in R$ such that $1 \neq 0$ and $1x = x$ for all $x \in R$;

D. (Distributive Law) $x(y + z) = xy + xz$ for all $x, y, z \in R$.

A large number of familiar properties of numbers can be proved using these axioms, and this means that these properties hold in any commutative ring. We will prove some of these in the examples and exercises.

Example 1.3.2. If F is a commutative ring and $x, y, z \in F$, prove that

(a) $x + z = y + z$ implies $x = y$;

(b) $x \cdot 0 = 0$;

(c) $(-x)y = -xy$;

Solution: Suppose $x + z = y + z$. On adding $-z$ to both sides, this becomes

$$(x + z) + (-z) = (y + z) + (-z).$$

Applying the associative law of addition (**A2**) yields

$$x + (z + (-z)) = y + (z + (-z)).$$

But $(z + (-z)) = 0$ by **A4** and $x + 0 = x$ by **A3** and **A1**. Similarly, $y + 0 = y$. We conclude that $x = y$. This proves (a).

By **A3**, $0 + 0 = 0$. By **D** and **A3**,

$$x \cdot 0 + x \cdot 0 = x \cdot (0 + 0) = x \cdot 0 = 0 + x \cdot 0.$$

Using (a) above, we conclude that $x \cdot 0 = 0$.

To prove (c), we first note that, by definition, $-xy$ is the additive inverse of xy (it follows from (a) that there is only one of these). We will show that $(-x)y$ is also an additive inverse for xy. By **D**, (b), and **A1**,

$$xy + (-x)y = (x + (-x))y = 0 \cdot y = 0.$$

This proves that $(-x)y$ is an additive inverse for xy and, hence, it must be $-xy$.

Subtraction in a commutative ring is defined in terms of addition and the additive inverse by setting

$$x - y = x + (-y).$$

The system of integers satisfies all the laws of Definition 1.3.1, and so it is a commutative ring. In fact, it is a commutative ring with an order relation, since the order relation on $\mathbb{N}$ can be used to define a compatible order relation on $\mathbb{Z}$. However, $\mathbb{Z}$ is still inadequate as a number system. This is due to our need to talk about fractional parts of things. This defect is fixed by passing from the integers to the rational numbers.

The Field of Rational Numbers

A *field* is a commutative ring in which division is possible as long as the divisor is not 0. That is,

Definition 1.3.3. A field is a commutative ring satisfying the additional axiom:

M4. (Multiplicative Inverses) for each non-zero element x there is an element x^{-1} such that $x^{-1}x = 1$.

In a field, an element y can be divided by any non-zero element x. The result is $x^{-1}y$, which can also be written as y/x or $\frac{y}{x}$.

The rational number system $\mathbb{Q}$ is a field that is constructed directly from the integers. The construction begins by considering all symbols of the form $\frac{n}{m}$, with $n, m \in \mathbb{Z}$ and $m \neq 0$. We identify two such symbols $\frac{n}{m}$ and $\frac{p}{q}$ whenever $nq = mp$. The resulting object is called a fraction. Thus, $\frac{4}{6}$ and $\frac{2}{3}$ represent the same fraction because $4 \cdot 3 = 6 \cdot 2$. The set $\mathbb{Q}$ is then the set of all fractions.

Addition and multiplication in $\mathbb{Q}$ are defined in the familiar way:

$$\frac{n}{m} + \frac{p}{q} = \frac{nq + mp}{mq} \quad \text{and} \quad \frac{n}{m} \cdot \frac{p}{q} = \frac{np}{mq}.$$

A fraction of the form $\frac{n}{1}$ is identified with the integer n. This makes the set of integers $\mathbb{Z}$ a subset of $\mathbb{Q}$.

The above construction yields a system that satisfies **A1** through **A4**, **M1** through **M4** and **D**. It is therefore a field. We call it the field of rational numbers and denote it by $\mathbb{Q}$. We won't prove here that $\mathbb{Q}$ satisfies all of the field axioms, but a few of them will be verified in the examples and exercises of this section. We will also use the examples and exercises to show how the field axioms can be used to prove other standard facts about arithmetic in fields such as $\mathbb{Q}$.

Example 1.3.4. Assuming that $\mathbb{Z}$ satisfies the axioms of a commutative ring , verify that $\mathbb{Q}$ satisfies **A3** and **M3**.

Solution: The additive identity in $\mathbb{Z}$ is the integer 0, which is identified with the fraction $\frac{0}{1}$. If we add this to another fraction $\frac{n}{m}$, the result is

$$\frac{0}{1} + \frac{n}{m} = \frac{0 \cdot m + 1 \cdot n}{1 \cdot m} = \frac{n}{m}.$$

Thus, $0 = \frac{0}{1}$ is an additive identity for $\mathbb{Q}$ and axiom **A3** is satisfied.

The multiplicative identity in $\mathbb{Z}$ is the integer 1 which is identified with the fraction $\frac{1}{1}$. If we multiply this by another fraction $\frac{n}{m}$, the result is

$$\frac{1}{1} \cdot \frac{n}{m} = \frac{1 \cdot n}{1 \cdot m} = \frac{n}{m}.$$

Thus, $1 = \frac{1}{1}$ is a multiplicative identity for $\mathbb{Q}$ and axiom **M3** is satisfied.

Example 1.3.5. Verify that $\mathbb{Q}$ satisfies **M4**.

Solution: We know that the elements of $\mathbb{Q}$ of the form $\frac{0}{m}$ represent the zero element of $\mathbb{Q}$. Thus, each non-zero element is represented by a fraction $\frac{n}{m}$ in which $n \neq 0$. Then $\frac{m}{n}$ is also a fraction, and

$$\frac{m}{n} \cdot \frac{n}{m} = \frac{nm}{nm} = \frac{1}{1} = 1.$$

Thus, $\frac{m}{n}$ is a multiplicative inverse for $\frac{n}{m}$. This proves that **M4** is satisfied in $\mathbb{Q}$.

The Ordered Field of Rational Numbers

Using the order relation on the integers, it is easy to define an order relation on $\mathbb{Q}$. If r is an element of $\mathbb{Q}$, then we declare $r \geq 0$ if r can be represented in the form $\frac{n}{m}$ for integers $n \geq 0$ and $m > 0$. The order relation is then defined by declaring

$$\frac{p}{q} \leq \frac{n}{m} \quad \text{if and only if} \quad \frac{n}{m} - \frac{p}{q} \geq 0.$$

With the order relation defined this way, $\mathbb{Q}$ becomes an ordered field. That is, it satisfies the axioms in the following definition.

Definition 1.3.6. A field F is called an *ordered field* if it has an order relation "$\leq$" such that the following are satisfied for all $x, y, z \in F$:

O1. either $x \leq y$ or $y \leq x$;

O2. if $x \leq y$ and $y \leq x$, then $x = y$:

O3. if $x \leq y$ and $y \leq z$, then $x \leq z$.

O4. if $x \leq y$, then $x + z \leq y + z$;

O5. if $x \leq y$ and $0 \leq z$, then $xz \leq yz$.

Remark 1.3.7. Given an order relation "$\leq$", we don't distinguish between the statements "$x \leq y$ and "$y \geq x$" – they mean the same thing. Also, If $x \leq y$ and $x \neq y$, then we write $x < y$ or, equivalently, $y > x$.

Example 1.3.8. Prove that if F is an ordered field, then

(a) if $x, y \in F$ and $x \leq y$, then $-y \leq -x$;

(b) if $x \in F$, then $x^2 \geq 0$;

(c) $0 < 1$.

Solution: If $x \leq y$, then $0 = x - x \leq y - x$ by **O4**. Using **O4** again, along with **A1** through **A4** yields $-y \leq (y - x) - y = -x$. This completes the proof of (a).

By **O1**, if $x \in F$, then $0 \leq x$ or $x \leq 0$. If $0 \leq x$, then we multiply this inequality by x and use **O4** to conclude that $0 \leq x^2$. On the other hand, suppose $x \leq 0$. Then, by Part (a), $0 \leq -x$. As above, we conclude that $0 \leq (-x)^2$. Since $(-x)^2 = x^2$ (Exercise 1.3.6), the proof of Part (b) is complete.

Since $1^2 = 1$, Part (b) implies that $0 \leq 1$. By **M3**, $1 \neq 0$ and so $0 < 1$.

Defects of the Rational Field

The rational number system is very satisfying in many ways and is highly useful. However, there are real world mathematic problems that appear to have real world numerical solutions, but these solutions cannot be rational numbers. For example, the Pythagorean Theorem tells us that if the legs of a right triangle have length a and b, then the length c of the hypotenuse satisfies the equation

$$c^2 = a^2 + b^2.$$

However, there are many examples of rational and even integer choices for a and b, such that this equation has no rational solution for c. The simplest example is $a = b = 1$. The Pythagorean Theorem says that a right triangle with legs of length 1 has a hypotenuse of length c satisfying $c^2 = 2$. However, there is no rational number whose square is 2. We will prove this using the following theorem:

Theorem 1.3.9. *If k is an integer and the equation $x^2 = k$ has a rational solution, then that solution is actually an integer.*

Proof. Suppose r is a rational number such that $r^2 = k$. Let $r = \frac{n}{m}$ be r expressed as a fraction in which n and m have no common factors. Then,

$$\left(\frac{n}{m}\right)^2 = k \quad \text{and so} \quad n^2 = m^2 k$$

This equation implies that m divides n^2. However, if $m \neq 1$, then m can be expressed as a product of primes, and each of these primes must also divide n^2. However, if a prime number divides n^2, it must also divide n (Exercise 1.3.15). Thus, each prime factor of m divides n. Since n and m have no common factors, this is impossible. We conclude that $m = 1$ and, hence, that $r = n$ is an integer. □

Now it is easy to see that 2 is not the square of a rational number. If it were, that number would have to be an integer, by the above theorem. The

only possibilities are $-1, 0, 1$ since all other integers have squares that are too large. Of course, none of the numbers $-1, 0, 1$ has square equal to 2.

  Other geometric objects also lead to the conclusion that the system of rational numbers is not sufficient for the measurement of objects that occur in the natural world. The area π of a circle of radius 1 is not a rational number, for example. In fact, the rational number system is riddled with holes where there ought to be numbers. This problem is fixed by the introduction of the system of real numbers which is the topic of the next section.

Exercise Set 1.3

1. Given that $\mathbb{N}$ has an operation of addition which is commutative and associative, how would you define such an addition operation in $\mathbb{Z}$?

2. Referring to the previous exercise, answer the same question for the operation of multiplication.

3. Prove that if $\mathbb{Z}$ satisfies the axioms for a commutative ring, then $\mathbb{Q}$ satisfies **A1** and **M1**.

4. Prove that if $\mathbb{Z}$ satisfies the axioms for a commutative ring, then $\mathbb{Q}$ satisfies **A2** and **M2**.

In the next three exercises you are to prove the given statement assuming x, y, z are elements of a field. You may use the results of examples and theorems from this section.

6. $(-x)(-y) = xy$.

7. $xz = yz$ implies $x = y$, provided $z \neq 0$.

8. $xy = 0$ implies $x = 0$ or $y = 0$.

In the next three exercises you are to prove the given statement assuming x, y, z are elements of an ordered field. Again, you may use the results of examples and theorems from this section.

9. $x > 0$ and $y > 0$ imply $xy > 0$.

10. $x > 0$ implies $x^{-1} > 0$.

11. $0 < x < y$ implies $y^{-1} < x^{-1}$.

12. Prove that the equation $x^2 = 5$ has no rational solution.

13. Generalize Theorem 1.3.9 by proving that every rational solution of a polynomial equation

$$x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 = 0,$$

with integer coefficients a_k, is an integer solution.

14. Prove that if m and n are positive integers with no common factors other than 1 (i. e. m and n are relatively prime), then there are integers a and b such that $1 = am + bn$. Hint: let S be the set of all positive integers of the form $am + bn$, where a and b are integers. This set has a smallest element by Exercise 1.2.19. Use the division algorithm (Exercise 1.2.20) to show that this smallest element divides both m and n.

15. Use the result of the preceding exercise to prove that if a prime p divides the product nm of two positive integers, then it divides n or it divides m.

1.4 The Real Numbers

As pointed out in the previous section, the set of rational numbers is riddled with "holes" where there ought to be numbers. Here we will try to make this statement more precise and then indicate how these holes can be "filled" resulting in the system of real numbers. In addition to the ordered field axioms, the real number system satisfies a new axiom **C** – the completeness axiom. Later in the section we will state it and explore its consequences.

The construction of the real numbers that we outline below is motivated by the idea that a "hole" in the rational numbers is a location along the rational number line where there should be a number but there is no rational number. What do we mean by a "location" along the rational number line? Well if this has meaning, then it should make sense to talk about the rational numbers that are to the left of this location and those that are to the right of this location. This should lead to a separation of the rational numbers into two sets – one to the left and one to the right of the given location. In fact, we can *define* a location on the rational line to *be* such a separation. This leads to the notion of a *Dedekind cut*.

Dedekind Cuts

If r is a rational number, consider the infinite interval L_r consisting of all rational numbers to the left of r. That is,

$$L_r = \{x \in \mathbb{Q} : x < r\}. \tag{1.4.1}$$

This set is a non-empty, proper subset of $\mathbb{Q}$. It has no largest element, since, for each $x < r$, there are rational numbers larger than x that are also less than r (for example, $(x + r)/2$ is one such number). It also has the property that if $x \in L_r$, then so is any rational number less than x. It turns out that there are also subsets of $\mathbb{Q}$ with these three properties that are not of the form L_r for some rational number. A subset of $\mathbb{Q}$ with these three properties is called a *Dedekind cut*. That is,

Definition 1.4.1. A subset L of $\mathbb{Q}$ is called a *Dedekind cut*, or simply a *cut* in the rationals, if it satisfies the following three conditions:
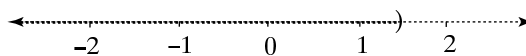
Figure 1.2: A Dedekind Cut in the Rationals.

(a) $L \neq \emptyset$ and $L \neq \mathbb{Q}$;

(b) L has no largest element;

(c) if $x \in L$ then so is every y with $y < x$.

The reason for calling such a set L a "cut" is that, if R is the complement of L, then each number in L is to the left of each number in R. Thus, the rational line is separated or *cut* into left and right halves. Since each half determines the other, we choose to focus on just the left half in this discussion.

Each rational number r determines a cut – the set L_r of (1.4.1). In this case, r is called the *cut number* for the Dedekind cut. Are there Dedekiind cuts that are not determined in this way? cuts that have no rational cut number?

Example 1.4.2. Describe a Dedekind cut that is not of the form L_r for a rational number r.

Solution: We are guided by the idea that there ought to be a number whose square is 2, but there is no such rational number. If there were a number $\sqrt{2}$ with square 2, then the set of rational numbers less than $\sqrt{2}$ could be described as

$$L = \{r \in \mathbb{Q} : r \geq 0 \text{ and } r^2 < 2\} \cup \{r \in \mathbb{Q} : r < 0\}.$$

We claim this a Dedekind cut not of the form L_r for any $r \in \mathbb{Q}$.

Certainly L is a non-empty, proper subset of $\mathbb{Q}$. It has no largest element because if $\frac{n}{m}$ is any positive element of L, then we can always choose a larger rational number which is still has square less than 2 as follows: $\frac{kn+1}{km} > \frac{n}{m}$ for every $k \in \mathbb{N}$ and

$$\left(\frac{kn+1}{km}\right)^2 = \left(\frac{n}{m}\right)^2 + \frac{1}{km}\left(2\frac{n}{m} + \frac{1}{km}\right).$$

By choosing k large enough, we can make the second term on the right less than $2 - \left(\frac{n}{m}\right)^2$ and this will imply that $\left(\frac{kn+1}{km}\right)^2 < 2$. Thus, L has no largest element.

If $x \in L$ and $y < x$, then either y is negative, in which case it is in L, or $0 \leq y < x$. In the latter case, $y^2 < x^2 < 2$, and so $y \in L$ in this case as well. Thus L is a Dedekind cut.

We next show that there is no rational number r such that $L = L_r$. If there is such a number r, then r is a positive rational number not in L and so $r^2 \geq 2$. However, there are numbers in L arbitrarily close to r and each of them has square less than 2. It follows that $r^2 \leq 2$. This means $r^2 = 2$, which is impossible for a rational number r.

Thus, although it might seem that every Dedekind cut ought to correspond to a cut number, the above example shows that this is not the case. In fact, there are a lot more cuts than there are rational cut numbers. However, we can fix this by enlarging the number system so that there is a cut number for every Dedekind cut. The way this is usually done is to define the new number system to actually *be* the set of all Dedekind cuts of the rationals. Below, we attempt to describe this idea in a way that is somewhat visually intuitive.

We will think of a Dedekind cut L as specifying a certain location (the location between L and its complement R) along the rational number line. We will think of the real number system $\mathbb{R}$ as being the set of all such locations. Then each real number x corresponds to a Dedekind cut L_x, which is to be thought of as the set of all rational numbers to the left of the location x. We next need to define an order relation and operations of addition and multiplication in $\mathbb{R}$.

The order relation on $\mathbb{R}$ is simple: We say $x \leq y$ if $L_x \subset L_y$. An element $x \in \mathbb{R}$ is, then, non-negative if $L_0 \subset L_x$. With this definition of order on $\mathbb{R}$ we can assert that

$$L_x = \{r \in \mathbb{Q} : r < x\}$$

for all $x \in \mathbb{R}$ (not just for $x \in \mathbb{Q}$).

Addition of real numbers is defined as follows: If $x, y \in \mathbb{R}$, then we set

$$L_x + L_y = \{r + s : r \in L_x, s \in L_y\}.$$

It is easily verified that this is also a Dedekind cut (Exercise 1.4.10) and, hence, it corresponds to an element of $\mathbb{R}$. We define $x + y$ to be this element.

The product of two non-negative numbers x and y is defined as follows: we set

$$K = \{rs : r \in L_x, r \geq 0, s \in L_y, s \geq 0\} \cup \{t \in \mathbb{Q} : t < 0\}.$$

This is a Dedekind cut (Exercise 1.4.11), and we define xy to be the corresponding element of $\mathbb{R}$. For pairs of numbers where one or both is negative, the definition of product is more complicated due to the fact that multiplication by a negative number reverses order.

Of course $\mathbb{Q} \subset \mathbb{R}$, since each rational number was already the cut number of a Dedekind cut. It is easily checked that the definitions of addition, multiplication and order given above agree with the usual ones in the case that the numbers are rational.

The numbers in $\mathbb{R}$ that are not in $\mathbb{Q}$ are called *irrational* numbers. It turns out that there are many more irrational numbers than there are rational numbers. To make sense of this statement requires a discussion of finite sets and infinite sets, and how some infinite sets are larger than others. We present such a discussion in the appendix.

The Completeness Axiom

This is the property of the real number system that distinguishes it from the rational number system. Without it, most of the theorems of calculus would

not be true.

A subset A of an ordered field F is said to be *bounded above* if there is an element $m \in F$ such that $x \leq m$ for every $x \in A$. The element m is called an *upper bound* for A. If, among all upper bounds for A, there is one which is smallest (less than all the others), then we say that A has a *least upper bound*.

Definition 1.4.3. An ordered field F is said to be *complete* if it satisfies:

C. each non-empty subset of F which is bounded above has a least upper bound.

If one defines the real number system $\mathbb{R}$ in terms of Dedekind cuts of the rationals and defines addition, multiplication, and order as above, then one can prove that the resulting system is an ordered field. To carry out all the details of this proof is a long and tedious process and it will not be done here. However, it is quite easy to prove that $\mathbb{R}$, as defined in this way, satisfies the completeness axiom **C**.

Theorem 1.4.4. *If $\mathbb{R}$ is defined using Dedekind cuts of $\mathbb{Q}$, as above, then every non-empty subset of $\mathbb{R}$ which is bounded above has a least upper bound.*

Proof. Let A be a bounded subset of $\mathbb{R}$ and let m be any upper bound for A. For each $x \in A$, let L_x be the corresponding cut in $\mathbb{Q}$. Then $x \leq m$ for all $x \in A$ means that $L_x \subset L_m$ for all $x \in A$. We set

$$L = \bigcup_{x \in A} L_x.$$

Then L is a proper subset of $\mathbb{Q}$ because $L \subset L_m$. If $r \in L$ and $s < r$, then $r \in L_x$ for some $x \in A$ and this implies $s \in L_x$ and, hence, $s \in L$. If L had a largest element t, then t would belong to L_x for some x, and it would have to be a largest element for L_x – a contradiction. Thus, L has no largest element. We have now proved that L satisfies (a), (b), and (c) of Definition 1.4.1 and, hence, that L is a Dedekind cut.

If y is the real number corresponding to L, that is if $L = L_y$, then, for all $x \in A$, $L_x \subset L_y$, and this means $x \leq y$. Thus, y is an upper bound for A. Also, $L_y \subset L_m$ means that $y \leq m$. Since m was an arbitrary upper bound for A, this implies that y is the least upper bound for A. This completes the proof. $\square$

This completes our outline of the construction of the real number system beginning with Peano's axioms for the natural numbers. The final result is the following theorem, which we will state without further proof. It will be the starting point for our development of calculus.

Theorem 1.4.5. *The real number system $\mathbb{R}$ is a complete ordered field.*

Example 1.4.6. Find all upper bounds and the least upper bound for the following sets:

$$A = (-1, 2) = \{x \in \mathbb{R} : -1 < x < 2\};$$
$$B = (0, 3] = \{x \in \mathbb{R} : 0 < x \leq 3\}.$$

Solution: The set of all upper bounds for the set A is $\{x \in \mathbb{R} : x \geq 2\}$. The smallest element of this set (the least upper bound of A) is 2. Note that 2 is not actually in the set A.

The set of all upper bounds for B is the set $\{x \in \mathbb{R} : x \geq 3\}$. The smallest element of this set is 3 and so it is the least upper bound of B. Note that, in this case, the least upper bound is an element of the set B.

If the least upper bound of a set A does belong to A, then it is called the *maximum* of A. Note that a non-empty set which is bounded above always has a least upper bound, by Axiom **C**. However, the preceding example shows that it need not have a maximum.

The Archimedean Property

An ordered field always contains a copy of the natural numbers and, hence, a copy of the integers (Exercise 1.4.5). Thus, the following definition makes sense.

Definition 1.4.7. An ordered field is said to have the Archimedean property if, for every $x \in \mathbb{R}$, there is a natural number n such that $x < n$. An ordered field with the Archimedean property is called an *Archimedean ordered field*.

Theorem 1.4.8. *The field of real numbers has the Archimedean property.*

Proof. We use the completeness property. Suppose there is an x such that $n \leq x$ for all $n \in \mathbb{N}$. Then $\mathbb{N}$ is a non-empty subset of $\mathbb{R}$ which is bounded above. By the completeness property, there is a least upper bound b for $\mathbb{N}$. Then b is an upper bound for $\mathbb{N}$, but $b - 1$ is not. This implies there is an $n \in \mathbb{N}$ such that $b - 1 < n$. Then $b < n + 1$, which contradicts the statement that b is an upper bound for $\mathbb{N}$. Thus, the assumption that $\mathbb{N}$ is bounded above by some $x \in \mathbb{R}$ has led to a contradiction. We conclude that every x in $\mathbb{R}$ is less than some natural number. This completes the proof. $\square$

The Archimedean property can be stated in any one of several equivalent ways. One of these is: for every real number $x > 0$, there is an $n \in N$ such that $1/n < x$ (Example 1.4.9). Another is: given real numbers x and y with $x > 0$, there is an $n \in \mathbb{N}$ such that $nx > y$ (Exercise 1.4.6).

Example 1.4.9. Prove that, in an Archimedean field, for each $x > 0$ there is an $n \in \mathbb{N}$ such that $1/n < x$.

Solution The Archimedean property tells us that there is a natural number $n > 1/x$. Since n and x are positive, this inequaltiy is preserved when we multiply it by x and divide it by n. This yields $1/n < x$, as required.

Another consequence of the Archimedean property is that there is a rational number between each distinct pair of real numbers (Exercise 1.4.7).

Exercise Set 1.4

1. For each of the following sets, describe the set of all upper bounds for the set :

    (a) the set of odd integers;

    (b) $\{1 - 1/n : n \in \mathbb{N}\}$;

    (c) $\{r \in \mathbb{Q} : r^3 < 8\}$;

    (d) $\{\sin x : x \in \mathbb{R}\}$.

2. For each of the sets in $(a), (b), (c)$ of the preceding exercise, find the least upper bound of the set, if it exists.

3. Prove that if a subset A of $\mathbb{R}$ is bounded above, then the set of all upper bounds for A is a set of the form $[x, \infty)$. What is x?

4. Show that the set $A = \{x : x^2 < 1 - x\}$ is bounded above, and then find its least upper bound.

5. If F is an ordered field, prove that there is a sequence of elements $\{n_k\}_{k \in \mathbb{N}}$, all different, such that $n_1 = 1$ (the identity element of F), and $n_{k+1} = n_k + 1$ for each $k \in \mathbb{N}$. Argue that the terms of this sequence form a subset of F which is a copy of the natural numbers, by showing that the correspondence $k \to n_k$ is a one-to-one function from $\mathbb{N}$ onto this subset. By definition it takes the successor $k + 1$ of an element $k \in \mathbb{N}$ to the successor $n_k + 1$ of its image n_k.

6. Let F be an ordered field, We consider $\mathbb{N}$ to be a subset of F as described in the preceding exercise. Prove that F is Archimedean if and only if, for each pair $x, y \in F$ with $x > 0$, there exists a natural number n such that $nx > y$.

7. Prove that if $x < y$ are two real numbers, then there is a rational number r with $x < r < y$. Hint: use the result of Example 1.4.9.

8. Prove that if x is irrational and r is a non-zero rational number, then $x + r$ and rx are also irrational.

9. We know that $\sqrt{2}$ is irrational. Use this fact and the previous exercise to prove that if $r < s$ are rational numbers, then there is an irrational number x with $r < x < s$.

The following exercises concern Dedekind cuts of the rationals and should be done using only properties of the rational number system and the definition of Dedekind cut.

10. Show that if L_x and L_y are Dedekind cuts defining real numbers x and y, then
$$L_x + L_y = \{r + s : r \in L_x \text{ and } s \in L_y\}$$

is also a Dedekind cut (this is the Dedekind cut determining the sum $x + y$).

11. If L_x and L_y are Dedekind cuts determining positive real numbers x and y, and if we set

$$K = \{rs : 0 \le r \in L_x \text{ and } 0 \le s \in L_y\} \cup \{t \in \mathbb{Q} : t < 0\},$$

then K is also a Dedekind cut (this is the Dedekind cut determining the product xy).

12. If L is the Dedekind cut of Example 1.4.2 and L determines the real number x (so that $L = L_x$), prove that $L_{x^2} = L_2$. Thus, the real number corresponding to L has square 2.

1.5   Sup and Inf

The concept of least upper bound, which appears in the completeness axiom, will be extremely important in this course. It will be examined in detail in this section. We first note that there is a companion concept for sets that are bounded below.

Greatest Lower Bound

We say a set A is bounded below if there is a number m such that $m \le x$ for every $x \in A$. The number m is called a *lower bound* for A. A *greatest lower bound* for A is a lower bound that is larger than any other lower bound.

Theorem 1.5.1. *Every non-empty subset of $\mathbb{R}$ that is bounded below has a greatest lower bound.*

Proof. Suppose A is a non-empty subset of $\mathbb{R}$ which is bounded below . We must show that there is a lower bound for A which is greater than any other lower bound for A. If m is any lower bound for A, then Example 1.3.8 (a) implies that, $-m$ is an upper bound for $-A = \{-a : a \in A\}$. Since $\mathbb{R}$ is a complete ordered field, there is a least upper bound r for $-A$. Then

$$-a \le r \text{ for all } a \in A \quad \text{and} \quad r \le -m.$$

Applying Example 1.3.8 (a) yields that

$$-r \le a \text{ for all } a \in A \quad \text{and} \quad m \le -r.$$

Thus, $-r$ is a lower bound for A and, since m was an arbitrary lower bound, the inequality $m \le -r$ implies that $-r$ is the greatest lower bound. $\square$

The Extended Real Numbers

For many reasons, it is convenient to extend the real number system by adjoining two new points ∞ and $-\infty$. The resulting set is called the *extended real number system*. We declare that ∞ is greater than and $-\infty$ less than every other extended real number. This makes the extended real number system an ordered set. We also define $x + \infty$ to be ∞ if x is any extended real number other than $-\infty$. Similarly, $x - \infty = x + (-\infty)$ is defined to be $-\infty$ if x is any extended real number other than ∞. Of course, there is no reasonable way to make sense of $\infty - \infty$.

The introduction of the extended real number system is just a convenient notational convention. For example, it allows us to make the following definition.

Sup and Inf

Definition 1.5.2. Let A be an arbitrary non-empy subset of $\mathbb{R}$. We define the *supremum* of A, denoted $\sup A$, to be the smallest extended real number M such that $a \leq M$ for every $a \in A$.

The infimum of A, denoted $\inf A$, is the largest extended real number m such that $m \leq a$ for all $a \in A$.

Note that, if A is bounded above, then $\sup A$ is the least upper bound of A. If A is not bounded above, then the only extended real number M with $a \leq M$ for all $a \in A$ is ∞, and so $\sup A = \infty$ in this case. Similarly, $\inf A$ is the greatest lower bound of A if A is bounded below and is $-\infty$ if A is not bounded below. Thus, $\sup A$ and $\inf A$ exist as extended real numbers for any non-empty set A, but they might not be finite. Also note that, even when they are finite real numbers, they may not actually belong to A, as Example 1.4.6 shows.

Example 1.5.3. Find the sup and inf of the following sets:

$$A = (-1, 1] = \{x \in \mathbb{R} : -1 < x \leq 1\};$$
$$B = (-\infty, 5) = \{x \in \mathbb{R} : x < 5\}.$$
$$C = \left\{ \frac{n^2}{n+1} : n \in \mathbb{N} \right\} \tag{1.5.1}$$
$$D = \left\{ \frac{1}{n} : n \in \mathbb{N} \right\} \tag{1.5.2}$$

Solution: Clearly, $\inf A = -1$ and $\sup A = 1$. These are finite, $\sup A$ belongs to A, but $\inf A$ does not.

Also, $\inf B = -\infty$ and $\sup B = 5$. In this case, the inf is not finite. The sup is finite but does not belong to B.

Since $\dfrac{n^2}{n+1} \geq \dfrac{n}{2}$, the set C is unbounded, and so $\sup C = \infty$. Also, we have $n + 1 \leq n^2 + n^2 = 2n^2$, and so

$$\frac{1}{2} \leq \frac{n^2}{n+1}$$

for all $n \in \mathbb{N}$. Thus, $1/2$ is a lower bound for C. It is the greatest lower bound, since it actually belongs to C, due to the fact that $\dfrac{n^2}{n+1} = \dfrac{1}{2}$ when $n = 1$. Thus, $\inf C = 1/2$.

Certainly 0 is a lower bound for the set D. It follows from the Archimedean property (see Example 1.4.9) that there is no $x \in F$ with $x > 0$ which is a lower bound for this set, and so 0 is the greatest lower bound. Thus, $\inf D = 0$. Clearly, $\sup D = 1$.

If A is a set of numbers and $\sup A$ actually belongs to A, then it is called the *maximum* of A and denoted $\max A$. Similarly, if $\inf A$ belongs to A, then it is called the *minimum* of A and is denoted $\min A$.

The following theorem is really just a restatement of the definition of sup, but it may give some helpful insight. It says that $\sup A$ is the dividing point between the numbers which are upper bounds for A (if there are any) and the numbers which are not upper bounds for A. A similar theorem holds for inf. Its formulation and proof are left to the exercises.

Theorem 1.5.4. *Let A be a non-empty subset of $\mathbb{R}$ and x an element of $\mathbb{R}$. Then*

(a) $\sup A \leq x$ if and only if $a \leq x$ for every $a \in A$;

(b) $x < \sup A$ if and only if $x < a$ for some $a \in A$.

Proof. (a) By definition $a \leq x$ for every $a \in A$ if and only if x is an upper bound for A. If x is an upper bound for A, then A is bounded above. This implies its *sup* is its least upper bound, which is necessarily less than or equal to x.

Conversely, if $\sup A \leq x$, then $\sup A$ is finite and is the least upper bound for A. Since $\sup A \leq x$, x is also an upper bound for A. Thus, $\sup A \leq x$ if and only if $a \leq x$ for every $a \in A$.

(b) If $x < \sup A$, then x is not an upper bound for A, which means that $x < a$ for some $a \in A$. Conversely, if $x < a$ for some $a \in A$, then $x < \sup A$, since $a \leq \sup A$. Thus, $x < \sup A$ if and only if $x < a$ for some $a \in A$. $\qquad \square$

Example 1.5.5. If $A = \left\{ \dfrac{4n-1}{6n+3} : n \in \mathbb{N} \right\}$, find the set of all upper bounds for A.

Solution: Long division yields

$$\frac{4n-1}{6n+3} = \frac{2}{3} - \frac{1}{2n+1} \leq \frac{2}{3}.$$

Thus, $2/3$ is an upper bound for A. If $x < 2/3$, then $\epsilon = 2/3 - x$ is positive, and the Archimedean Property implies we can choose n large enough that

$$\frac{1}{2n+1} < \frac{1}{n} < \epsilon.$$

Then

$$x < \frac{2}{3} - \frac{1}{2n+1} = \frac{4n-1}{6n+3}$$

for such an n, which means that x is not an upper bound for A.

We conclude that $2/3$ is the least upper bound for A – that is $\sup A = 2/3$. By the previous theorem, the set of all upper bounds for A is the interval $[2/3, \infty)$.

Example 1.5.6. If $A = \left\{ \dfrac{n^2}{n+1} : n \in \mathbb{N} \right\}$, find $\sup A$ and the set of all upper bounds for A.

   Solution: Long division yields

$$\frac{n^2}{n+1} = n - 1 + \frac{1}{n+1} \geq n - 1.$$

Then the Archmedean Property implies that there are no upper bounds for A, since, for every $x \in \mathbb{R}$, there is an $n \in \mathbb{N}$ for which $n - 1$ is larger than x. Thus, the set of upper bounds for A is the empty set and $\sup A = \infty$.

Properties of Sup and Inf

The next theorem uses the following notation concerning subsets A and B of $\mathbb{R}$:

$$-A = \{-a : a \in A\};$$
$$A + B = \{a + b : a \in A, b \in B\}$$
$$A - B = \{a - b : a \in A, b \in B\}.$$

Theorem 1.5.7. *Let A and B be non-empty subsets of $\mathbb{R}$. Then*

(a) $\inf A \leq \sup A$;

(b) $\sup(-A) = -\inf A$ *and* $\inf(-A) = -\sup A$;

(c) $\sup(A + B) = \sup A + \sup B$ *and* $\inf(A + B) = \inf A + \inf B$;

(d) $\sup(A - B) = \sup A - \inf B$;

(e) if $A \subset B$, then $\sup A \leq \sup B$ *and* $\inf B \leq \inf A$.

Proof. We will prove (a), (b), and (c) and leave (d) and (e) to the exercises.

   (a) If A is non-empty, then there is an element $a \in A$. Since $\inf A$ is a lower bound and $\sup A$ an upper bound for A, we have $\inf A \leq a \leq \sup A$.

   (b) A number x is a lower bound for the set A ($x \leq a$ for all $a \in A$) if and only if $-x$ is an upper bound for the set $-A$ ($-a \leq -x$ for all $a \in A$). Thus, if L is the set of all lower bounds for A, then $-L$ is the set of all upper bounds for $-A$. Furthermore, the largest member of L and the smallest member of $-L$ are negatives of each other. That is, $-\inf A = \sup(-A)$. This is the first equality

in (b) If we apply this result with $-A$ replacing A, we have $-\inf(-A) = \sup A$. If we multiply this by -1, we get the second equality in (b).

(c) Since $a \le \sup A$ and $b \le \sup B$ for all $a \in A$, $b \in B$, we have

$$a + b \le \sup A + \sup B \quad \text{for all} \quad a \in A, \ b \in B.$$

It follows that
$$\sup(A + B) \le \sup A + \sup B.$$

Let x be any number less than $\sup A + \sup B$. We claim that there are elements $a \in A$ and $b \in B$ such that

$$x < a + b. \tag{1.5.3}$$

Once proved, this will imply that no number less than $\sup A + \sup B$ is an upper bound for $A + B$. Thus, proving this claim will establish that $\sup(A + B) = \sup A + \sup B$.

There are two cases to consider: $\sup B$ finite and $\sup B = \infty$. If $\sup B$ is finite, then $x - \sup B < \sup A$, and Theorem 1.5.4 implies there is an $a \in A$ with $x - \sup B < a$. Then $x - a < \sup B$. Applying Theorem 1.5.4 again, we conclude there is an $b \in B$ with $x - a < b$. This implies (1.5.3), and proves our claim in the case where $\sup B$ is finite.

Now suppose $\sup B = \infty$. Let a be any element of A. Then $x - a < \sup B = \infty$ and so, as above, we conclude from Theorem 1.5.4 that there is a $b \in B$ satisfying $x - a < b$. This implies (1.5.3), which establishes our claim in this case and completes the proof. $\qquad\square$

Sup and Inf for Functions

If f is a real valued function defined on some set X and if A is a subset of X, then

$$f(A) = \{f(x) : x \in A\}$$

is a set of real numbers, and so we can take its sup and inf.

Definition 1.5.8. If $f : X \to \mathbb{R}$ is a function and $A \subset X$, then we set

$$\sup_A f = \sup\{f(x) : x \in A\} \quad \text{and} \quad \inf_A f = \inf\{f(x) : x \in A\}.$$

Thus, $\sup_A f$ is the supremum of the set of values that f assumes on A and $\inf_A f$ is the infimum of this set. They themselves may or may not be values that f assumes on A. If $\sup_A f$ is a value that f assumes on A, then it is called the *maximum* of f on A. Similarly, if $\inf_A f$ is a value assumed by f somewhere on A, then it is called the *minimum* of f on A.

Example 1.5.9. Find $\sup_I f$ and $\inf_I f$ if

(a) $f(x) = \sin x$ and $I = [-\pi/2, \pi/2)$;

(b) $f(x) = 1/x$ and $I = (0, \infty)$.

   Solution: (a) The function $\sin x$ takes on all values in the interval $[-1,1)$ on I, but does not take on the value 1. Thus, $\inf_I f = -1$ and $\sup_I f = 1$. In this case, $\inf_I f$ is a value assumed by f on I, but $\sup_I f$ is not.

   (b) the function $1/x$ takes on all values in the open interval $(0, \infty)$. Thus, $\inf_I f = 0$ and $\sup_i f = \infty$ in this case. Neither one of these extended real numbers is a value taken on by f on I.

   The following theorem concerning sup and inf for functions follows easily from Theorem 1.5.7. We leave the details to the exercises.

Theorem 1.5.10. *Let f and g be functions defined on a set containing A as a subset, and let $c \in \mathbb{R}$ be a positive constant. Then*

   (a) $\sup_A cf = c\sup_A f$ *and* $\inf_A cf = c\inf_A f$;

   (b) $\sup_A(-f) = -\inf_A f$;

   (c) $\sup_A(f + g) \le \sup_A f + \sup_A g$ *and* $\inf_A f + \inf_A g \le \inf_A(f + g)$;

   (d) $\sup\{f(x) - f(y) : x, y \in A\} = \sup_A f - \inf_A f$.

Exercise Set 1.5

1. For each of the following sets, find the set of all extended real numbers x that are greater than or equal to every element of the set. Then find the sup of the set. Does the set have a maximum?

   (a) $(-10, 10)$;

   (b) $\{n^2 : n \in \mathbb{N}\}$;

   (c) $\left\{ \dfrac{2n+1}{n+1} \right\}$.

2. Find the sup and inf of the following sets. Tell whether each set has a maximum or a minimum.

   (a) $(1, 8]$;

   (b) $\left\{ \dfrac{n+2}{n^2+1} \right\}$;

   (c) $\{n/m : n, m \in \mathbb{Z},\ n^2 < 5m^2\}$;

3. Prove that if $\sup A < \infty$, then for each $n \in \mathbb{N}$ there is an element $a_n \in A$ such that $\sup A - 1/n < a_n \le \sup A$.

4. Prove that if $\sup A = \infty$, then for each $n \in \mathbb{N}$ there is an element $a_n \in A$ such that $a_n > n$.

5. Formulate and prove the analog of Theorem 1.5.4 for inf.

6. Prove part (d) of Theorem 1.5.7.

7. Prove (e) of Theorem 1.5.7.

8. if A and B are two non-empty sets of real numbers, then prove that

$$\sup(A \cup B) = \max\{\sup A, \sup B\} \quad \text{and} \quad \inf(A \cup B) = \min\{\inf A, \inf B\}.$$

9. Find $\sup_I f$ and $\inf_I f$ for the following functions f and sets I. Which of these is actually the maximum or the minimum of the function f on I.

   (a) $f(x) = x^2$, $I = [-1, 1]$;
   (b) $f(x) = \dfrac{x+1}{x-1}$, $I = (1, 2)$;
   (c) $f(x) = 2x - x^2$, $I = [0, 1)$.

10. Prove (a) of Theorem 1.5.10

11. Prove (b) of Theorem 1.5.10

12. Prove (c) of Theorem 1.5.10

13. Prove (d) of Theorem 1.5.10

Chapter 2

Sequences

In this chapter we have our first encounter with the concept of limit – the concept that lies at the heart of the calculus. We first study limits of sequences of real numbers. Limits of functions will be studied in the next chapter.

2.1 Limits of Sequences

Limits make sense in any context in which we have a notion of distance between objects. Thus, we begin with a discussion of the notion of distance between two real numbers.

Distance and Absolute Value

Recall that the absolute value $|x|$ of a number x is defined by

$$|x| = \left\{ \begin{array}{lll} x & \text{if} & x \geq 0 \\ -x & \text{if} & x < 0. \end{array} \right\}$$

Thus, $|x|$ is always a non-negative number. It can be thought of as the distance from x to 0. For example,

$$|3| = |-3| = 3,$$

just means that the distance from 3 to 0 and the distance from -3 to 0 are the same, namely 3. More generally, if x and y are any two real numbers, the distance from x to y is $|x - y|$.

We will often need to specify that a number x is close to another number a. However, this doesn't mean anything unless we specify how close. If ϵ is a positive number, then the statement "x is within ϵ of a" does have meaning. It means that the distance between x and a is less than ϵ – that is

$$|x - a| < \epsilon.$$

This statement also means that x is in the open interval of radius ϵ, centered at a, as pointed out in Part (b) of the following theorem.

Theorem 2.1.1. *If x, y, a and ϵ are real numbers with $\epsilon > 0$, then*

 (a) $|y| < \epsilon$ *if and only if* $-\epsilon < y < \epsilon$;

 (b) $|x - a| < \epsilon$ *if and only if* $a - \epsilon < x < a + \epsilon$.

These statements remain true if "$<$" is replaced by "$\leq$".

Proof. To prove (a), we consider two cases:

 1. Suppose $y \geq 0$. Then $|y| = y$, and so $|y| < \epsilon$ if and only if $y < \epsilon$. The latter statement means the same as $-\epsilon < y < \epsilon$, because $-\epsilon < y$ is automatically true in this case.

 2. Suppose $y < 0$. Then $|y| = -y$, and so $|y| < \epsilon$ if and only if $-y < \epsilon$. This is true if and only if $-\epsilon < y$, which is true if and only if $-\epsilon < y < \epsilon$, because $y < \epsilon$ is automatically true in this case.

   Part (b) follows from Part (a). That is, if we apply Part (a) with $y = x - a$, then we conclude that $|x - a| < \epsilon$ if and only if $-\epsilon < x - a < \epsilon$, and this is true if and only if $a - \epsilon < x < a + \epsilon$.
   If "$<$" is replaced by "$\leq$" the proofs of (a) and (b) remain the same. □

   The following theorem will be used extensively throughout the text.

Theorem 2.1.2. (Triangle Inequality) *If a and b are real numbers, then*

 (a) $|a + b| \leq |a| + |b|$; *and*

 (b) $||a| - |b|| \leq |a - b|$.

Proof. For part (a), we observe that $-|a| \leq a \leq |a|$ and $-|b| \leq b \leq |b|$. If we add these inequalities, the result is

$$-(|a| + |b|) \leq a + b \leq |a| + |b|$$

By the preceding theorem (with "$<$" replaced by "$\leq$"), this is equivalent to $|a + b| \leq |a| + |b|$. This proves Part (a).
   For part (b), we note that part (a) implies $|a| = |b + (a - b)| \leq |b| + |a - b|$ and this yields

$$|a| - |b| \leq |a - b| \tag{2.1.1}$$

when we subtract $|b|$ from both sides. If we interchange b and a, then the right side of this inequality stays the same and the left side becomes $|b| - |a|$. Thus, the inequality

$$|b| - |a| \leq |b| + |a - b|$$

also holds. This, and (2.1.1) together imply Part (b). □

Sequences

A sequence of real numbers is a function from the natural numbers to the real numbers. That is, it is an assignment of a real number a_n to each natural number n. Traditionally, we use the notation

$$\{a_n\}_{n=1}^{\infty} \quad \text{or simply} \quad \{a_n\},$$

to denote a sequence, rather than using standard function notation. Alternatively, we may describe a sequence by writing out its first few terms and possibly its nth term:

$$a_1, a_2, a_3 \cdots \quad \text{or} \quad a_1, a_2, a_3, \cdots, a_n, \cdots .$$

Example 2.1.3. Write each of the following sequences in the form

$$a_1, a_2, a_3, \cdots, a_n, \cdots .$$

(a) the sequence $\{(-1)^n 1/n\}$;

(b) the sequence of positive even integers;

(c) the sequence defined inductively by $a_1 = 2$ and $a_{n+1} = \dfrac{a_n + 1}{2}$.

   Solution: The answers are

(a) $-1, 1/2, -1/3, \cdots, (-1)^n 1/n, \cdots$;

(b) $2, 4, 6, \cdots, 2n, \cdots$;

(c) $2, 3/2, 5/4, \cdots, 1 + 1/2^{n-1}, \cdots$.

The first two are obvious. For (c), we prove that $a_n = 1 + 1/2^{n-1}$ by induction. This is certainly true for $n = 1$. If it is correct for an integer n, then $a_n = 1 + 1/2^{n-1}$ and so

$$a_{n+1} = (a_n + 1)/2 = (1 + 1/2^{n-1} + 1)/2 = 1 + 1/2^n.$$

Thus, our formula for a_n is true for $n + 1$ if it is true for n. By induction, it is true for all natural numbers.

   It is sometimes convenient to begin the indexing of a sequence at some integer k other than 1. For example, the sequence

$$1, 2, 4, 8, \cdots, 2^n, \cdots$$

has description $n \to 2^{n-1}$ as a function from the natural numbers to the real numbers, or, using standard sequence notation, $\{2^{n-1}\}_{n=1}^{\infty}$, but it is usually more convenient to think of it as the function $n \to 2^n$ from the non-negative integers to the reals, and denote it $\{2^n\}_{n=0}^{\infty}$. Similarly, the sequence

$$8/3, 4, 32/5, 32/3, 128/7, \cdots$$

can be described as the sequence $\left\{ \dfrac{2^{n+2}}{n+2} \right\}_{n=1}^{\infty}$, but it may be more convenient

to describe it as $\left\{ \dfrac{2^n}{n} \right\}_{n=3}^{\infty}$. Passing from one notation to the other is a *change of variables* in the index – that is, n is replaced by $n-2$ and the starting point for the sequence is changed from $n=1$ to $n=3$ (since $n-2$ is 1 when n is 3).

Limits of Sequences

A sequence $\{a_n\}$ converges to a number a if the distance from a_n to a can be made less than any given positive number by insisting that n be sufficiently large. More precisely:

Definition 2.1.4. A sequence $\{a_n\}$ of real numbers is said to *converge* to the number a, or have *limit* equal to a, if, for each $\epsilon > 0$, there is a real number N such that

$$|a_n - a| < \epsilon \quad \text{whenever} \quad n > N.$$

In this case, we will write $\lim\limits_{n \to \infty} a_n = a$ or $\lim a_n = a$ or simply $a_n \to a$.

Remark 2.1.5. If we compare what would be required by the above definition for $\lim a_n = a$ and what would be required for $\lim |a_n - a| = 0$, then we find that the requirements are identical. Thus, $a_n \to a$ if and only if $|a_n - a| \to 0$.

The limit of a sequence (if it exists) is well defined – that is, a sequence cannot have more than one limit.

Theorem 2.1.6. *If $a_n \to a$ and $a_n \to b$, then $a = b$.*

Proof. If $a_n \to a$ and $a_n \to b$, then, for each $\epsilon > 0$ there are numbers N_1 and N_2 such that

$$n > N_1 \quad \text{implies} \quad |a_n - a| < \epsilon/2, \text{and}$$
$$n > N_2 \quad \text{implies} \quad |a_n - b| < \epsilon/2.$$

If n is an integer larger than both N_1 and N_2, then

$$|b - a| = |(a_n - a) + (b - a_n)| \le |a_n - a| + |b - a_n| < \epsilon/2 + \epsilon/2 = \epsilon.$$

This implies that $|b-a|$ is smaller than every positive number ϵ. Since $|b-a| \ge 0$, this is possible only if $|b - a| = 0$ – that is, only if $a = b$. (In this argument we used an important property of the real number system without comment. In Exercise 2.1.12 you are asked to figure out what property that is.) $\square$

Finding the limit of a sequence often involves two steps: (1) make a good intuitive guess as to what the limit should be, and (2) prove that your guess is correct by using the above definition or theorems that have been proved using it. The following example illustrates the first of these steps.

Example 2.1.7. Make an educated guess as to what the limits are for the following sequences.

(a) $\{1/n\}$;

(b) $\left\{ \dfrac{n}{2n+1} \right\}$;

(c) $\{(-1)^n\}$;

(d) $\{\sqrt{4+1/n}\}$.

Solution:(a) The larger n becomes, the smaller $1/n$ becomes. Thus, it appears that $\lim 1/n = 0$.

(b) If we divide the numerator and denominator of $\dfrac{n}{2n+1}$ by n, the result is $\dfrac{1}{2+1/n}$. If $1/n \to 0$, then it should be the case that $\dfrac{1}{2+1/n} \to 1/2$. Thus, we choose $1/2$ as our guess.

(c) Since the sequence $\{(-1)^n\}$ alternates between -1 and 1, it does not appear to converge to any one number. Thus, we guess that it does not converge.

(d) If $1/n \to 0$, then it should be the case that $\sqrt{4+1/n} \to \sqrt{4} = 2$. Thus, our guess is 2.

Example 2.1.8. Use the definition of limit to verify that the guesses in the preceding example are correct:

Solution: (a) Given $\epsilon > 0$, we must show that there is an N such that $n > N$ implies $1/n < \epsilon$. However, since $1/n < \epsilon$ if and only if $n > 1/\epsilon$, if we choose $N = 1/\epsilon$, then indeed, $n > N$ implies $1/n < \epsilon$.

(b) Given $\epsilon > 0$ we must show that there is an N such that

$$n > N \quad \text{implies} \quad \left| \frac{n}{2n+1} - 1/2 \right| < \epsilon.$$

Some work with the expression in absolute values shows us how to do this:

$$\left| \frac{n}{2n+1} - 1/2 \right| = \left| \frac{2n - 2n + 1}{4n+2} \right| = \frac{1}{4n+2} < \frac{1}{4n}.$$

Thus, $\left| \dfrac{n}{2n+1} - 1/2 \right| < \epsilon$ whenever $\dfrac{1}{4n} < \epsilon$ – that is, whenever $n > \dfrac{1}{4\epsilon}$. Thus, it suffices to choose $N = \dfrac{1}{4\epsilon}$.

(c) We will show that there is no number a which satisfies the definition of the statement $\lim(-1)^n = a$. Let a be any real number and choose $\epsilon = 1/2$. If $\lim(-1)^n = a$, then there must be an N such that

$$n > N \quad \text{implies} \quad |(-1)^n - a| < 1/2.$$

Since there are both even and odd integers $n > N$, this means that

$$|1 - a| < 1/2 \quad \text{and} \quad |-1 - a| < 1/2.$$

Then the triangle inequality (Theorem 2.1.2 (a)) implies

$$2 = |1 - a + 1 + a| \leq |1 - a| + |1 + a| = |1 - a| + |-1 - a| < 1/2 + 1/2 = 1.$$

Since it is not true that $2 < 1$, our assumption that $\lim(-1)^n = a$ must be false. Since this is the case no matter what real number we choose for a, we conclude that $\{(-1)^n\}$ has no limit. (Once again, as in the proof of Theorem 2.1.6, we used here, without comment, a special property of the real number system. Exercise 2.1.12 asks you to state what property that is.)

(d) Given $\epsilon > 0$, we must show there is an N such that

$$n > N \quad \text{implies} \quad |\sqrt{4 + 1/n} - 2| < \epsilon.$$

We simplify this problem by rationalizing the positive expression $\sqrt{4 + 1/n} - 2$:

$$
\begin{aligned}
|\sqrt{4 + 1/n} - 2| = \sqrt{4 + 1/n} - 2 &= \frac{(\sqrt{4 + 1/n} - 2)(\sqrt{4 + 1/n} + 2)}{\sqrt{4 + 1/n} + 2} \\
&= \frac{4 + 1/n - 4}{\sqrt{4 + 1/n} + 2} < \frac{1/n}{\sqrt{4} + 2} = \frac{1}{4n}.
\end{aligned}
\tag{2.1.2}
$$

Thus, if $N = 1/(4\epsilon)$, then $n > N$ implies $|\sqrt{4 + 1/n} - 2| < \epsilon$.

Exercise Set 2.1

1. Show that

   (a) if $|x - 5| < 1$, then x is a number greater than 4 and less than 6.;

   (b) if $|x - 3| < 1/2$ and $|y - 3| < 1/2$, then $|x - y| < 1$;

   (c) if $|x - a| < 1/2$ and $|y - b| < 1/2$, then $|x + y - (a + b)| < 1$.

2. Use the triangle inequality to prove that there is no number x which satisfies both $|x - 1| < 1/2$ and $|x - 2| < 1/2$.

3. Put each of the following sequences in the form $a_1, a_2, a_3, \cdots, a_n, \cdots$. This requires that you compute the first 3 terms and find an expression for the nth term.

   (a) the sequence of positive odd integers;

   (b) the sequence defined inductively by $a_1 = 1$ and $a_{n+1} = -\dfrac{a_n}{2}$;

   (c) the sequence defined inductively by $a_1 = 1$ and $a_{n+1} = \dfrac{a_n}{n + 1}$.

   In each of the next six exercises, first make an educated guess as to what you think the limit is, then use the definition of limit to prove that your guess is correct.

4. $\lim 1/n^2$.

5. $\lim \dfrac{2n-1}{3n+1}$.

6. $\lim(-1)^n/n$.

7. $\lim \dfrac{n}{n^3+4}$.

8. $\lim\{\sqrt{n+1}-\sqrt{n}\}$.

9. Prove that $\lim(1/n+(-1)^n/n^2)=0$.

10. Prove that $\lim 2^{-n}=0$. Hint: prove first that $2^n \geq n$ for all natural numbers n.

11. Prove that if $a_n \to 0$ and k is any constant, then $ka_n \to 0$.

12. In the proof of Theorem 2.1.6 we failed to point out that one step is true only because we are working in the real number system and not some other ordered field. What special property of the real number system makes this argument work? This same property is also used without comment in Example 2.1.8 (c).

2.2 Using the Definition of Limit

It is important that mathematics students become comfortable with the notion of limit of a sequence. Unfortunately, it is a difficult concept to grasp. Students almost always have difficulty with it at first and learn to understand it only through repeated exposure and extensive practice in its use. This section is designed to provide some of this practice.

Using Identities and Inequalities

In each of the following examples, we wish to show that a certain sequence $\{a_n\}$ has limit a. The strategy for doing this, in each case, is to use identities and inequalities on the expression $|a_n - a|$ until we can show that it is less than or equal to some much simpler expression in n that can clearly be made less than any given ϵ by choosing n large enough.

Example 2.2.1. Prove that $\lim \dfrac{n}{2n-3}=1/2$.

Solution: We have

$$\left|\frac{n}{2n-3}-1/2\right|=\left|\frac{2n-2n+3}{4n-6}\right|=\left|\frac{3}{4n-6}\right|.$$

Now $4n-6 = n+(3n-6) \geq n$ whenever $n > 1$. Thus,

$$\left|\frac{n}{2n-3}-1/2\right| \leq \frac{3}{4n-6} \leq \frac{3}{n}$$

provided $n > 1$. Given $\epsilon > 0$, if we choose $N = \max\{1, 3/\epsilon\}$, then

$$\left| \frac{n}{2n - 3} - 1/2 \right| \le \frac{3}{n} < \epsilon \quad \text{whenever} \quad n > N.$$

This completes the proof that $\lim \dfrac{n}{2n - 1} = 1/2$.

Example 2.2.2. Prove that $\lim(2 + 1/n)^2 = 4$.
    Solution: We have

$$|(2 + 1/n)^2 - 4| = |2 + 1/n + 2||2 + 1/n - 2| = \frac{4 + 1/n}{n} \le \frac{5}{n}.$$

Thus, given $\epsilon > 0$, if we set $N = 5/\epsilon$ we have

$$|(2 + 1/n)^2 - 4| \le \frac{5}{n} < \epsilon \quad \text{whenever} \quad n > N.$$

This proves that $\lim(2 + 1/n)^2 = 4$.

Using Information About a Limit

Knowing that a sequence converges or that it converges to a specific number always provides a great deal of other information. We give some examples below.

Theorem 2.2.3. *If* $\lim a_n = a$ *and* $a < c$, *then there exists an* N *such that*

$$a_n < c \quad \text{for all} \quad n > N.$$

Similarly, if $b < a$, *then there is an* N *such that*

$$b < a_n \quad \text{for all} \quad n > N.$$

Proof. If $a < c$, then $c - a > 0$. Since $\lim a_n = a$, for each $\epsilon > 0$, there is an N such that

$$|a_n - a| < \epsilon \quad \text{whenever} \quad n > N.$$

If we use this in the case where $\epsilon = c - a$ it tells us there is an N such that

$$|a_n - a| < c - a \quad \text{whenever} \quad n > N.$$

This implies

$$a - c + a < a_n < a + c - a \quad \text{whenever} \quad n > N,$$

by Theorem 2.1.1(b). Thus, $a_n < c$ for all $n > N$.
    The second statement of the theorem is proved in the same way. $\square$

    A sequence $\{a_n\}$ is bounded above (or below) if the set of numbers which appear as terms of $\{a_n\}$ is bounded above (or below) as a set of numbers. A sequence which is bounded above and bounded below is simply said to be bounded.
    The following corollary follows directly from the preceding theorem. We leave the details to the exercises.

Corollary 2.2.4. *If a sequence $\{a_n\}$ converges, then it is bounded.*

Theorem 2.2.5. *If $\{a_n\}$ is a sequence and $\lim a_n = a$, then $\lim |a_n| = |a|$.*

Proof. We use the second form of the triangle inequality (Theorem 2.1.2(b)) to write

$$||a_n| - |a|| \leq |a_n - a|. \tag{2.2.1}$$

Since $\lim a_n = a$, given $\epsilon > 0$, there is an N such that

$$|a_n - a| < \epsilon \quad \text{whenever} \quad n > N.$$

Then, by (2.2.1), it is also true that

$$||a_n| - |a|| < \epsilon \quad \text{whenever} \quad n > N,$$

Thus, $\lim |a_n| = |a|$. □

Example 2.2.6. For a sequence $\{a_n\}$ with $\lim a_n = a$, prove $\lim a_n^2 = a^2$.
  Solution: We first note that

$$|a_n^2 - a^2| = |a + a_n||a_n - a| \leq (|a_n| + |a|)|a_n - a| \tag{2.2.2}$$

We know that $\lim |a_n| = |a|$ by the previous theorem. Since $|a| < |a| + 1$, Theorem 2.2.3 implies that there is an N_1 such that $|a_n| < |a| + 1$ for all $n > N_1$. This and (2.2.2) together imply that

$$|a_n^2 - a^2| < (2|a| + 1)|a_n - a| \quad \text{whenever} \quad n > N_1.$$

Given $\epsilon > 0$ we choose N_2 such that $|a_n - a| < \dfrac{\epsilon}{2|a| + 1}$ whenever $n > N_2$. We can do this because $\lim a_n = a$. If we set $N = \max(N_1, N_2)$, then

$$|a_n^2 - a^2| < \epsilon \quad \text{whenever} \quad n > N.$$

Hence, $\lim a_n^2 = a^2$.

An Equivalent Definition of Limit

The following theorem rephrases the definition of limit in a way that may provide some additional insight.

Theorem 2.2.7. *A sequence $\{a_n\}$ converges to a if and only if, for each $\epsilon > 0$, there are only finitely many n for which $|a_n - a| \geq \epsilon$.*

Proof. Given $\epsilon > 0$, set

$$A_\epsilon = \{n \in \mathbb{N} : |a_n - a| \geq \epsilon\}.$$

If $\lim a_n = a$ and $\epsilon > 0$, there is an N such that $|a_n - a| < \epsilon$ whenever $n > N$. This means that A_ϵ is contained in the set $\{1, 2, \cdots, N\}$ and, hence, is finite.

  Conversely, suppose that, for each $\epsilon > 0$, the set A_ϵ is finite. Then given $\epsilon > 0$, the set A_ϵ has a largest element N. This means $n \notin A_\epsilon$ if $n > N$ – that is, $|a_n - a| < \epsilon$ if $n > N$. This implies that $\lim a_n = a$. □

Negating the Limit Definition

What does it mean for it not to be true that $\lim a_n = a$? That is, what is the *negation* of the statement "for each $\epsilon > 0$ there is an N such that $|a_n - a| < \epsilon$ whenever $n > N$" ? If it is not true that for each $\epsilon > 0$, there is an N such that $\cdots$, then for some $\epsilon > 0$, there is no N such that $\cdots$. If we fill in the dots we get the following statement:

The sequence $\{a_n\}$ does not converge to a if and only for some $\epsilon > 0$ there is no N such that $|a_n - a| < \epsilon$ for all $n > N$.

We may rephrase the second half of this statement to obtain:

The sequence $\{a_n\}$ does not converge to a if and only for some $\epsilon > 0$ and for every N there is an $n > N$ such that $|a_n - a| \geq \epsilon$.

Negating the equivalent definition of limit given in Theorem 2.2.7 yields a somewhat simpler statement:

The sequence $\{a_n\}$ does not converge to a if and only for some $\epsilon > 0$ there are infinitely many $n \in \mathbb{N}$ for which $|a_n - a| \geq \epsilon$.

Example 2.2.8. Show that the sequence $\{2^{-n} + (1 + (-1)^n)2^{-50}\}$ does not converge to 0.

Solution: Try computing a few terms of this sequence on a calculator. It appears to be converging to 0. However, if we choose $\epsilon = 2^{-49}$, then for every even $n \in N$

$$|2^{-n} + (1 + (-1)^n)2^{-50} - 0| = 2^{-n} + 2 \cdot 2^{-50} \geq 2^{-49}.$$

Since this inequality holds for infinitely many n, the sequence does not converge to 0.

Exercise Set 2.2

In each of the following six exercises, first make an educated guess as to what you think the limit is, then use the definition of limit to prove that your guess is correct.

1. $\lim \dfrac{3n^2 - 2}{n^2 + 1}$.

2. $\lim \dfrac{n}{n^2 + 2}$.

3. $\lim \dfrac{1}{\sqrt{n}}$.

4. $\lim \left(\dfrac{n}{n+1} \right)^2$.

5. $\lim(\sqrt{n^2+n}-n)$.

6. $\lim(1+1/n)^3$.

7. Prove Corollary 2.2.4.

8. Prove that if $\lim a_n = a$, then $\lim a_n^3 = a^3$;

9. Does the sequence $\{\cos(n\pi/3)\}$ have a limit? Justify your answer.

10. Give an example of a sequence $\{a_n\}$ which does not converge, but the sequence $\{|a_n|\}$ does converge.

11. Prove that if $\{a_n\}$ and $\{b_n\}$ are sequences with $|a_n| \le b_n$ for all n and if $\lim b_n = 0$, then $\lim a_n = 0$ also.

12. Prove the following partial converse to Theorem 2.2.3: Suppose $\{a_n\}$ is a convergent sequence. If there is an N such that $a_n \le c$ for all $n > N$, then $\lim a_n \le c$. Also, if there is an N such that $b \le a_n$ for all $n > N$, then $b \le \lim a_n$.

2.3   Limit Theorems

We reiterate that the strategy to use in proving a statement of the form

$$\lim a_n = a$$

directly from the definition is to use a string of identities and inequalities to conclude that $|a_n - a|$ is less than or equal to a simpler expression in n that we can easily force to be less than ϵ by making n sufficiently large. This strategy was used throughout the previous two sections. The following theorem formalizes this strategy in a way that will lead us to use the right approach to many limit proofs.

Theorem 2.3.1. *Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers and suppose $\lim b_n = 0$. If $a \in \mathbb{R}$ and there is an N_1 such that*

$$|a_n - a| \le b_n \quad \textit{for all} \quad n > N_1, \tag{2.3.1}$$

then $\lim a_n = a$.

Proof. Since $\lim b_n = 0$, given any $\epsilon > 0$, there is an N_2 such that

$$b_n = |b_n - 0| < \epsilon \quad \text{whenever} \quad n > N_2.$$

It now follows from (2.3.1) that

$$|a_n - a| < \epsilon \quad \text{whenever} \quad n > N = \max\{N_1, N_2\}.$$

Thus, $\lim a_n = a$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Of course, to prove that $\lim a_n = a$ using this theorem one must establish an inequality of the form (2.3.1), where $\{b_n\}$ is a sequence of non-negative terms that we know converges to 0. The proof of the next theorem uses this technique. The proof is easy and is left to the exercises.

A sequence $\{b_n\}$ for which there is a number k such that $b_n \leq k$ for all n is said to be *bounded above*. If there is a number m such that $m \leq b_n$ for all n, then the sequence is said to be *bounded below*. A sequence which is bounded above and below is simply said to be *bounded*. Note that a sequence $\{b_n\}$ is bounded if and only if $\{|b_n|\}$ is bounded above (Exercise 2.3.6). Recall from Corollary 2.2.4 that convergent sequences are bounded.

Theorem 2.3.2. *Let $\{a_n\}$ be a sequence of real numbers such that $\lim a_n = 0$, and let $\{b_n\}$ be a bounded sequence. Then $\lim a_n b_n = 0$.*

The following theorem is often called the *squeeze principle*.

Theorem 2.3.3. *If $\{a_n\}$, $\{b_n\}$, and $\{c_n\}$ are sequences for which there is a number K such that*

$$b_n \leq a_n \leq c_n \quad \text{for all} \quad n > K,$$

and if $b_n \to a$ and $c_n \to a$, then $a_n \to a$.

Proof. Since $b_n \to a$ and $c_n \to a$, given $\epsilon > 0$ there are numbers N_1 and N_2 such that

$$
\begin{aligned}
a - \epsilon < b_n < a + \epsilon \quad &\text{for all} \quad n > N_1; and \\
a - \epsilon < c_n < a + \epsilon \quad &\text{for all} \quad n > N_2.
\end{aligned}
\tag{2.3.2}
$$

Then for $n > N = \max\{N_1, N_2, K\}$ we have

$$a - \epsilon < b_n \leq a_n \leq c_n < a + \epsilon.$$

This implies $|a_n - a| < \epsilon$. Thus, $\lim a_n = a$. $\square$

Example 2.3.4. Prove that if $\{a_n\}$ is a sequence of positive numbers converging to a positive number a, then $\lim \sqrt{a_n} = \sqrt{a}$.

Solution: We will use Theorem 2.3.1. Rationalizing the numerator gives us

$$|\sqrt{a_n} - \sqrt{a}| = \frac{|a_n - a|}{\sqrt{a_n} + \sqrt{a}} < \frac{1}{\sqrt{a}}|a_n - a|.$$

Since $a_n \to a$, Remark 2.1.5 implies $|a_n - a| \to 0$. Then Theorems 2.3.2 and 2.3.1 imply $\sqrt{a_n} \to \sqrt{a}$.

Example 2.3.5. Prove that if $|a| < 1$, then $\lim a^n = 0$.

Solution: The result is trivial in the case $a = 0$. If $a \neq 0$, we set $b = |a|^{-1} - 1$. Then $b > 0$ and $|a|^{-1} = 1 + b$. We use the Binomial Theorem (Theorem 1.2.12) to expand $|a|^{-n} = (1 + b)^n$:

$$(1 + b)^n = 1 + nb + \frac{n(n-1)}{2}b^2 + \cdots + b^n.$$

Since all the terms involved are positive, it follows that $|a|^{-n} = (1+b)^n \geq nb$. Inverting this yields

$$|a^n| \leq \frac{1}{nb} = \frac{1}{b}\frac{1}{n}$$

Since $1/n \to 0$, it follows from Theorems 2.3.2 and 2.3.1 that $a^n \to 0$.

The Main Limit Theorem

This is the theorem that tells us that the limit concept behaves well with regard to the usual algebraic operations.

Theorem 2.3.6. *Suppose $a_n \to a$, $b_n \to b$, c is a real number, and k is a natural number. Then*

(a) $ca_n \to ca$;

(b) $a_n + b_n \to a + b$;

(c) $a_n b_n \to ab$;

(d) $a_n/b_n \to a/b$ if $b \neq 0$ and $b_n \neq 0$ for all n;

(e) $a_n^k \to a^k$;

(f) $a_n^{1/k} \to a^{1/k}$ if $a_n \geq 0$ for all n;

Proof. Part (a) follows immediately from Theorem 2.3.2 applied to the sequence $\{c(a_n - a)\}$. We will prove (c), and (e) and leave (b), (d), and (f) to the exercises.

(c) We use the strategy suggested by Theorem 2.3.1. We have

$$|a_n b_n - ab| = |a_n b_n - ab_n + ab_n - ab| \leq |a_n - a||b_n| + |a||b_n - b|,$$

by the triangle inequality. Furthermore, we have $\{b_n\}$ is bounded by 2.2.4, and so $\{|b_n|\}$ is bounded above. We also have $|a_n - a| \to 0$, by Remark 2.1.5. Therefore, by Theorem 2.3.2, $|a_n - a||b_n| \to 0$. By Part (a), $|a||b_n - b| \to 0$. By Part (b) the sum $|a_n - a||b_n| + |a||b_n - b|$ converges to 0 and, hence, $a_n b_n \to ab$ by Theorem 2.3.1.

(e) We use the identity

$$a_n^k - a^k = (a_n - a)(a_n^{k-1} + a_n^{k-2}a + a_n^{k-3}a^2 + \cdots + a^{k-1}) = (a_n - a)b_n,$$

where

$$b_n = a_n^{k-1} + a_n^{k-2}a + a_n^{k-3}a^2 + \cdots + a^{k-1}.$$

Now, because the sequence $\{a_n\}$ converges, it is bounded and, hence, $\{|a_n|\}$ is bounded above. We choose an upper bound m for $\{|a_n|\}$ which also satisfies $|a| \leq m$. Then

$$|b_n| \leq km^k.$$

Since k and m are fixed, the sequence $\{|b_n|\}$ is bounded above.

We conclude from Theorem 2.3.2 that $|a_n - a||b_n| \to 0$ and from Theorem 2.3.1 that $a_n^k \to a^k$. $\square$

Example 2.3.7. Use the main limit theorem to find $\lim \dfrac{n^2 + 3n + 1}{3n^2 - 7n + 2}$.

   Solution: In a problem of this type, we divide the numerator and denominator by the highest power of n that appears in either one. In this case, that is the second power. The result is

$$\frac{1 + 3/n + 1/n^2}{3 - 7/n + 2/n^2}.$$

The main limit theorem then tells us that

$$\begin{aligned}
\lim \frac{1 + 3/n + 1/n^2}{3 - 7/n + 2/n^2} &= \frac{\lim(1 + 3(1/n) + 2(1/n)^2)}{\lim(3 - 7(1/n) + 2(1/n)^2)} \\
&= \frac{1 + 3\lim(1/n) + 2\lim(1/n)^2}{3 - 7\lim(1/n) + 2\lim(1/n)^2} = \frac{1 + 3\lim(1/n) + 2(\lim 1/n)^2}{3 - 7\lim(1/n) + 2(\lim 1/n)^2} \quad\quad (2.3.3) \\
&= \frac{1 + 3 \cdot 0 + 2(0)^2}{3 - 7 \cdot 0 + 2(0)^2} = 1/3.
\end{aligned}$$

Here, we didn't explicitly refer to the parts of the Main Limit Theorem as we used them, but it is clear that the first equality uses (d), the second (a) and (b), the third (e), and the fourth the fact that $\lim 1/n = 0$ (Example 2.1.8).

Theorem 2.3.8. *If $\{a_n\} \to a$ and $\{b_n \to b\}$ are convergent sequences and if there is a number K such that $a_n \le b_n$ whenever $n > K$, then $a \le b$.*

Proof. The sequence $c_n = b_n - a_n$ is a sequence with $b - a$ as its limit and with terms that are non-negative for $n > K$. If $b - a$ were negative, then Theorem 2.2.3 would imply $b_n - a_n < 0$ for all sufficiently large n. Since this is not the case, we conclude that $a \le b$.   □

Exercise Set 2.3

1. Use the Main Limit Theorem to find $\lim \dfrac{2n^3 - n + 1}{3n^3 + n^2 + 6}$.

2. Use the Main Limit Theorem to find $\lim \dfrac{n^2 - 5}{n^3 + 2n^2 + 5}$.

3. Use the Main Limit Theorem to find $\lim \dfrac{2^n}{2^n + 1}$.

4. Prove that $\lim \dfrac{\sin n}{n} = 0$.

5. Prove Theorem 2.3.2.

6. Prove that a sequence $\{a_n\}$ is both bounded above and bounded below if and only if its sequence of absolute values $\{|a_n|\}$ is bounded above.

7. Prove Part (b) of Theorem 2.3.6.

8. Prove that if $\{b_n\}$ is a sequence of positive terms and $b_n \to b > 0$, then there is a number $m > 0$ such that $b_n \geq m$ for all n.

9. Prove Part (d) of Theorem 2.3.6. Hint: use the previous exercise.

10. Prove Part (f) of Theorem 2.3.6. Hint: use the identity

$$x^k - y^k = (x - y)(x^{k-1} + x^{k-2}y + \cdots + y^{k-1})$$

with $x = a_n^{1/k}$ and $y = a^{1/k}$.

11. For each natural number n, let $b_n = n^{1/n} - 1$. Then b_n is positive and $n = (1 + b_n)^n$. Use the Binomial Theorem (Theorem 1.2.12) to prove that $n \geq \dfrac{n(n-1)}{2} b_n^2$ and, hence, that $b_n \leq \sqrt{\dfrac{2}{n-1}}$.

12. Prove that $\lim n^{1/n} = 1$. Hint: use the result of the previous exercise.

13. Prove that if $a > 0$, then $\lim a^{1/n} = 1$. Hint: do this first for $a \geq 1$; use the result of the previous exercise and the squeeze principle.

2.4   Monotone Sequences

A sequence of real numbers $\{a_n\}$ is said to be *non-decreasing* if $a_{n+1} \geq a_n$ for each n. The sequence is said to be *non-increasing* if $a_{n+1} \leq a_n$ for each n. If it is one or the other (either non-decreasing or non-increasing), the sequence is said to be *monotone*.

Convergence of Monotone Sequences

In this section and the next, we will develop powerful tools for proving that a sequence converges. These tools work even in situations where we have no idea what the limit might be. It is the completeness axiom for the real number system that makes these results possible.

Theorem 2.4.1. (Monotone Convergence Theorem) *Each bounded monotone sequence converges.*

Proof. A non-decreasing sequence $\{a_n\}$ is bounded if and only if it is bounded above, since it is automatically bounded below by a_1. Similarly, a non-increasing sequence is bounded if and only if it is bounded below.

We will prove that every non-decreasing sequence that is bounded above converges. The proof that every non-increasing sequence that is bounded below converges is the same but with all the inequalities reversed.

Thus, suppose $\{a_n\}$ is non-decreasing and bounded above. Then the set

$$A = \{a_n : n \in \mathbb{N}\}$$

is a non-empty set which is bounded above. By the completeness axiom **C**, this set has a least upper bound a. That is,

$$\sup_n a_n = \sup A = a$$

is finite. We will show that a is the limit of the sequence $\{a_n\}$.

Given $\epsilon > 0$, the number $a - \epsilon$ is less than a and so it is not an upper bound for A. This means there is some natural number N such that $a - \epsilon < a_N$. If $n > N$, then $a_N \leq a_n$ since $\{a_n\}$ is a non-decreasing sequence. This implies $a - \epsilon < a_n$. We also have $a_n \leq a < a + \epsilon$, since a is an upper bound for $\{a_n\}$. Combining these inequalities yields

$$a - \epsilon < a_n < a + \epsilon \quad \text{for all} \quad n > N.$$

By Theorem 2.1.1(b), this is equivalent to

$$|a_n - a| < \epsilon \quad \text{for all} \quad n > N.$$

We conclude that $\lim a_n = a$.   □

Example 2.4.2. Let a sequence be defined inductively by $a_1 = 0$ and

$$a_{n+1} = \frac{a_n + 1}{2}. \tag{2.4.1}$$

Prove that this sequence converges and find its limit.

Solution: This is a non-decreasing sequence (Exercise 1.2.13). Also, a simple induction argument shows that it is bounded above by 1. Therefore it is a bounded monotone sequence, and it converges by the previous theorem. Let $\lim a_n = a$. If we take the limit of both sides of (2.4.1), the result is $a = (a+1)/2$, or $a/2 = 1/2$. Thus, $a = 1$.

A less trivial example is the following:

Example 2.4.3. Let a sequence $\{a_n\}$ be defined inductively by $a_1 = 2$ and

$$a_{n+1} = \frac{a_n^2 + 2}{2a_n}. \tag{2.4.2}$$

Prove that this sequence converges and then find its limit.

Solution: We first note that a trivial induction argument shows that $a_n > 0$ for all n. This is true when $n = 1$ and true for $n + 1$ whenever it is true for n by (2.4.2).

We will prove that the sequence is non-increasing. To show that $a_{n+1} \leq a_n$, we must show that $\dfrac{a_n^2 + 2}{2a_n} \leq a_n$. If we assume that $a_n > 0$, then we may multiply this inequality by $2a_n$ to obtain the equivalent inequality

$$a_n^2 + 2 \leq 2a_n^2 \quad \text{or} \quad a_n^2 \geq 2.$$

We conclude that $a_{n+1} \le a_n$ as long as a_n is positive and $a_n^2 \ge 2$ – that is, as long as $a_n \ge \sqrt{2}$. Now $a_1 = 2$ and so the sequence starts out with a number greater than or equal to $\sqrt{2}$. Every other number in this sequence has the form

$$\frac{x^2 + 2}{2x}.$$

for some positive x. We claim every such number is greater than or equal to $\sqrt{2}$. In fact

$$0 \le (x - \sqrt{2})^2 = x^2 - 2\sqrt{2}\,x + 2, \quad \text{and so} \quad 2\sqrt{2}\,x \le x^2 + 2.$$

This implies $\sqrt{2} \le \dfrac{x^2 + 2}{2x}$. Thus every a_n is greater than or equal to $\sqrt{2}$.

We now know that the sequence $\{a_n\}$ is non-increasing and bounded below by $\sqrt{2}$. Thus, it is a bounded monotone sequence and has a limit by the previous theorem. Call the limit a. By (2.4.2), we have

$$2a_n a_{n+1} = a_n^2 + 2.$$

If we take the limit of both sides of this equation and note that $\lim a_n = \lim a_{n+1} = a$, then the result is

$$2a^2 = a^2 + 2 \quad \text{or} \quad a^2 = 2.$$

Thus, $a = \sqrt{2}$.

Infinite Limits

Definition 2.4.4. If $\{a_n\}$ is a sequence of real numbers, then $\lim a_n = \infty$ if, for every real number M, there is a number N such that

$$a_n > M \quad \text{whenever} \quad n > N.$$

Similarly, we say $\lim a_n = -\infty$ if for every real number M there is an N such that

$$a_n < M \quad \text{whenever} \quad n > N.$$

Example 2.4.5. If $r > 0$ prove that $\lim n^r = \infty$.

Solution: To prove that $\lim n^r = \infty$ we must show that for every M there is an N such that

$$n^r > M \quad \text{whenever} \quad n > N.$$

Clearly, we need only choose N to be $M^{1/r}$.

With $+\infty$ and $-\infty$ as possible limits of a sequence, we can now assert that:

Theorem 2.4.6. *Every monotone sequence has a limit.*

The proof of this is left to the exercises.

Note that we must now make a distinction between a sequence *converging* and a sequence *having a limit*. A sequence may have a limit which is infinite, but a sequence which converges must have a finite limit.

Theorem 2.4.7. *Let $\{a_n\}$ and $\{b_n\}$ be sequences of real numbers. Then*

(a) *if $a_n > 0$ for all n, then $\lim a_n = \infty$ if and only if $\lim 1/a_n = 0$;*

(b) *if $\{b_n\}$ is bounded below, then $\lim a_n = \infty$ implies $\lim(a_n + b_n) = \infty$.*

(c) *$\lim a_n = \infty$ if and only if $\lim(-a_n) = -\infty$;*

(d) *if $a_n \leq b_n$ for all n, then $\lim a_n = \infty$ implies $\lim b_n = \infty$;*

(e) *if there is a positive constant k such that $k \leq b_n$ for all n, then $\lim a_n = \infty$ implies $\lim a_n b_n = \infty$;*

Proof. We will prove (a) and (b) and leave (c), (d), and (e) to the exercises.

(a) If we are given an ϵ, we will set $M = 1/\epsilon$. Conversely, if we are given an M, we will set $\epsilon = 1/M$. Then the statements

$$|1/a_n| < \epsilon \quad \text{and} \quad a_n > M$$

mean the same thing (since a_n is positive) so that, if there is an N such that one of these statements is true for all $n > N$ then the other statement is also true for all $n > N$. Thus, $\lim 1/a_n = 0$ if and only if $\lim a_n = \infty$.

(b) Let b_n be bounded below by, say, K. Assuming $\lim a_n = \infty$, we wish to show that $\lim(a_n + b_n) = \infty$. Given $M \in \mathbb{R}$, the number $M - K$ is also in $\mathbb{R}$ and so, by our assumption that $\lim a_n = \infty$, we know there is an N such that

$$a_n > M - K \quad \text{whenever} \quad n > N.$$

then

$$a_n + b_n > M - K + K = M \quad \text{whenever} \quad n > N.$$

Thus, $\lim(a_n + b_n) = \infty$. $\square$

Example 2.4.8. Find the following limits:

(a) $\lim \dfrac{2n^2 + 3}{n + 1}$;

(b) $\lim a^n$ for $a > 1$;

(c) $\lim(\sqrt{n} + (-1)^n)$.

Solution: (a) We factor the largest power of n that occurs out of each of the denominator and the numerator. The result is

$$\frac{2n^2 + 3}{n + 1} = \frac{n^2(2 + 3/n^2)}{n(1 + 1/n)} = n\frac{2 + 3/n^2}{1 + 1/n}.$$

Now $\lim n = \infty$ and $\dfrac{2 + 3/n^2}{1 + 1/n} \geq 1$ for all n. Thus,

$$\lim \frac{2n^2 + 3}{n + 1} = \infty,$$

by Theorem 2.4.7 (e).

(b) Since $|1/a| < 1$, it follows from Example 2.3.5 that $\lim 1/a^n = 0$. Then $\lim a^n = +\infty$ by Theorem 2.4.7(a). Another proof of this fact is suggested in Exercise 2.4.7.

(c) Since $\sqrt{n} = n^{1/2}$, Example 2.4.5 implies that $\lim \sqrt{n} = \infty$. Then Theorem 2.4.7 (b) implies that $\lim(\sqrt{n} + (-1)^n) = \infty$.

Exercise Set 2.4

1. Tell which of the following sequences are non-increasing, non-decreasing, bounded? Justify your answers.

   (a) $\{n^2\}$;

   (b) $\left\{\dfrac{1}{\sqrt{n}}\right\}$;

   (c) $\left\{\dfrac{(-1)^n}{n}\right\}$;

   (d) $\left\{\dfrac{n}{2^n}\right\}$;

   (e) $\left\{\dfrac{n}{n + 1}\right\}$.

2. Prove that the sequence of Example 1.2.11 converges and decide what number it converges to.

3. If $a_1 = 1$ and $a_{n+1} = (1 - 2^{-n})a_n$, prove that $\{a_n\}$ converges.

4. Let $\{d_n\}$ be a sequence of 0's and 1's and define a sequence of numbers $\{a_n\}$ by
   $$a_n = d_1 2^{-1} + d_2 2^{-2} + \cdots + d_n 2^{-n}.$$
   Prove that this sequence converges to a number between 0 and 1.

5. Let $\{s_n\}$ be the sequence of partial sums of a series with positive terms. That is,
   $$s_n = \sum_{k=1}^{n} a_k \quad \text{with all} \quad a_k \geq 0.$$
   Prove that $\lim s_n$ exists (though it may not be finite).

6. Give an alternate proof to the result of Example 2.3.5 that does not use the Binomial theorem. Instead, first show that $\{|a|^n\}$ is a non-increasing sequence. Then show that 0 is the only possible value for the limit.

7. Give an alternate proof of the result of Example 2.4.8(b) that does not use Example 2.3.5. Use the method of the previous exercise.

8. Prove that $\lim \dfrac{n^5 + 3n^3 + 2}{n^4 - n + 1} = \infty$.

9. Prove that $\lim \dfrac{2^n}{n} = \infty$.

10. Prove Theorem 2.4.6.

11. Prove Part (c) of Theorem 2.4.7.

12. Prove Part (d) of Theorem 2.4.7.

13. Prove Part (e) of Theorem 2.4.7.

14. Suppose $\{a_n\}$ and $\{b_n\}$ are non-decreasing sequences that are interlaced in the sense that each term of the sequence $\{a_n\}$ is less than or equal to some term of the sequence $\{b_n\}$ and vice-versa. Prove that $\lim a_n = \lim b_n$.

2.5   Cauchy Sequences

In this section we will prove two of the most important theorems about convergence of sequences. The proofs are based on the nested interval property, which we describe below.

Nested Intervals

A *nested sequence of closed bounded intervals* is a sequence

$$I_1 \supset I_2 \supset I_3 \supset \cdots$$

in which each I_n is a closed bounded interval, and each interval in the sequence contains the next one. Thus, each of the intervals I_n has the form $[a_n, b_n]$ for real numbers $a_n < b_n$. The nested condition means that $I_n \supset I_{n+1}$ for each n – that is,

$$a_n \leq a_{n+1} < b_{n+1} \leq b_n$$

for each n.

Theorem 2.5.1. (Nested Interval Property) *If $I_1 \supset I_2 \supset I_3 \supset \cdots$ is a nested sequence of closed bounded intervals, then $\bigcap_n I_n \neq \emptyset$. That is, there is at least one point x that is in all the intervals I_n.*

Proof. Let $I_n = [a_n, b_n]$, as above. Then the sequence $\{a_n\}$ of left endpoints is a non-decreasing sequence which is bounded above (by b_1), and the sequence $\{b_n\}$ of right endpoints is a non-increasing sequence which is bounded below (by a_1). The Monotone Convergence Theorem (2.4.1) implies that both sequences converge.

If $a = \lim a_n$ and $b = \lim b_n$, then $a \leq b$ by Theorem 2.3.8. In fact,

$$a_n \leq a \leq b \leq b_n$$

for each n. This means that $[a, b] \subset I_n$ for every n and, hence, that $[a, b] \subset \bigcap_n I_n$.

The set $[a, b]$ is a closed interval if $a < b$ and a single point if $a = b$. In either case, it is non-empty. $\square$

We leave to the exercises the problem of showing that this theorem is false if we don't insist that the intervals are closed or if we don't insist that they are bounded.

The Bolzano-Weierstrass Theorem

A sequence $\{b_k\}$ is a *subsequence* of the sequence $\{a_n\}$ if it is made up of some of the terms of $\{a_n\}$, taken in the order that they appear in $\{a_n\}$. More precisely:

Definition 2.5.2. A sequence $\{b_k\}$ is a *subsequence* of the sequence $\{a_n\}$ if there is a strictly increasing sequence of natural numbers $\{n_k\}$ such that $b_k = a_{n_k}$.

Example 2.5.3. Give three examples of subsequences of the sequence

$$0, \ 3/2, \ -2/3, \ 5/4, \ -4/5, \ 7/6, \ -6/7, \ 9/8 \cdots, (-1)^n + 1/n, \cdots.$$

Does the original sequence converge? How about the three subsequences?

Solution:

(a) $3/2, \ 5/4, \ 7/6, \cdots, 1 + 1/(2k), \cdots$;

(b) $0, \ -2/3, \ -4/5, \cdots, -1 + 1/(2k - 1), \cdots$;

(c) $3/2, \ 5/4, \ 9/8, \cdots, 1 + 1/2^k, \cdots$.

The original sequence clearly does not converge, but sequence (a) converges to 1, (b) converges to -1 and (c) converges to 1.

Theorem 2.5.4. *If $\{a_n\}$ has a limit (possibly infinite), then each of its subsequences has the same limit.*

Proof. We will prove this in the case of a finite limit, the other cases are similar and are covered in the exercises.

Suppose $\{a_{n_k}\}$ is a subsequence of $\{a_n\}$. Then $\{n_k\}$ is an increasing sequence of natural numbers, and this implies that $n_k \geq k$ for all k (Exercise 2.5.4).

Now suppose $\lim a_n = a$. Given $\epsilon > 0$, there is an N such that

$$|a_n - a| < \epsilon \quad \text{whenever} \quad n > N.$$

Then $k > N$ implies $n_k > N$, since $n_k \geq k$. Thus,

$$|a_{n_k} - a| < \epsilon \quad \text{whenever} \quad k > N.$$

By definition, this means that $\lim a_{n_k} = a$. $\square$

Theorem 2.5.5. (Bolzano-Weierstrass Theorem) *Every bounded sequence of real numbers has a convergent subsequence.*

Proof. If $\{a_n\}$ is a bounded sequence, then it has an upper bound M and a lower bound m. This means that every a_n is contained in the interval $I_1 = [m, M]$. We will construct a nested sequence of closed bounded intervals

$$I_1 \supset I_2 \supset I_3 \supset \cdots \tag{2.5.1}$$

such that I_k contains infinitely many of the terms of $\{a_n\}$ for each k and I_{k+1} is either the left or the right half of the interval I_k, for each n. We do this by induction.

Certainly I_1 contains infinitely many terms of $\{a_n\}$ – in fact, it contains all of them. Suppose $I_1 \supset I_2 \supset I \cdots \supset I_k$ can be chosen with the required properties. Then we cut I_k into two closed intervals by dividing it at its midpoint. One of the two halves must contain infinitely many terms of $\{a_n\}$ since I_k does. Let I_{k+1} be the right half if it has this property; otherwise let it be the left half. This shows that a nested sequence of $k + 1$ intervals with the required properties can be chosen provide one with k terms can be chosen. By induction, there exists an infinite sequence 2.5.1 with the required properties.

By the Nested Interval Theorem, there is a point a that is in every one of the intervals I_k. Also, each interval I_k contains infinitely many terms of the sequence $\{a_n\}$. We will inductively define a subsequence $\{a_{n_k}\}$ of $\{a_n\}$ with the property that $a_{n_k} \in I_k$ for each k. We choose $n_1 = 1$ and define n_{k+1} in terms of n_k by the rule that n_{k+1} is the first integer greater than n_k such that $a_{n_{k+1}} \in I_{k+1}$. This is the basis for an inductive definition of the sequence we seek. Once this sequence of integers has been chosen, then $\{a_{n_k}\}$ is a subsequence of $\{a_n\}$. We will show that this subsequence converges to a.

For each k, a and a_{n_k} both belong to I_k. This means the distance between them can be no greater than the length of I_k, which is $(M - m)2^{1-k}$. That is,

$$|a_{n_k} - a| \le \frac{M - m}{2^{k-1}}.$$

Since $\dfrac{M - m}{2^{k-1}} \to 0$, Theorem 2.3.1 implies that $\lim a_{n_k} = a$.   □

Example 2.5.6. Construct a sequence $\{a_n\}$ as follows: for each n let a_n be the number obtained by replacing by 0 all digits to the left of the decimal point in the decimal expansion of $10^n \pi$. Does this sequence have a convergent subsequence?

Solution: This is a crazy sequence and it certainly does not appear to converge. However, each number in this sequence lies between 0 and 1 and so it is a bounded sequence. By the Bolzano-Weierstrass Theorem it has a convergent subsequence.

Cauchy Sequences

Definition 2.5.7. A sequence $\{a_n\}$ is said to be a *Cauchy Sequence* if, for every $\epsilon > 0$, there is an N such that

$$|a_n - a_m| < \epsilon \quad \text{whenever} \quad n, m > N.$$

Intuitively, this means we can make the terms of the sequence arbitrarily close to each other by going far enough out in the sequence. It is by no means obvious that this means that the sequence converges, but it does.

Theorem 2.5.8. *A sequence of real numbers $\{a_n\}$ is a Cauchy sequence if and only if it converges.*

Proof. There are two things to prove here – the "if" and the "only if". First we do the "if" – that is, we will prove that a sequence is Cauchy if it converges.

Assume $\{a_n\}$ converges to a number a. Then, given $\epsilon > 0$ there is an N such that
$$|a_n - a| < \epsilon/2 \quad \text{whenever} \quad n > N.$$
If $n, m > N$, then

$$|a_n - a_m| = |a_n - a + a - a_m| \le |a_n - a| + |a_m - a| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Therefore, $\{a_n\}$ is Cauchy.

Now for the "only if". Suppose $\{a_n\}$ is Cauchy. We first prove that $\{a_n\}$ is bounded. In fact, there is an N such that

$$|a_n - a_m| < 1 \quad \text{whenever} \quad n, m > N.$$

In particular, $|a_n - a_{N+1}| < 1$ for all $n > N$. This implies that

$$a_{N+1} - 1 < a_n < a_{N+1} + 1 \quad \text{whenever} \quad n > N.$$

Then $\max\{a_1, \cdots, a_N, a_{N+1}+1\}$ is an upper bound for $\{a_n\}$. Similarly, we have $\min\{a_1, \cdots, a_N, a_{N+1}-1\}$ is a lower bound for $\{a_n\}$. Thus, $\{a_n\}$ is a bounded sequence.

We next use the Bolzano-Weierstrass Theorem to conclude there is a subsequence $\{a_{n_k}\}$ of $\{a_n\}$ which converges to a number a. Finally, we use the definition of Cauchy sequence and what it means for a_{n_k} to converge to a. Given $\epsilon > 0$, there are numbers N_1 and N_1 such that

$$|a_n - a_m| < \epsilon/2 \quad \text{whenever} \quad n > N_1,$$

and,

$$|a_{n_k} - a| < \epsilon/2 \quad \text{whenever} \quad k > N_2.$$

If $n > N_1$ and we choose a $k > \max\{N_1, N_2\}$, then

$$|a_n - a| = |a_n - a_{n_k} + a_{n_k} - a| \le |a_n - a_{n_k}| + |a_{n_k} - a| < \epsilon/2 + \epsilon/2 = \epsilon.$$

This completes the proof that every Cauchy sequence is convergent. $\qquad\square$

Example 2.5.9. Show that the sequence $\{s_n\}$ of partial sums of the series $\sum_{k=1}^{\infty}(-1)^k \dfrac{k}{4^k}$ converges.

Solution: We have $s_n = \sum_{k=1}^{n}(-1)^k \dfrac{k}{4^k}$ and so, for $m > n$,

$$|s_m - s_n| = \left| \sum_{k=n+1}^{m}(-1)^k \frac{k}{4^k} \right| \leq \sum_{k=n+1}^{m}\frac{1}{2^k} \leq \frac{1}{2^{n+1}}\sum_{k=0}^{\infty}\frac{1}{2^k} = \frac{1}{2^n}.$$

Here we have used the fact that $k \leq 2^k$ for all k and the fact that the geometric series $\sum_{k=0}^{\infty}2^{-k}$ has sum $\dfrac{1}{1-1/2} = 2$.

Since $\lim 1/2^n = 0$, by Example 2.3.5, given $\epsilon > 0$, there is an N such that $n > N$ implies $1/2^n < \epsilon$. Then $|s_m - s_n| < \epsilon$ for all n, m with $m > n > N$. This means that $\{s_n\}$ is Cauchy and, hence, converges.

Exercise Set 2.5

1. Give an example of a nested sequence of bounded open intervals that does not have a point in its intersection.

2. Give an example of a nested sequence of closed but unbounded intervals which does not have a point in its intersection.

3. Prove that if I is a closed, bounded interval which is contained in the union of some collection of open intervals, then I is contained in the union of some finite subcollection of these open intervals.

4. Prove by induction that if $\{n_k\}$ is an increasing sequence of natural numbers, then $n_k \geq k$ for all k.

5. Which of the following sequences $\{a_n\}$ have a convergent subsequence? Justify your answer.

   (a) $a_n = (-2)^n$;
   (b) $a_n = \dfrac{5 + (-1)^n n}{2 + 3n}$;
   (c) $a_n = 2^{(-1)^n}$

6. For each of the following sequences $\{a_n\}$, find a subsequence which converges. Justify your answer.

   (a) $a_n = (-1)^n$;
   (b) $a_n = \sin n\pi/4$;
   (c) $a_n = \dfrac{n}{2^k} - 1$ with k an integer chosen so that $2^k \leq n < 2^{k+1}$.

7. For each of the following sequences, determine how many different limits of subsequences there are. Justify your answer.

   (a) $\{1 + (-1)^n\}$;
   (b) $\{\cos n\pi/3\}$;
   (c) $1, 1/2, 1, 1/2, 1/3, 1, 1/2, 1/3, 1/4, 1, 1/2, 1/3, 1/4, 1/5, \cdots$;

8. Does the sequence $\sin n$ have a convergent subsequence? Why?

9. Prove that a sequence which satisfies $|a_{n+1} - a_n| < 2^{-n}$ for all n is a Cauchy sequence.

10. Suppose a sequence $\{a_n\}$ has the property that for every $\epsilon > 0$, there is an N such that

$$|a_{n+1} - a_n| < \epsilon \quad \text{whenever} \quad n > N.$$

   Is $\{a_n\}$ necessarily Cauchy? Prove it is or give an example where it is not.

11. Let $s_n = \sum_{k=1}^{n} \frac{1}{k2^k}$ be the sequence of partial sums of the series $\sum_{k=1}^{\infty} \frac{1}{k2^k}$. Prove that $\{s_n\}$ converges. Hint: show that it is a Cauchy sequence.

12. Given a series $\sum_{k=1}^{\infty} a_k$, set $s_n = \sum_{k=1}^{n} a_k$ and $t_n = \sum_{k=1}^{n} |a_k|$. Prove that $\{s_n\}$ converges if $\{t_n\}$ is bounded.

2.6   lim inf and lim sup

A bounded sequence has a convergent subsequence according to the Bolzano-Weierstrass Theorem. In fact, a bounded sequence has many convergent subsequences and these may converge to many different limits, as is illustrated by some of the exercises in the previous section. Here we will show that there is a smallest closed interval that contains all of these limits. The endpoints of this interval are the lim inf and the lim sup of the sequence.

Given a sequence $\{a_n\}$, we construct two monotone sequences $\{i_n\}$ and $\{s_n\}$ with $\{a_n\}$ trapped in between. They are defined as follows:

$$\begin{aligned} i_n &= \inf\{a_k : k \geq n\} \\ s_n &= \sup\{a_k : k \geq n\}. \end{aligned} \tag{2.6.1}$$

Note the i_n will all be $-\infty$ if $\{a_n\}$ is not bounded below and the s_n will all be $+\infty$ if $\{a_n\}$ is not bounded above. However, if $\{a_n\}$ is bounded, say $m \leq a_n \leq M$ for all n, then $m \leq i_n \leq s_n \leq M$ for each n. Hence, in this case, the numbers i_n and s_n are all finite and $\{i_n\}$ and $\{s_n\}$ are bounded sequences.

Theorem 2.6.1. *Given a bounded sequence $\{a_n\}$, if $\{i_n\}$ and $\{s_n\}$ are defined as above, then*

(a) *$\{i_n\}$ is a non-decreasing sequence;*

(b) *$\{s_n\}$ is a non-increasing sequence;*

(c) *$i_n \leq a_n \leq s_n$ for all n.*

Proof. If $A_n = \{a_k : k \geq n\}$, then $A_{n+1} \subset A_n$ for each n. It follows from Theorem 1.5.7(e) that, for all n,

$$
\begin{aligned}
s_{n+1} &= \sup A_{n+1} \leq \sup A_n = s_n \text{ and} \\
i_{n+1} &= \inf A_{n+1} \geq \inf A_n = i_n.
\end{aligned}
\tag{2.6.2}
$$

Also, since $a_n \in A_n$, $i_n = \inf A_n \leq a_n \leq \sup A_n = s_n$. $\qquad\square$

Since the sequences $\{i_n\}$ and $\{s_n\}$ are monotone, their limits exist.

Definition 2.6.2. If $\{a_n\}$ is a sequence and $\{i_n\}$ and $\{s_n\}$ are defined as above, then we set

$$
\begin{aligned}
\liminf a_n &= \lim i_n, \\
\limsup a_n &= \lim s_n.
\end{aligned}
\tag{2.6.3}
$$

Note that If $\{a_n\}$ is not bounded below, then $\liminf a_n = -\infty$, while if $\{a_n\}$ is not bounded above, then $\limsup a_n = +\infty$.

Example 2.6.3. Find $\liminf a_n$ and $\limsup a_n$ if $a_n = (-1)^n + 1/n$.
   Solution: As before, we let $i_n = \inf\{a_k : k \geq n\}$ and $s_n = \sup\{a_k : k \geq n\}$. We claim $i_n = -1$ for all n. In fact,

$$
-1 \leq (-1)^k + 1/k \quad \text{for all} \quad k
$$

implies

$$
i_k = \inf\{(-1)^k + 1/k : k \geq n\} \geq -1.
$$

Furthermore, $(-1)^k + 1/k$ approaches -1 for large odd k, so no number greater than -1 is a lower bound for $\{a_k : k \geq n\}$. Thus, $i_n = -1$, as claimed. This implies that $\liminf a_n = \lim i_n = -1$.
   We claim, $1 \leq s_n \leq 1 + 1/n$. In fact, the set $\{(-1)^k + 1/k : k \geq n\}$ contains numbers greater than 1 no matter what n is, and so

$$
s_n = \sup\{(-1)^k + 1/k : k \geq n\} \geq 1.
$$

Furthermore, $(-1)^k + 1/k \leq 1 + 1/n$ if $k \geq n$. Thus, $1 \leq s_n \leq 1 + 1/n$. This implies that $\limsup a_n = \lim s_n = 1$.

Subsequential Limits

If $\{a_n\}$ is a sequence, then by a *subsequential limit* of $\{a_n\}$ we mean a number which is the limit of some subsequence of $\{a_n\}$.

Theorem 2.6.4. *Every subsequential limit of $\{a_n\}$ lies between* $\liminf a_n$ *and* $\limsup a_n$.

Proof. If $\{a_{n_k}\}$ is a convergent subsequence of $\{a_n\}$, Theorem 2.6.1 (c) implies

$$i_{n_k} \leq a_{n_k} \leq s_{n_k},$$

where $i_n = \inf\{a_k : k \geq n\}$ and $s_n = \sup\{a_k : k \geq n\}$. The sequences $\{i_{n_k}\}$ and $\{s_{n_k}\}$ are subsequences of $\{i_n\}$ and $\{s_n\}$, respectively, and, hence, have the same limits, namely $\liminf a_n$ and $\limsup a_n$, by Theorem 2.5.4. It follows from Theorem 2.3.8 and the above inequalities that

$$\liminf a_n \leq \lim a_{n_k} \leq \limsup a_n.$$

$\square$

Theorem 2.6.5. *If $\{a_n\}$ is a sequence, then $\limsup a_n$ and $\liminf a_n$ are subsequential limits of $\{a_n\}$.*

Proof. We will show that $\limsup a_n$ is a subsequential limit of $\{a_n\}$. The same statement for $\liminf$ has a similar proof. We will assume that $\limsup a_n$ is a finite number s. The case where $\limsup a_n = \infty$ is left as an exercise.

We must show that there is some subsequence of $\{a_n\}$ which converges to $s = \limsup a_n$. We will construct such a sequence inductively. As before, we let $s_n = \sup\{a_k : k \geq n\}$. For each $\epsilon > 0$, the number $s_n - \epsilon$ is less than s_n and so it is not an upper bound for $\{a_k : k \geq n\}$. This means there is an element of $\{a_k : k \geq n\}$ which is greater than $s_n - \epsilon$ but less than or equal to s_n. We will choose a sequence of such elements by induction.

We choose n_1 such that $s_1 - 1 < a_{n_1} \leq s_1$. Suppose $n_1 < n_2 < \cdots < n_m$ have been chosen so that

$$s_j - 1/j < a_{n_j} \leq s_j \quad \text{for} \quad j = 1, \cdots, m. \tag{2.6.4}$$

We may then choose $n_{m+1} > n_m$ such that $s_{n_{m+1}} - 1/(m+1) < a_{n_{m+1}} \leq s_{n_{m+1}}$. However, $n_{m+1} \geq m + 1$ and so $s_{n_{m+1}} \leq s_{m+1}$. In other words (2.6.4) holds with m replaced by $m + 1$. This completes the induction step and proves that there is an increasing sequence of natural numbers $\{n_j\}$ such that (2.6.4) holds for all j.

Since both $s_j - 1/j \to s$ and $s_j \to s$, the subsequence $\{a_{n_j}\}$ also converges to s by the squeeze principle. $\square$

A Criterion for Convergence

Theorem 2.6.6. *A sequence $\{a_n\}$ has limit a if and only if* $\limsup a_n = \liminf a_n = a$.

Proof. We first prove that if $\limsup a_n = \liminf a_n = a$, then $\lim a_n$ exists and equals a. By Theorem 2.6.1(c),

$$i_n \leq a_n \leq s_n,$$

where i_n and s_n are as before. Since $\lim i_n = \lim s_n = a$, it follows from the squeeze principle that $\lim a_n = a$.

Next we assume $\lim a_n = a$. By Theorem 2.5.4 each subsequence of $\{a_n\}$ also has limit a. Since $\limsup a_n$ and $\liminf a_n$ are subsequential limits of $\{a_n\}$, they must both be equal to a. This completes the proof. $\qquad\square$

Exercise Set 2.6

1. Find $\limsup a_n$ and $\liminf a_n$ for the following sequences:

    (a) $a_n = (-1)^n$;
    (b) $a_n = (-1/n)^n$;
    (a) $a_n = \sin n\pi/3$.

2. Find $\liminf$ and $\limsup$ for the sequence of Exercise 2.5.6(c).

3. Find $\liminf$ and $\limsup$ for the sequence of Exercise 2.5.7(c).

4. If $\limsup a_n$ and $\limsup b_n$ are finite, prove that

    $$\limsup (a_n + b_n) \leq \limsup a_n + \limsup b_n.$$

5. If $\limsup a_n$ is finite, prove that $\liminf(-a_n) = -\limsup a_n$.

6. If $k \geq 0$ and $\limsup a_n$ is finite, prove that $\limsup ka_n = k\limsup a_n$.

7. If $a_n \geq 0$ and $b_n \geq 0$, prove that $\limsup a_n b_n \leq (\limsup a_n)(\limsup b_n)$.

8. If $\{a_n\}$ and $\{b_n\}$ are non-negative sequences and $\{b_n\}$ converges, prove that $\limsup a_n b_n = (\limsup a_n)(\lim b_n)$.

9. Let $\{r_n\}_{n=1}^{\infty}$ be an enumeration of the rational numbers between 0 and 1. Show that, for each $x \in [0,1]$, there is a subsequence of this sequence which converges to x. Hint: use Exercise 1.4.7.

10. Prove Theorem 2.6.5 for $\limsup$ in the case where $\limsup a_n = +\infty$.

11. Prove that c is $\limsup a_n$ if and only if there is a subsequence of $\{a_n\}$ which converges to c, but there is no subsequence of $\{a_n\}$ which converges to a number greater than c.

12. Which numbers do you think are subsequential limits of the sequence $\{\sin n\}_{n=1}^{\infty}$? Can you prove that your guess is correct?

Chapter 3

Continuous Functions

In this chapter we begin our study of functions of a real variable. The concepts of limit and continuity for such functions are of critical importance.

3.1 Continuity

We will be dealing with functions from a subset of $\mathbb{R}$ to $\mathbb{R}$. Usually in this chapter, the domain of a function will be an interval – closed, open, or half-open, bounded or unbounded – or a finite union of intervals. However, it is certainly possible to consider functions which have much more complicated subsets of $\mathbb{R}$ as domain.

To define a function from a subset of $\mathbb{R}$ to $\mathbb{R}$, we must specify a domain for the function and the rule or formula that specifies the value of the function at each point of that domain. For example, the following are descriptions of functions:

1. $f(x) = 1/x$ on $(0, \infty)$;

2. $g(x) = 1/x$ on $\mathbb{R} \setminus \{0\}$;

3. $h(x) = \sin x$ on $[0, 2\pi]$;

4. $k(x) = \sin x$ on $\mathbb{R}$;

5. $e(x) = \mathrm{e}^x$ on $[0, 1)$.

Although a function may have a *natural domain* – that is, a largest subset of $\mathbb{R}$ on which the formula describing it makes sense – we are at liberty to choose a smaller domain for the function if we wish.

There are a number of special types of functions that we will deal with on a regular basis

1. **Polynomials:** functions of the form $a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0$, where the a_k are constants for $k = 0, \cdots, n$. If $a_n \neq 0$, then the *degree* of the polynomial is n. The natural domain of a polynomial is $\mathbb{R}$;

2. **Rational functions:** functions of the form p/q with p and q polynomials. The natural domain of a function of this form is the set of all real numbers where the denominator q is non-zero;

3. **Trigonometric functions:** sin, cos, tan, cot, sec, csc;

4. **Inverse trigonometric functions:** $\sin^{-1}$, $\tan^{-1}$, etc;

5. **Exponential and log functions:** e^x and $\ln x$.

6. **Power functions:** x^a for $a \in \mathbb{R}$. The natural domain is $\{x \in \mathbb{R}; x \geq 0\}$ unless a is a rational number with an odd denominator – in this case x^a is defined for all real numbers x.

Elementary functions are functions that can be constructed from functions of the above types using addition, multiplication, quotients and composition. It is *not* the case that all the functions we wish to consider are elementary functions.

Continuity

Definition 3.1.1. Let f be a function with domain $D \subset \mathbb{R}$ and let a be an element of D. We will say that f is *continuous* at a if, for each $\epsilon > 0$, there is a $\delta > 0$, such that

$$|f(x) - f(a)| < \epsilon \quad \text{whenever} \quad x \in D \quad \text{and} \quad |x - a| < \delta. \qquad (3.1.1)$$

There is a subtle difference between the definition of continuity given above and the one that is usually given in calculus courses. The difference is that our definition depends on the domain of the function. A given expression may not be continuous at a point a if given one domain containing a, and yet it may be continuous at a if it is given a smaller domain.

Example 3.1.2. Give an example of a function which is not continuous at a certain point of its domain, but it is continuous at this point if a smaller domain is chosen for the function.
    Solution: Each $x \in \mathbb{R}$ is in exactly one of the intervals $[n, n+1)$ for $n \in \mathbb{Z}$. Consider the function defined on $\mathbb{R}$ by

$$f(x) = x - n \quad \text{if} \quad x \in [n, n+1), \ n \in \mathbb{Z}.$$

The graph of this function is shown in Figure 3.1, which shows why this function is called the *sawtooth function*. We will show that this function is not continuous at 0 (or at any other integer for that matter). However, if its domain is restricted to be the interval $[0, 1)$, then it is continuous at 0.
    Now $f(x) = x$ on $[0, 1)$ and $f(x) = x + 1$ on $[-1, 0)$. Suppose ϵ is greater than 0 but less than $1/2$. Then, for any $\delta > 0$, the interval $(-\delta, \delta)$ will contain points of $(-1/2, 0)$ and for any such point x,

$$|f(x) - f(0)| = |x + 1 - 0| > 1/2 > \epsilon.$$
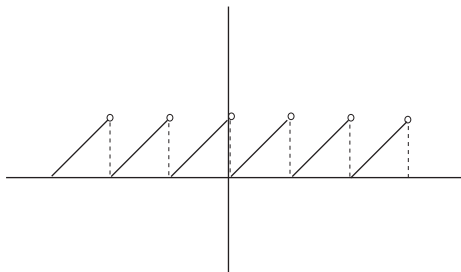
Figure 3.1: The Sawtooth Function.

Thus, there is no way to choose δ such that $|f(x)-f(0)| < \epsilon$ whenever $|x-0| < \delta$. This means that f is not continuous at 0. The same argument works at any other integer n.

On the other hand, suppose we define a new function g which is the same as f, but with domain cut down to be just $D = [0,1)$. Then $g(x) = x$ on D. If, for a given $\epsilon > 0$, we choose $\delta = \epsilon$, then

$$|g(x) - g(0)| = |x| < \epsilon \quad \text{whenever} \quad x \in D, \text{ and } |x - 0| = |x| < \delta.$$

Thus, g is continuous at 0.

Definition 3.1.3. We will simply say that a function with domain D is *continuous* if it is continuous at every point of D.

Example 3.1.4. Prove that $f(x) = x^2$ is continuous at $x = 2$.
    Solution: We have

$$|f(x) - f(2)| = |x^2 - 4| = |x + 2||x - 2|.$$

If we insist that $|x - 2| < 1$, then $1 < x < 3$ and so $|x + 2| < 5$. Thus, given $\epsilon > 0$, if we choose $\delta = \min\{1, \epsilon/5\}$, then

$$|f(x) - f(2)| = |x + 2||x - 2| < 5|x - 2| < \epsilon \quad \text{whenever} \quad |x - 2| < \delta.$$

This proves that f is continuous at 2.

An Alternate Characterization of Continuity

There is an alternate characterization of continuity that will allow us to use the theorems of the previous chapter to easily prove the standard theorems concerning continuous functions:

Theorem 3.1.5. *Let f be a function with domain D and suppose $a \in D$. Then f is continuous at a if and only if, whenever $\{x_n\}$ is a sequence in D which converges to a, then the sequence $\{f(x_n)\}$ converges to $f(a)$.*

Proof. We first prove the "only if" – that is, we assume f is continuous and proceed to prove the statement about sequences. Let $\{x_n\}$ be a sequence in D with $x_n \to a$. Given $\epsilon > 0$, there is a $\delta > 0$ such that

$$|f(x) - f(a)| < \epsilon \quad \text{whenever} \quad x \in D \quad \text{and} \quad |x - a| < \delta.$$

For this δ, there is an N such that

$$|x_n - a| < \delta \quad \text{whenever} \quad n > N.$$

On combining these statements, we conclude

$$|f(x_n) - f(a)| < \epsilon \quad \text{whenever} \quad n > N.$$

Thus, $f(x_n) \to f(a)$. This completes the proof of the "only if" half of the theorem.

We will prove the "if" part, by proving the contrapositive – that is, we will prove that if f is not continuous at a, then there is a sequence $\{x_n\}$ in D such that $x_n \to a$ but $\{f(x_n)\}$ does not converge to $f(a)$.

The assumption that f is not continuous at a means that there is an $\epsilon > 0$ for which no δ can be found for which (3.1.1) is true. This means that, no matter what δ we choose, there is always an $x \in D$ such that

$$|x - a| < \delta \quad \text{but} \quad |f(x) - f(a)| \geq \epsilon.$$

In particular, for each of the numbers $1/n$ for $n \in \mathbb{N}$ we may choose an $x_n \in D$ such that

$$|x_n - a| < 1/n \quad \text{but} \quad |f(x_n) - f(a)| \geq \epsilon.$$

These numbers form a sequence $\{x_n\}$ which converges to a (since $1/n \to 0$), but the image sequence $\{f(x_n)\}$ does not converge to $f(a)$. This completes the proof of the "if" part of the theorem. $\qquad\square$

Combining this with the Main Limit Theorem yields the following:

Theorem 3.1.6. *If r is a positive rational number, then the function $f(x) = x^r$ is continuous on its natural domain.*

Proof. The natural domain D of $f(x) = x^r$ is $\mathbb{R}$ if r has an odd denominator and is the set of non-negative real numbers if r has an even denominator when written in lowest terms. In either case, if $a \in D$ and $\{x_n\}$ is a sequence in D which converges to a, then $\{x_n^r\}$ converges to a^r by parts (e) and (f) of the Main Limit Theorem (Theorem 2.3.6). This implies that x^r is continuous by the previous theorem. $\qquad\square$

Remark 3.1.7. We will eventually prove that the functions x^a for $a \in \mathbb{R}$, e^x, $\ln x$, and the inverse trigonometric functions are all continuous. In the meantime, we will assume this is true whenever it is convenient to do so in an exercise or example. The continuity of the trigonometric functions is usually proved adequately in elementary calculus and so we will use the continuity of these functions whenever it is needed.

Combinations of Continuous Functions

If f and g are functions with domains D_f and D_g, then $f + g$ and fg have domain $D = D_f \cap D_g$, and f/g has domain $\{x \in D : g(x) \neq 0\}$.

Theorem 3.1.8. *Let f and g be functions with domains D_f and D_g. Assume f and g are both continuous at a point $a \in D = D_f \cap D_g$, and let c be a constant. Then*

(a) *cf is continous at a;*

(b) *$f + g$ is continous at a;*

(c) *fg is continous at a;*

(d) *f/g is continous at a, provided $g(a) \neq 0$;*

Proof. These are all proved using the same technique used to prove the previous theorem – combine Theorem 3.1.5 with the corresponding part of the Main Limit Theorem. We will do (b) to illustrate this technique, pose part (d) as an exercise, and let it go at that.

If f and g are continuous at a and $\{x_n\}$ is any sequence in D which converges to a, then Theorem 3.1.5 tells us that $\{f(x_n)\}$ converges to $f(a)$ and $\{g(x_n)\}$ converges to $g(a)$. By part (b) of the Main Limit Theorem (Theorem 2.3.6), $\{f(x_n) + g(x_n)\}$ converges to $f(a) + g(a)$. Therefore, by Theorem 3.1.5 again, $f + g$ is continuous at a. $\qquad\square$

Example 3.1.9. Prove that each polynomial is continuous on all of $\mathbb{R}$ and each rational function is continuous at all points where its denominator is not zero.

Solution: Every positive integral power of x is continuous on $\mathbb{R}$ by Theorem 3.1.6. By (a) of the above theorem, each constant times a power of x is also continuous. Then (b) of the theorem implies that every polynomial is continuous on $\mathbb{R}$ and (d) implies that every rational function is continuous at points where its denominator is not zero.

Composition of Continuous Functions

If f is a function with domain D_f and g is a function with domain D_g, then the composite function $f \circ g$ has domain $D_{f \circ g} = \{x \in D_g : g(x) \in D_f\}$. Suppose a is in this set, so that $a \in D_g$ and $g(a) \in D_f$. Then we can ask if $f \circ g$ is continuous at a. The following theorem answers this question. Its proof is left to the exercises.

Theorem 3.1.10. *With f and g as above, let a be in the domain of $f \circ g$. Then $f \circ g$ is continuous at a if g is continuous at a and f is continuous at $g(a)$.*

Example 3.1.11. Prove that $f(x) = \dfrac{1}{\sqrt{1 - x^2}}$ is continuous as a function on its natural domain.

   Solution: The function f has as natural domain the interval $(-1, 1)$, since it is for points in this interval and those points alone that $\sqrt{1 - x^2}$ is defined and non-zero. The function $1 - x^2$ is continuous on $(-1, 1)$ because it is a polynomial. The square root function is continuous on $[0, \infty)$ by Theorem 3.1.6. Thus, the composition $\sqrt{1 - x^2}$ is continuous by Theorem 3.1.10. Finally, f is continuous by part (d) of Theorem 3.1.8.

Exercise Set 3.1

1. If f is a function with domain $[0, 1]$, what is the domain of $f(x^2 - 1)$?

2. What is the natural domain of the function $\dfrac{x^2 + 1}{x^2 - 1}$. With this as its domain, is this function continuous? Why?

3. We know $\sqrt{x}$ is continuous at all $a \geq 0$, by Theorem 3.1.6. Give another proof of this fact using only the definition of continuity (Definition 3.1.1).

4. Prove that $\dfrac{1}{1 + x^2}$ has natural domain $\mathbb{R}$ and is continuous.

5. Show that the function $f(x) = |x|$ is continuous on all of $\mathbb{R}$.

6. Assuming sin is continuous, prove that $\sin(x^3 - 4x)$ is continuous.

7. Prove (d) of Theorem 3.1.8.

8. Prove Theorem 3.1.10.

9. Consider the function

$$f(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ -1 & \text{if} \quad x < 0. \end{cases}$$

   Is this function continuous if its domain is $\mathbb{R}$?   Is it continuous if its domain is cut down to $\{x \in \mathbb{R} : x \geq 0\}$?   How about if its domain is $\{x \in \mathbb{R} : x \leq 0\}$?

10. Let f be a function with domain D and suppose f is continuous at some point $a \in D$. Prove that, for each $\epsilon > 0$, there is a $\delta > 0$ such that

$$|f(x) - f(y)| < \epsilon \quad \text{whenever} \quad x, y \in D \cap (a - \delta, a + \delta).$$

11. Prove that the function $f(x) = \begin{cases} \sin 1/x & \text{if} \quad x \neq 0 \\ 0 & \text{if} \quad x = 0 \end{cases}$ is not continuous at 0.

12. Prove that the function $f(x) = \begin{cases} x \sin 1/x & \text{if} \quad x \neq 0 \\ 0 & \text{if} \quad x = 0 \end{cases}$ is continuous at 0.

3.2 Properties of Continuous Functions

Continuous functions on closed bounded intervals have a number of highly useful properties. We explore some of these in this section.

Maximum and Minimum Values

A function f with domain D is said to be be *bounded* above on $S \subset D$ if and only if the set $f(S) = \{f(x) : x \in S\}$ is bounded above. This is true if and only if

$$\sup_S f = \sup\{f(x) : x \in S\}$$

is finite. Similarly, f is bounded below on S if $f(S)$ is bounded below and this is true if and only if

$$\inf_S f = \inf\{f(x) : x \in S\}$$

is finite. If f is bounded above and below on S, then we say f is *bounded* on S. If f is bounded on its domain D, then it is said to be a bounded function.

Just as a bounded set may have a finite sup, but may not have a maximum element (the sup may not belong to the set), a function f may be bounded above on S without having a maximum value (this happens if $\sup_S f$ is not a value that f assumes on S). However, if f is a continuous function on a closed bounded interval, then the situation is particularly nice.

Theorem 3.2.1. *If f is a continuous function on a closed bounded interval I, then f is bounded on I and, in fact, it assumes both a minimum and a maximum value on I.*

Proof. We will prove that $M = \sup_{x \in I} f(x)$ is finite and, in fact, is a value that f takes on somewhere on I. The proof of the analogous fact for $\inf_{x \in I} f(x)$ has the same proof.

We will inductively construct a nested sequence of closed intervals $\{I_n\}$ with the following properties:

(1) $I_1 = I$;

(2) I_k is the closed left or right half of I_{k-1} for each $k > 1$;

(3) $\sup_{I_k} f(x) = M$ for each k.

The first condition tells us how to pick I_1. Suppose that $I_1, \cdots, I_n$ have been chosen satisfying (1), (2), (3) for $k \leq n$. we choose I_{n+1} as follows: If I_n is cut in half at its midpoint, yielding two closed intervals with union I_n and with intersection the midpoint of I_n, then the sup of f on at least one of these intervals must be the same as the sup of f on I_n. This is M by our induction assumption. If this is true of only one of the two halves of I_n, we choose this half to be I_{n+1}. If it is true of both halves, then we choose I_{n+1} to be the right half of I_n. This completes the induction step of the definition and proves that a sequence $\{I_n\}$ satisfying (1), (2), (3) can be constructed.

Given a nest of intervals $\{I_n\}$ as above, the Nested Interval Property (Theorem 2.5.1) implies that there is a point $a \in \bigcap_n I_n$. This is, in particular, a point of $I = I_1$. We know f is continuous at this point and so, given $\epsilon > 0$, there is a $\delta > 0$ such that

$$f(a) - \epsilon < f(x) < f(a) + \epsilon \quad \text{whenever} \quad x \in I, \ |x - a| < \delta. \qquad (3.2.1)$$

Now the length of I_n is $L/2^{n-1}$, where L is the length of I. Since $\lim L/2^{n-1} = 2L \lim (1/2)^n = 0$, the length of I_n will be less than δ for n sufficiently large. Suppose n is this large. Then $|x - a| < \delta$ for all $x \in I_n$, since $a \in I_n$. By (3.2.1)

$$f(a) - \epsilon < \sup_{I_n} f \leq f(a) + \epsilon.$$

That is,

$$f(a) - \epsilon < M \leq f(a) + \epsilon.$$

This implies that M is finite and that $|f(a) - M| \leq \epsilon$ for every positive ϵ. This is possible only if $f(a) = M$. Thus we have proved that $\sup_{x \in I} f(x)$ is finite and that it is a value assumed by f at some point a of I. $\qquad \square$

Each of the hypotheses of the above theorem is necessary in order for the conclusion to hold. This is illustrated by the following example and some of the exercises.

Example 3.2.2. Give examples of functions on $[0, 1]$ which are

(1) unbounded;

(2) bounded, but with no maximum value.

   Solution: (1) Let

$$f(x) = \begin{cases} 1 & \text{if} \quad x \leq 1/2 \\ \dfrac{1}{2x - 1} & \text{if} \quad x > 1/2. \end{cases};$$

this function is clearly unbounded on $[0, 1]$ since it blows up as x approaches $1/2$ from the right. Note that f is not continuous at $1/2$.
   (2) Let

$$f(x) = \begin{cases} 2x & \text{if} \quad x < 1/2 \\ 0 & \text{if} \quad x \geq 1/2. \end{cases};$$

this function is bounded on $[0, 1]$ and its sup on this interval is 1, but it never takes on the value 1 on the interval. Again, this function is not continuous at $1/2$.

Exercises 3.2.4 and 3.2.5 ask the student to come up with examples showing that the conclusion of the theorem fails for a function which is continuous on an interval I, but I is not closed or is not bounded.

Intermediate Value Theorem

The next theorem says that if a continuous function on an interval takes on two values, then it takes on every value in between. Its proof is almost identical to the proof of the previous theorem.

Theorem 3.2.3. (Intermediate Value Theorem) *Let f be defined and continuous on an interval containing the points a and b and assume that $a < b$. If y is any number between $f(a)$ and $f(b)$, then there is a number c with $a \leq c \leq b$ such that $f(c) = y$.*

Proof. Let $a_1 = a$ and $b_1 = b$ and consider the closed interval $I_1 = [a_1, b_1]$. We are given that y lies between $f(a_1)$ and $f(b_1)$. We will construct a nested sequence of closed intervals with the same property. That is, we will prove by induction that there is a sequence of closed intervals $\{I_k = [a_k, b_k]\}$ such that, for all $k > 1$,

(1) $[a_k, b_k]$ is the closed left or right half of the interval $[a_{k-1}, b_{k-1}]$;

(2) y lies between $f(a_k)$ and $f(b_k)$.

Suppose it is possible to choose $\{I_1, \cdots, I_n\}$ so that (1) and (2) hold for $k \leq n$. Then we cut I_n into two halves that have only the midpoint c_n of I_n in common. If y lies between $f(a_n)$ and $f(b_n)$ then it either lies between $f(a_n)$ and $f(c_n)$ or it lies between $f(c_n)$ and $f(b_n)$. If only one of these is true, then choose I_{n+1} to be the corresponding half of I_n. If both are true, then choose I_{n+1} to be the right half of I_n. This results in a choice for I_{n+1} that satisfies (1) and (2) for $k = n + 1$. This completes the induction step of the construction and, hence, the proof that a nested sequence of intervals satisfying (1) and (2) can be constructed.

By the Nested Interval Property, there is a point c in the intersection of all the intervals I_n. By hypothesis f is continuous at c and so, given $\epsilon > 0$, there is a $\delta > 0$ such that

$$f(c) - \epsilon < f(x) < f(c) + \epsilon \quad \text{whenever} \quad x \in I, \ |x - c| < \delta. \qquad (3.2.2)$$

Now the length of I_n is $L/2^{n-1}$, where L is the length of I. Since $\lim L/2^{n-1} = 2L \lim(1/2)^n = 0$, the length of I_n will be less than δ for n sufficiently large. Suppose n is this large. Then $|x - c| < \delta$ for all $x \in I_n$, since $c \in I_n$. By (3.2.2)

$$f(c) - \epsilon < f(a_n) < f(c) + \epsilon \quad \text{and} \quad f(c) - \epsilon < f(b_n) < f(c) + \epsilon.$$

Taken together with the fact that y lies between $f(a_n)$ and $f(b_n)$, these inequalities imply that

$$f(c) - \epsilon < y < f(c) + \epsilon \quad \text{or} \quad |f(c) - y| < \epsilon.$$

This is only possible for all positive ϵ if $f(c) = y$. This completes the proof. $\square$

This is another example of a theorem which is not true if the function is not required to be continuous (see Exercise 3.2.6).

Image of an Interval

Theorem 3.2.4. *If f is a continuous function defined on a closed bounded interval $I = [a, b]$, then $f(I)$ is also a closed, bounded interval or it is a single point.*

Proof. By Theorem 3.2.1, f has a maximum value M and a minimum value m on I. By Theorem 3.2.3 f takes on every value between m and M on I. Therefore the image of I is exactly $[m, M]$. This is a closed interval if $m \neq M$, and is a point otherwise. □

Inverse Functions

We learn in calculus that a function which is monotone increasing or monotone decreasing on an interval has an inverse function. Here a function f is *monotone increasing* on I if $f(x) < f(y)$ whenever $x, y \in I$ and $x < y$. A function f is *monotone decreasing* on I if $f(x) > f(y)$ whenever $x, y \in I$ and $x < y$. A function which is monotone increasing or monotone decreasing on I is said to be *strictly monotone* on I. For strictly monotone functions, there is a converse to the previous theorem.

Theorem 3.2.5. *If f is strictly monotone on I and its range $f(I)$ is an interval, then f is continuous on I.*

Proof. Suppose f is monotone increasing. Let $f(I) = [s, t]$. Given $c \in I$, we will prove that f is continuous at c. We do this first in the case where c is not an endpoint of $I = [a, b]$.

Given $\epsilon > 0$, let $u = \max\{s, f(c) - \epsilon\}$ and $v = \min\{t, f(c) + \epsilon\}$. Then u and v are points of $[s, t]$ and

$$f(c) - \epsilon \leq u \leq f(c) \leq v \leq f(c) + \epsilon.$$

Note that the only way one of the inequalities $u \leq f(c) \leq v$ can be an equality is if $f(c)$ is one of the endpoints s or t. However, this cannot happen, since c is not an endpoint of I. Thus, $u < f(c) < v$.

Since $f(I) = [s, t]$, there are points $p, q \in I$ such that $f(p) = u$ and $f(q) = v$. Since f is monotone increasing,

$$p < c < q.$$

We choose $\delta = \min\{q - c, c - p\}$. Then $|x - c| < \delta$ implies $p < x < q$ and this implies

$$f(c) - \epsilon \leq u < f(x) < v \leq f(c) + \epsilon \quad \text{that is} \quad |f(x) - f(c)| < \epsilon.$$

This proves that f is continuous at c in the case where c is not an endpoint of I.

If c is an endpoint of I, then the argument is the same except that we only have to concern ourselves with points that lie to one side of c and of $f(c)$. The details are left to the exercises.

It remains to prove that a monotone *decreasing* function on I with a closed interval for its range is continuous. However, if g is monotone decreasing, then $f = -g$ is monotone increasing, also has a closed interval as image and, hence, is continuous by the above. But if $-g$ is continuous, then so is $g = (-1)(-g)$. $\square$

Theorem 3.2.6. *A continuous, strictly monotone function f on a closed interval I has a continuous inverse function defined on $J = f(I)$. That is, there is a continuous function g, with domain J, such that $g(f(x)) = x$ for all $x \in I$ and $f(g(y)) = y$ for all $y \in J$.*

Proof. Since f is strictly monotone, for each $y \in J$ there is exactly one $x \in I$ such that $f(x) = y$. We set $g(y) = x$. Then, by the choice of x, we have $f(g(y)) = f(x) = y$ and $g(f(x)) = g(y) = x$.

The function g is strictly monotone because f is strictly monotone. Furthermore, the range of g is I. By the previous theorem, this implies that g is continuous. $\square$

Exercise Set 3.2

1. Find the maximum and minumum values of the function $f(x) = x^2 - 2x$ on the interval $[0, 3)$.

2. Prove that if f is a continuous function on a closed bounded interval I and if $f(x)$ is never 0 for $x \in I$, then there is a number $m > 0$ such that $f(x) \geq m$ for all $x \in I$ or $f(x) \leq -m$ for all $x \in I$.

3. Prove that if f is a continuous function on a closed bounded interval $[a, b]$ and if (x_0, y_0) is any point in the plane, then there is a closest point to (x_0, y_0) on the graph of f.

4. Find an example of a function which is continuous on a bounded (but not closed) interval I, but is not bounded. Then find an example of a function which is continuous and bounded on a bounded interval I, but does not have a maximum value.

5. Find an example of a function which is continuous on a closed (but not bounded) interval I, but is not bounded. Then find an example of a function which is continuous and bounded on a closed interval I, but does not have a maximum value.

6. Give an example of a function defined on the interval $[0, 1]$, which does not take on every value between $f(0)$ and $f(1)$.

7. Show that if f and g are continuous functions on the interval $[a, b]$ such that $f(a) < g(a)$ and $g(b) < f(b)$, then there is a number $c \in (a, b)$ such that $f(c) = g(c)$.

8. Let f be a continuous function from $[0, 1]$ to $[0, 1]$. Prove there is a point $c \in [0, 1]$ such that $f(c) = c$ – that is, show that f has a *fixed point*. Hint: apply the Intermediate Value Theorem to the function $g(x) = f(x) - x$.

9. Use the Intermediate Value Theorem to prove that, if n is a natural number, then every positive number a has a positive nth root.

10. Prove that a polynomial of odd degree has at least one real root.

11. Use the Intermediate Value Theorem to prove that if f is a continuous function on an interval $[a, b]$ and if $f(x) \leq m$ for every $x \in [a, b)$, then $f(b) \leq m$.

12. Prove that if f is strictly increasing on $[a, b]$, then its inverse function is strictly increasing on $[f(a), f(b)]$.

3.3   Uniform Continuity

Compare the definition of continuity given in Definition 3.1.1 with the following definition.

Definition 3.3.1. If f is a function with domain D, then f is said to be *uniformly continuous* on D if for each $\epsilon > 0$ there is a $\delta > 0$ such that

$$|f(x) - f(a)| < \epsilon \quad \text{whenever} \quad x, a \in D \text{ and } |x - a| < \delta. \qquad (3.3.1)$$

By contrast, Definition 3.1.1 tells us that f is continuous on D if for each $a \in D$ and each $\epsilon > 0$ there is a $\delta > 0$ such that

$$|f(x) - f(a)| < \epsilon \quad \text{whenever} \quad x \in D \text{ and } |x - a| < \delta.$$

These two definitions appear to be identical until one examines them closely. The difference is subtle but extremely important. In the definition of uniform continuity, given ϵ, a single δ must be chosen that works for all points $a \in D$, while in the definition of continuity, δ is allowed to depend on a.

Example 3.3.2. Find a function which is continuous on its domain, but not uniformly continuous.

Solution: We claim that the functiion $f(x) = 1/x$ with domain $(0, 1]$ is continuous but not uniformly continuous on $(0, 1]$.

It is continuous because x is continuous on $(0, 1]$ and is never 0 on this set. Thus, Theorem 3.1.8(d) implies that $1/x$ is continuous at each point of $(0, 1]$.

On the other hand, if we attempt to verify that f is uniformly continuous, we run into trouble. Given $\epsilon > 0$, we try to find a $\delta > 0$ such that

$$|1/x - 1/a| < \epsilon \quad \text{whenever} \quad a, x \in (0, 1] \quad \text{and} \quad |x - a| < \delta.$$

However, for any $\delta > 0$, if x and a are chosen so that $0 < x < a < \delta$, then it will be true that

$$|x - a| < \delta.$$

However, we can make $1/x$ and, hence, $|1/x - 1/a|$ as large as we want by simply keeping $a < \delta$ fixed and choosing $x < a$ small enough. In particular, $|1/x - 1/a|$ can be made larger than ϵ regardless of what ϵ we start with. Thus, $f(x) = 1/x$ is not uniformly continuous on $(0, 1]$
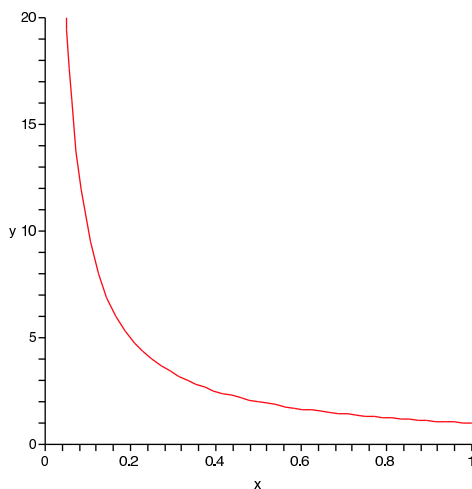
Figure 3.2: The Function $1/x$ on $(0, 1]$.

Example 3.3.3. Prove that $f(x) = 1/x$ is uniformly continuous on any interval of the form $[r, 1]$, where $r > 0$.

Solution: If x and a are in the interval $[r, 1]$, then

$$\left| \frac{1}{x} - \frac{1}{a} \right| = \frac{|x - a|}{ax} \leq \frac{|x - a|}{r^2}.$$

Thus, given $\epsilon > 0$, if we choose $\delta = r^2 \epsilon$, then

$$\left| \frac{1}{x} - \frac{1}{a} \right| < \epsilon \quad \text{whenever} \quad |x - a| < \delta.$$

This implies that $f(x) = 1/x$ is uniformly continuous on $[r, 1]$.

Conditions Ensuring Uniform Continuity

In the last example, the domain of the function f was a closed, bounded interval. It turns out that, in this case, continuity implies uniform continuity. This is the main theorem of the section.

Theorem 3.3.4. *If f is a continuous function on a closed, bounded interval I, then f is uniformly continuous on I.*

Proof. We will prove the contrapositive. Suppose f is not uniformly continuous on $[a, b]$. Then there is an $\epsilon > 0$ for which no δ can be found which satisfies (3.3.1). In particular, none of the numbers $1/n$ for $n \in \mathbb{N}$ will suffice for δ. This means that, for each n, there are numbers $x_n, a_n \in I$ such that

$$|x_n - a_n| < 1/n \quad \text{but} \quad |f(x_n) - f(a_n)| \geq \epsilon.$$

By the Bolzano-Weierstrass Theorem, some subsequence $\{x_{n_k}\}$ of the sequence $\{x_n\}$ converges to a point x of I. The inequality $|x_{n_k} - a_{n_k}| < 1/n_k \leq 1/k$ implies that $\{a_{n_k}\}$ converges to the same number. Since $|f(x_{n_k}) - f(a_{n_k})| \geq \epsilon$, the sequences $\{f(x_{n_k})\}$ and $\{f(a_{n_k})\}$ cannot converge to the same number. However, they would both have to converge to $f(x)$ if f were continuous at x, by Theorem 3.1.5. Thus, we conclude that f is not continuous at every point of I. □

Consequences of Uniform Continuity

Theorem 3.3.5. *If f is uniformly continuous on its domain D, and if $\{x_n\}$ is any Cauchy sequence in D, then $\{f(x_n)\}$ is also a Cauchy sequence.*

Proof. Given $\epsilon > 0$, by uniform continuity there is a $\delta > 0$ such that

$$|f(x) - f(y)| < \epsilon \quad \text{whenever} \quad x, y \in D \text{ and } |x - y| < \delta.$$

Since $\{x_n\}$ is Cauchy, there is an N such that

$$|x_n - x_m| < \delta \quad \text{whenever} \quad n, m > N.$$

Combining these two statements tells us that

$$|f(x_n) - f(x_m)| < \epsilon \quad \text{whenever} \quad n, m > N.$$

Thus, $\{f(x_n)\}$ is a Cauchy sequence. □

An interval may be closed, open or half open. If I is an interval, we denote by $\overline{I}$ the closed interval consisting of I along with any endpoints of I that may be missing from I. If I is a bounded interval, then $\overline{I}$ is a closed, bounded interval.

Given a continuous function f on a bounded interval I that is not closed, it may or may not be possible to extend f to a continuous function on $\overline{I}$. That is, it may or may not be possible to give f values at the missing endpoint(s) that make the new function continuous. The next theorem tells when this can be done.

Theorem 3.3.6. *If f is a continuous function on a bounded interval I , which may not be closed, then f has a continuous extension to $\overline{I}$ if and only if f is uniformly continuous on I.*

Proof. If f has a continuous extension $\tilde{f}$ to $\overline{I}$, then $\tilde{f}$ is uniformly continuous on $\overline{I}$ by Theorem 3.3.4. But if a function is uniformly continuous on a set, then it is also uniformly continuous when restricted to any smaller set. Since f is just $\tilde{f}$ restricted to the smaller domain I, f is uniformly continuous on I.

Conversely, suppose f is uniformly continuous on I. Let a be a missing endpoint of I (left or right). There are lots of sequences in I which converge to a. Let $\{a_n\}$ be one of these. Then $\{a_n\}$ is a Cauchy sequence in I and so the previous theorem implies that $\{f(a_n)\}$ is also a Cauchy sequence. Since Cauchy sequences converge, we know that there is a y such that $f(a_n) \to y$.

We claim that if $\{b_n\}$ is any other sequence in I converging to a, then $\{f(b_n)\}$ converges to the same number y. We prove this by constructing a new sequence $\{c_n\}$ in I, which also converges to a, by interlacing the terms of $\{a_n\}$ and $\{b_n\}$. That is, we set

$$c_{2k-1} = a_k;$$
$$c_{2k} = b_k.$$

Since $c_n \to a$, we may argue as before, that $\{f(c_n)\}$ converges to some number. But one of its subsequences, $\{f(c_{2k-1})\}$, converges to y. This implies that $\{f(c_n)\}$ must converge to y as must any of its subsequences. In particular $\{c_{2k}\} = \{b_k\}$ converges to y. This proves our claim. That is, the number $y = \lim f(a_n)$ is the same no matter what sequence $\{a_n\}$ in I converging to a is chosen.

We now define a new function $\tilde{f}$ on $I \cup \{a\}$, by setting $\tilde{f}(a) = y$ and $\tilde{f}(x) = f(x)$ for each $x \in I$. It is clear from the construction that $\tilde{f}$ will be continuous at a, since $\tilde{f}(x_n) \to y = f(a)$ for every sequence $\{x_n\}$ in $I \cup \{a\}$ that converges to a.

This proves that a uniformly continuous function on a bounded interval I can be extended to be continuous on the interval obtained by adjoining one missing endpoint to I. If the other endpoint is also missing, we simply repeat the process to get an extension to all of $\overline{I}$. $\square$

This theorem often provides a quick way to see that a function on a bounded interval is not uniformly continuous.

Example 3.3.7. Show that the function $f(x) = \dfrac{1}{1 - x^2}$ is not uniformly continuous on the interval $(-1, 1)$.

Solution: If f is uniformly continuous on this interval, then the previous theorem implies that f has a continuous extension to $[-1, 1]$. However, a continuous function on a closed bounded interval is bounded. The function f is not bounded on $(-1, 1)$, and so no extension of it to $[-1, 1]$ can be bounded. Thus, f is not uniformly continuous.

If the interval I is unbounded, then it is possible for a function on I to be uniformly continuous and yet unbounded.

Example 3.3.8. Show that the function $f(x) = \sqrt{x}$ is uniformly continuous on $[1, +\infty)$.

Solution: If $x, y \in [1, +\infty)$, then

$$|\sqrt{x} - \sqrt{y}| = \frac{|x - y|}{\sqrt{x} + \sqrt{y}} < |x - y|,$$

since $\sqrt{x} \geq 1 > 1/2$ and $\sqrt{x} \geq 1 > 1/2$ if $x, y \in [1, +\infty)$. This clearly implies that f is uniformly continuous on $[1, +\infty)$. In fact, given $\epsilon > 0$, it suffices to choose $\delta = \epsilon$ to obtain

$$|f(x) - f(y)| < \epsilon \quad \text{whenever} \quad x, y \in [1, +\infty) \text{ and } |x - y| < \delta.$$

Exercise Set 3.3

1. Is the function $f(x) = x^2$ uniformly continuous on $(0,1)$? Justify your answer.

2. Is the function $f(x) = 1/x^2$ uniformly continuous on $(0,1)$? Justify your answer.

3. Is the function $f(x) = x^2$ uniformly continuous on $(0,+\infty)$? Justify your answer.

4. Using only the $\epsilon-\delta$ definition of uniform continuity, prove that the function $f(x) = \dfrac{x}{x+1}$ is uniformly continuous on $[0,\infty)$.

5. In Example 3.3.8 we showed that $\sqrt{x}$ is uniformly continuous on $[1,+\infty)$. Show that it is also uniformly continuous on $[0,1]$.

6. Prove that if I and J are overlapping intervals in $\mathbb{R}$ $(I \cap J \neq \emptyset)$ and f is a function, defined on $I \cup J$, which is uniformly continuous on I and uniformly continuous on J, then it is also uniformly continuous on $I \cup J$. Use this and the previous exercise to prove that $\sqrt{x}$ is uniformly continuous on $[0,+\infty)$

7. Prove that if I is a bounded interval and f is an unbounded function defined on I, then f cannot be uniformly continuous.

8. Let f be a function defined on an interval I and suppose that there are positive constants K and r such that

$$|f(x) - f(y)| \leq K|x - y|^r \quad \text{for all} \quad x, y \in I.$$

   Prove that f is uniformly continuous.

9. Is the function $f(x) = \sin 1/x$ continuous on $(0,1)$? Is it uniformly continuous on $(0,1)$. Justify your answers.

10. Is the function $f(x) = x \sin 1/x$ uniformly continuous on $(0,1)$? Justify your answer.

3.4 Uniform Convergence

Uniform convergence is a subject that is both similar to and very different from uniform continuity. Uniform continuity is a condition on the continuity of a single function, while uniform convergence is a condition on the convergence of a sequence of functions.

Sequences of Functions

In calculus we often encounter sequences of functions as opposed to sequences of numbers. They occur as partial sums of power series, for example. Other examples are the following (note that x is a variable):

1. $\{x/n\}$, $x \in \mathbb{R}$;

2. $\{x^n\}$, $x \in \mathbb{R}$;

3. $\left\{ \dfrac{1}{1 + nx} \right\}$, $x > 0$;

4. $\left\{ \dfrac{1 - x^n}{1 - x} \right\}$, $x \in (-1, 1)$;

5. $\{\sin nx\}$, $x \in [0, 2\pi)$.

It is important to have methods to show that various things are preserved by passing to the limit of a sequence of functions. If the functions in the sequence are all continuous on a certain set D, is the limit continuous on D? Is the integral of the limit equal to the limit of the integrals if we are integrating over some interval on which all the functions are defined? The answer to both of these questions is "yes" provided the convergence is *uniform*.

Uniform Convergence

Let $\{f_n\}$ be a sequence of functions on a set $D \subset \mathbb{R}$. We say that $\{f_n\}$ converges pointwise to a function f on D if, for each $x \in D$, the sequence of numbers $\{f_n(x)\}$ converges to the number $f(x)$. If we write out what this means in terms of the definition of convergence of a sequence of numbers we get the statement in (a) of the following definition. Statement (b) is the definition of uniform convergence.

Definition 3.4.1. Let $\{f_n\}$ be a sequence of functions on a set $D \subset \mathbb{R}$. Then

(a) $\{f_n\}$ is said to converge *pointwise* to a function f on D if, for each $x \in D$ and each $\epsilon > 0$, there is an N such that

$$|f(x) - f_n(x)| < \epsilon \quad \text{whenever} \quad n > N.$$

(b) $\{f_n\}$ is said to converge *uniformly* on D to a function f if, for each $\epsilon > 0$, there is an N such that

$$|f(x) - f_n(x)| < \epsilon \quad \text{whenever} \quad x \in D \text{ and } n > N;$$

As with continuity and uniform continuity, the definitions of pointwise convergence and uniform convergence seem identical until one studies them closely. In fact, they are very different. In the case of pointwise convergence, x is given along with ϵ before N is chosen. Here N may well depend on both ϵ and x. In the case of uniform convergence, only ϵ is given initially; then an N must be chosen which works for all x. That is, N does not depend on x in this case.
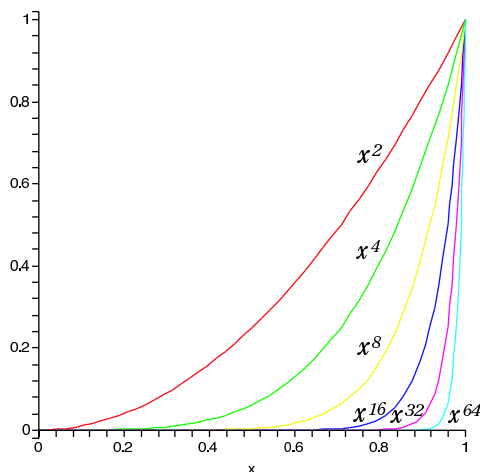
Figure 3.3: The Sequence $\{x^n\}$ does not Converge Uniformly on $[0,1]$.

Example 3.4.2. Give an example of a sequence of functions defined on $[0,1]$ which converges pointwise on $[0,1]$ but not uniformly.

   Solution: An example is the sequence $\{f_n\}$ on $[0,1]$ defined by $f_n(x) = x^n$, which is illustrated in Figure 3.3. This sequence of functions converges to the function f which is 0 if $x < 1$ and is 1 if $x = 1$. Since the sequence $\{f_n(x)\}$ converges to $f(x)$ for each value of x, the sequence $\{f_n\}$ converges pointwise to f on $[0,1]$. However, the convergence is not uniform on $[0,1]$. In fact,

$$|f_n(x) - f(x)| = x^n \quad \text{if} \quad x \in [0,1),$$

and so, given $\epsilon > 0$, in order for it to be true that $|f_n(x) - f(x)| < \epsilon$ for all $x \in [0,1]$ and some n, we would need that

$$x^n < \epsilon \quad \text{for all} \quad x \in [0,1).$$

However, since x^n is continuous on $[0,1]$, this would imply that $1 = 1^n \leq \epsilon$ (Exercise 3.2.11). Obviously, there are positive numbers ϵ for which this is not true (any positive $\epsilon < 1$). This shows that the convergence of $\{f_n\}$ on $[0,1]$ is not uniform.

   The problem in the above example is due to what is happening near $x = 1$. If we stay away from 1, the situation improves.

Example 3.4.3. If $0 < r < 1$, prove that the sequence $\{f_n\}$, defined by $f_n(x) = x^n$, converges uniformly to 0 on $[0, r]$.

   Solution: We have

$$|x^n - 0| = x^n \leq r^n \quad \text{for all} \quad x \in [0, r]. \tag{3.4.1}$$

Now, given $\epsilon > 0$, we choose N so that

$$r^n < \epsilon \quad \text{whenever} \quad n > N,$$

This is possible because $r^n \to 0$ if $0 \le r < 1$. Combining this with (3.4.1) yields

$$|x^n - 0| < \epsilon \quad \text{whenever} \quad x \in [0, r] \text{ and } n > N.$$

This proves that $\{x^n\}$ converges uniformly to 0 on $[0, r]$.

Uniform Convergence and Continuity

Theorem 3.4.4. *Let $\{f_n\}$ be a sequence of functions, all of which are defined and continuous on a set D. If $\{f_n\}$ converges uniformly to a function f on D, then f is continuous on D.*

Proof. If $a \in D$, we will show that f is continuous at a. Given $\epsilon > 0$, we first use the uniform convergence to choose an N such that

$$|f_n(x) - f(x)| < \epsilon/3 \quad \text{whenever} \quad x \in D, n > N.$$

We then fix a natural number $n > N$ and use the fact that each f_n is continuous at a to choose a $\delta > 0$ such that

$$|f_n(x) - f_n(a)| < \epsilon/3 \quad \text{whenever} \quad x \in D \text{ and } |x - a| < \delta.$$

On combining these and using the triangle inequality, we conclude that

$$|f(x) - f(a)| \le |f(x) - f_n(x)| + |f_n(x) - f_n(a)| + |f_n(a) - f(a)|$$
$$< \epsilon/3 + \epsilon/3 + \epsilon/3 = \epsilon,$$

whenever $x \in D$ and $|x - a| < \delta$. This proves that f is continuous at a. Since a was an arbitrary point of D, f is continuous on D. $\qquad\blacksquare$

Example 3.4.5. Analyze the convergence of the sequence of functions $\{f_n\}$ defined on $[0, \infty)$ by

$$f_n(x) = \frac{1}{1 + nx}$$

Does the sequence converge pointwise? Does it converge uniformly?

Solution: Since $f_n(0) = 1$ for all n, the sequence $\{f_n(x)\}$ converges to 1 at $x = 0$. Since each f_n can be re-written as

$$f_n(x) = \frac{1/n}{1/n + x},$$

and the denominator of this expression converges to x, the sequence $\{f_n(x)\}$ converges to 0 if $x \ne 0$. Thus, $\{f_n(x)\}$ converges pointwise to the function f on $[0, \infty)$ defined by $f(x) = 0$ if $x > 0$ and $f(0) = 1$.

It follows from the previous theorem that the convergence is not uniform, because f is not continuous on $[0, \infty)$ although each of the functions f_n is continuous on this interval.

Tests For Uniform Convergence

A sequence $\{f_n\}$ converges uniformly to f on a set D if and only if $\{|f_n - f|\}$ converges uniformly to 0 on D. Thus, it is useful to have simple tests for when a sequence converges uniformly to 0. We will give two such tests. One gives conditions which guarantee that a sequence converges uniformly to 0 and the other gives a condition, which if not true, guarantees that a sequence does not converge uniformly to 0. Both theorems have very simple proofs which are left to the exercises.

The following theorem is useful for showing that a sequence converges uniformly.

Theorem 3.4.6. *Let $\{f_n\}$ be a sequence of functions defined on a set D. If there is a sequence of numbers b_n, such that $b_n \to 0$, and*

$$|f_n(x)| \le b_n \quad \text{for all} \quad x \in D,$$

then $\{f_n\}$ converges uniformly to 0 on D.

The following theorem provides a useful test for proving a sequence does not converge uniformly.

Theorem 3.4.7. *Let $\{f_n\}$ be a sequence of functions defined on a set D. If $\{f_n\}$ converges uniformly to 0 on D, then $\{f_n(x_n)\}$ converges to 0 for every sequence $\{x_n\}$ of points of D.*

Example 3.4.8. If $f_n(x) = \dfrac{n}{x + n}$, prove that $\{f_n\}$ converges uniformly to 1 on the interval $[0, r]$ for each positive number r, but does not converge uniformly on $[0, \infty)$.

Solution: We have

$$|f_n(x) - 1| = \frac{x}{x + n} \le \frac{x}{n} \le \frac{r}{n},$$

if $x \in [0, r]$. Since $r/n \to 0$, Theorem 3.4.6 implies that $\dfrac{x}{x + n}$ converges uniformly to 0 on $[0, r]$ and, hence, that $\{f_n\}$ converges uniformly to 1 on $[0.r]$.

On the other hand if we set $x_n = n$, then $\{x_n\}$ is a sequence of numbers in $[0, \infty)$ and $f_n(x_n) = 1/2$. Since $f_n(x_n) - 1$ does not converge to 0, Theorem 3.4.7 implies that $\{f_n - 1\}$ does not converge uniformly to 0 on $[0, \infty)$ and, hence, that $\{f_n\}$ does not converge uniformly to 1 on $[0, \infty)$.

Uniformly Cauchy Sequences

Definition 3.4.9. A sequence of functions $\{f_n\}$ on a set D is said to be *uniformly Cauchy* on D if for each $\epsilon > 0$, there is an N such that

$$|f_n(x) - f_m(x)| < \epsilon \quad \text{whenever} \quad x \in D \text{ and } n, m > N.$$

If $\{f_n\}$ is a uniformly Cauchy sequence, then $\{f_n(x)\}$ is a Cauchy sequence for each $x \in D$. By Theorem 2.5.8, $\{f_n(x)\}$ converges. Thus, $\{f_n\}$ converges pointwise to some function f on D. The next theorem tells us that the convergence is uniform. Its proof is left to the exercises.

Theorem 3.4.10. *A sequence of functions $\{f_n\}$ on D is uniformly convergent on D if and only if it is uniformly Cauchy on D.*

Exercise Set 3.4

1. Prove that the sequence $\{x/n\}$ converges uniformly to 0 on each bounded interval, but does not converge uniformly on $\mathbb{R}$.

2. Prove that the sequence $\dfrac{1}{x^2 + n}$ converges uniformly to 0 on $\mathbb{R}$.

3. Prove that the sequence $\{\sin(x/n)\}$ converges to 0 pointwise on $\mathbb{R}$, but it does not converge uniformly on $\mathbb{R}$.

4. Prove that the sequence $\dfrac{\sin nx}{n}$ converges uniformly to 0 on $[0, 1]$.

5. Prove that $\{x^n(1-x)\}$ converges uniformly to 0 on $[0, 1]$. Hint: find where each of these functions has its maximum on $[0, 1]$.

6. Prove Theorem 3.4.6.

7. Prove Theorem 3.4.7.

8. Prove that if $\{f_n\}$ is a sequence of uniformly continuous functions on a set D and if this sequence converges uniformly to f on D, then f is also uniformly continuous.

9. For $x \in (-1, 1)$ set $s_n(x) = \displaystyle\sum_{k=0}^{n} x^k$. This is the nth partial sum of a geometric series. Prove that $s_n(x) = \dfrac{1 - x^{n+1}}{1 - x}$.

10. Prove that the sequence $\{s_n\}$ of the previous exercise converges uniformly to $\dfrac{1}{1 - x}$ on each interval of the form $[-r, r]$ with $r < 1$, but it does not converge uniformly on $(-1, 1)$.

11. Prove Theorem 3.4.10. Hint: use an argument like the one in the proof of Theorem 2.5.8.

12. Prove that if $\{a_k\}$ is a bounded sequence of numbers and a sequence $\{s_n\}$ is defined on $(-1, 1)$ by

$$s_n(x) = \sum_{k=0}^{n} a_k x^k,$$

then $\{s_n\}$ converges to a continuous function on $(-1,1)$. Hint: prove this sequence is uniformly Cauchy on each interval $[-r,r]$ for $0 < r < 1$.

Chapter 4

The Derivative

In this chapter we will prove the standard theorems from calculus concerning differentiation – theorems such as the Chain Rule, the Mean Value Theorem, and L'Hôpital's Rule.

We begin with the concept of the limit of a function.

4.1 Limits of Functions

Definition 4.1.1. Let I be an open interval, a a point of I, and f a function defined on I except possibly at a itself. Then we will say the limit of $f(x)$ as x approaches a is L and write

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

if, for each $\epsilon > 0$, there is a $\delta > 0$ such that

$$|f(x) - L| < \epsilon \quad \text{whenever} \quad x \in I \text{ and } 0 < |x - a| < \delta.$$

Note that the condition $0 < |x - a|$ in the above definition means that, in defining the limit of f as x approaches a, we only care about values of f at points of I other than a itself.

Note also, that the domain of f may be larger than I and may not be an interval at all, but, in order to define the limit of f at a we want f to be defined at least at all points, except a itself, in some open interval containing a.

Remark 4.1.2. On comparing the above definition with the definition of continuity (Definition 3.1.1), we conclude that, if f is defined on an open interval containing a, then f is continuous at a if and only if $\lim_{x \to a} f(x) = f(a)$.

This means that if f is not continuous at a (or not defined at a), but it has a limit L as x approaches a, then we can make f continuous at a by redefining (or defining) it at a by setting $f(a) = L$.

Example 4.1.3. Find $\lim_{x \to 1} f(x)$ if $f(x)$ is the function $\dfrac{x^3 - 1}{x - 1}$ on $\mathbb{R} \setminus \{1\}$.

87

Solution: For $x \in \mathbb{R} \setminus \{1\}$, we have

$$f(x) = \frac{x^3 - 1}{x - 1} = x^2 + x + 1.$$

The function on the right is continuous at 1 (since it is a polynomial) and has the value 3 there. Thus, if we extend f to all of $\mathbb{R}$ by giving it the value 3 at $x = 1$, then it becomes the continuous function $x^2 + x + 1$. By the above remark, $\lim_{x \to 1} f(x) = 3$.

Example 4.1.4. Can the function $\dfrac{\sin x}{x}$ on $\mathbb{R} \setminus \{0\}$ be defined at 0 in such a way that it becomes continuous at 0?

   Solution: We learned in calculus that $\lim_{x \to 0} \dfrac{\sin x}{x} = 1$. Thus, if $\dfrac{\sin x}{x}$ is given the value 1 at $x = 0$, it will be continuous there.

One Sided Limits, Limits at $\pm\infty$

Example 4.1.5. Give an intuitive discussion of the behavior of the function $f(x) = x/|x|$ as x approaches 0.
   Solution: We have $f(x) = 1$ if $x > 0$ and $f(x) = -1$ if $x < 0$. Thus, as x approaches 0, $f(x)$ approaches 1 if we keep x to the right of 0, while $f(x)$ approaches -1 if we keep x to the left of 0. However, $\lim_{x \to 0} f(x)$ does not exist, since in the definition of limit, we allow x to be on either side of a.

   The above example suggests that it may be useful to define one-sided limits that depend only on the behavior of the function on one side of the point a. If a function is defined on an unbounded interval, then it may also be useful to discuss its limit at $+\infty$ or $-\infty$. Correctly formulated, the same definition can be used to cover the cases of one sided limits and of limits at $\pm\infty$.

Definition 4.1.6. Let f be a function defined on an open interval (a, b), where a could be $-\infty$ and b could be $+\infty$. We say that the limit from the right of $f(x)$ as x approaches a is L and write

$$\lim_{x \to a^+} f(x) = L$$

if for every $\epsilon > 0$ there is an $m \in (a, b)$ such that

$$|f(x) - L| < \epsilon \quad \text{whenever} \quad a < x < m.$$

   Similarly, we say the limit of $f(x)$ as x approaches b from the left is L, and write

$$\lim_{x \to b^-} f(x) = L$$

if for every $\epsilon > 0$ there is a $m \in (a, b)$ such that

$$|f(x) - L| < \epsilon \quad \text{whenever} \quad m < x < b.$$

Note that, if a is finite, then to say that there is a $m \in (a, b)$ such that $|f(x) - L| < \epsilon$ whenever $a < x < m$ is the same thing as saying there is a $\delta > 0$ such that $|f(x) - L| < \epsilon$ whenever $|x - a| < \delta$ and $x \in (a, b)$ (this is clear if we let m and δ determine each other by the formula $\delta = m - a$). This is just like the ordinary definition of limit of f at a except x is restricted to lie to the right of a. A similar analysis holds for the limit from the left at b in the case where b is finite.

In the case where $b = \infty$, the condition $m < x < b$ just means that $m < x$, while in the case where $a = -\infty$, the condition $a < x < m$ just means that $x < m$. Stated this way, the above definition is the traditional definition of limit at ∞ or at $-\infty$.

For limits at ∞ or $-\infty$, we will simply write " $\lim_{x \to \infty} f(x)$" or " $\lim_{x \to -\infty} f(x)$" rather than " $\lim_{x \to \infty^-} f(x)$" or " $\lim_{x \to -\infty^+} f(x)$" .

In view of the above discussion, the following theorem is almost obvious. Its proof is left to the exercises.

Theorem 4.1.7. *Let I be an open interval and a a point of I. If f is defined on I except possibly at a then*

$$\lim_{x \to a} f(x) = L \quad \text{if and only if} \quad \lim_{x \to a^+} f(x) = L = \lim_{x \to a^-} f(x).$$

In other words the limit of $f(x)$ as x approaches a exists if and only if the limits from the left and the right both exist and are equal. Of course, the limit is then this common value of the limits from the left and right.

Example 4.1.8. For the function

$$f(x) = \begin{cases} 1 - x & \text{if} \quad x < 0 \\ \sin x & \text{if} \quad x > 0, \end{cases}$$

Find $\lim_{x \to 0^-} f(x)$, $\lim_{x \to 0^+} f(x)$, and $\lim_{x \to 0} f(x)$ if they exist.

Solution: Since, to the left of 0, f agrees with the continuous function $1 - x$, its limit from the left is $\lim_{x \to 0}(1 - x) = 1$. On the other hand, to the right of 0, f agrees with the continuous function $\sin x$, and so its limit from the right is $\lim_{x \to 0} \sin x = \sin 0 = 0$. Because the limits from the left and the right are not the same, $\lim_{x \to 0} f(x)$ does not exist.

Example 4.1.9. Find $\displaystyle\lim_{x \to \infty} \frac{x^2 + 3x + 1}{2x^2 - 4}$.

Solution: We do this just as we would if we were finding the limit of a sequence as $n \to \infty$. We divide both numerator and denominator by the highest power of x that occurs. This yields

$$\frac{x^2 + 3x + 1}{2x^2 - 4} = \frac{1 + 3/x + 1/x^2}{2 - 4/x^2}.$$

From this, we guess that the limit is $1/2$. If we want to prove this is true, using only the above definition, we proceed as follows:

$$\left| \frac{x^2 + 3x + 1}{2x^2 - 4} - \frac{1}{2} \right| = \left| \frac{3x + 3}{2x^2 - 4} \right|.$$

Now if $x \geq 3$, then $2x^2 - 4 \geq x^2$ and $3x + 3 < 4x$. In this case, it follows from the above that

$$\left| \frac{x^2 + 3x + 1}{2x^2 - 4} - \frac{1}{2} \right| \leq \frac{4x}{x^2} = \frac{4}{x}.$$

Thus, given $\epsilon > 0$, if we choose $m = \max(3, 4/\epsilon)$, then

$$\left| \frac{x^2 + 3x + 1}{2x^2 - 4} - \frac{1}{2} \right| \leq \frac{4}{x} < \epsilon \quad \text{whenever} \quad m < x.$$

This proves that the limit is $1/2$, as we expected.

Of course, once we prove some theorems about limits, it becomes much easier to do limit problems like the one above. It turns out that all the theorems about limits of sequences, proved in the last chapter, have analogues for limits of functions.

Limit Theorems

As was the case with continuity, the limit of a function can be characterized in terms of limits of sequences. The following theorem is just like Theorem 3.1.5 and is proved the same way. The only difference is that L replaces $f(a)$. We will not repeat the proof

Theorem 4.1.10. *Let (a, b) be a (possibly infinite) interval and let u be a^+ or b^- or a point in the interval (a, b). If f is a function, defined on (a, b), then*

$$\lim_{x \to u} f(x) = L$$

if and only if $f(a_n) \to L$ whenever $\{a_n\}$ is a sequence of points in (a, b), distinct from u, with $a_n \to u$.

As was the case with continuity in section 3.1, this theorem means that each theorem about convergence of sequences yields a theorem about limits of functions. For example, the Main Limit Theorem for sequences, together with the previous theorem implies the Main Limit Theorem for functions:

Theorem 4.1.11. (Main Limit Theorem) *Let (a, b) be a (possibly infinite) interval, let $u = a^+$ or b^- or a point in the interval (a, b), and let c be a constant. Let f and g be functions defined on (a, b). If $\lim_{x \to u} f(x) = K$ and $\lim_{x \to u} g(x) = L$, then*

(a) $\lim_{x \to u} c = c$;

(b) $\lim_{x \to u} cf(x) = cK$;

(c) $\lim_{x \to u} (f(x) + g(x)) = K + L$;

(d) $\lim_{x \to u} f(x)g(x) = KL$;

(e) $\lim_{x \to u} f(x)/g(x) = K/L$, *provided* $L \neq 0$.

There is also a theorem about the limit of a composite function which is similar to Theorem 3.1.10 and has the same proof.

Theorem 4.1.12. *Let* (a, b) *be a (possibly infinite) interval and let* $u = a^+$ *or* b^-. *If* g *is defined on* (a, b) *and* $\lim_{x \to u} g(x) = L$, f *is defined on an interval containing* L *and the image of* g, *and* f *is continuous at* L, *then*

$$\lim_{x \to u} f(g(x)) = f(L).$$

Proof. Let $\{a_n\}$ be a sequence in I converging to u. Then, by Theorem 4.1.10, $\lim_{x \to u} g(x) = L$ implies $g(a_n) \to L$. Then, by Theorem 3.1.5, the continuity of f at L implies that $f(g(a_n)) \to f(L)$. Again using Theorem 4.1.10, we conclude that $\lim_{x \to u} f(g(x)) = f(L)$. □

Example 4.1.13. Prove that if g is a non-negative function, defined on an interval I except possibly at one point $a \in I$, and if $\lim_{x \to a} g(x) = L$, then

$$\lim_{x \to a} g^r(x) = L^r \quad \text{for all rational} \quad r > 0.$$

Solution: If $r > 0$ is rational and we set $f(x) = x^r$, then f is continuous on $[0, \infty)$ by Theorem 3.1.6. Since $g^r(x) = f(g(x))$, it follows immediately from the previous theorem that $\lim_{x \to a} g^r(x) = L^r$.

Infinite Limits

Just as with sequences, for a function f it is sometimes useful to know that, even though f may not have a finite limit as $x \to u$, it does approach either $+\infty$ or $-\infty$. In analogy with Definition 2.4.4, we define infinite limits as follows.

Definition 4.1.14. If f is a function defined on an interval (a, b), then we say $\lim_{x \to a^+} f(x) = \infty$ if, for each M, there is an $m \in (a, b)$ such that

$$f(x) > M \quad \text{whenever} \quad a < x < m.$$

Infinite limits at b^- and what it means for the limit to be $-\infty$ are defined analogously (see the exercises).

If $c \in (a, b)$ and $\lim_{x \to c^-} f(x)$ and $\lim_{x \to c^+} f(x)$ are both ∞, then we write $\lim_{x \to c} f(x) = \infty$. The analogous statement holds if the limits are both $-\infty$.

The following theorem reduces statements about infinite limits to statements about finite limits. Its proof is left to the exercises.

Theorem 4.1.15. *Let f be defined on (a, b) and let $u = a^+$ or b^- or a point in the interval (a, b). If f is positive on (a, b), then*

$$\lim_{x \to u} f(x) = \infty \quad \text{if and only if} \quad \lim_{x \to u} \frac{1}{f(x)} = 0.$$

Similarly, if f is negative on (a, b), then

$$\lim_{x \to u} f(x) = -\infty \quad \text{if and only if} \quad \lim_{x \to u} \frac{1}{f(x)} = 0.$$

Example 4.1.16. Analyze the behavior of $f(x) = \dfrac{x}{1 - x}$ as x approaches 1.

   Solution: We have $\lim_{x \to 1} \dfrac{1}{f(x)} = \lim_{x \to 1} \dfrac{1 - x}{x} = 0$, and so the limits of this function from the left and the right at 1 are both 0. On $(0, 1)$ the function f is positive and so $\lim_{x \to 1^-} f(x) = \infty$ by the previous theorem. On $(1, \infty)$ the function f is negative and so $\lim_{x \to 1^+} f(x) = -\infty$, also by the previous theorem.

Exercise Set 4.1

   In each of the next 6 exercises find the indicated limit and prove that your answer is correct.

1. $\lim_{x \to 1} \dfrac{x^2 - 1}{x - 1}.$

2. $\lim_{x \to 2} \dfrac{x^2 + x - 2}{x - 1}.$

3. $\lim_{x \to 2} \left(\dfrac{x^2 - 4}{x - 2} \right)^{3/2}.$

4. $\lim_{x \to 0} \cos(x^2 - x).$

5. $\lim_{x \to 2} \dfrac{x^2 - 3x + 1}{2x^2 + 1}.$

6. $\lim_{x \to \infty} \dfrac{x^2 - 3x + 1}{2x^2 + 1}.$

7. If $f(x) = \dfrac{\sin x}{|x|}$, find $\lim_{x \to 0^+} f(x)$ and $\lim_{x \to 0^-} f(x)$. Does $\lim_{x \to 0} f(x)$ exist?

8. If $f(x) = \sin 1/x$, do $\lim_{x \to 0^+} f(x)$ and $\lim_{x \to 0^-} f(x)$ exist?

9. If, in Example 4.1.8, f is defined to be $-x$ for $x < 0$ instead of $1 - x$, does $\lim_{x \to 0} f(x)$ exist? Why?

10. Prove Theorem 4.1.7.

11. Let f be defined on a bounded interval (a, b) and let u be a^+, b^- or a point of (a, b). Prove that if $\lim_{x \to u} f(x)$ exists and is positive, then there is a $\delta > 0$ such that $f(x) > 0$ whenever $|x - u| < \delta$ and $x \in (a, b)$. Hint: recall the proof of Theorem 2.2.3.

12. Let f be a non-negative function on an interval (a, b) and let $u = a^+$ or b^-. If $\lim_{x \to u} f(x)$ exists, prove that it is a non-negative number.

13. Prove that if f is a bounded, non-decreasing function on the interval (a, b), then $\lim_{x \to a^+} f(x)$ and $\lim_{x \to b^-} f(x)$ both exist and are finite.

14. State an appropriate definition for the statement $\lim_{x \to b^-} f(x) = -\infty$.

15. Prove Theorem 4.1.15

4.2 The Derivative

The definition of the derivative is familiar from calculus.

Definition 4.2.1. Let f be a function defined on an open interval containing $a \in \mathbb{R}$. If

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

exists and is finite, then we denote it by $f'(a)$, and we say f is differentiable at a with derivative $f'(a)$. If f is defined and differentiable at every point of an open interval I, then we say that f is differentiable on I.

The derivative f' of f is a new function with domain consisting of those points in the domain of f at which f is differentiable.

Remark 4.2.2. When convenient, we will make the change of variables $h = x - a$ and write the derivative in the form

$$f'(a) = \lim_{h \to 0} \frac{f(a + h) - f(a)}{h}. \tag{4.2.1}$$

Equivalently, when it is convenient to use x for the independent variable in the function f', we will write the derivative in the form

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

We don't intend to repeat the computation of the derivatives of all the elementary functions. This is done in calculus. We will assume the student knows how to differentiate polynomials, rational functions, trigonometric functions, inverse trigonometric functions, and exponentials and logarithms. We will, however, compute a couple of derivatives directly from the above definition, just to remind the student of how this is done, and we will occasionally compute a derivative, as an example, to illustrate the use of some theorem.

Example 4.2.3. If $f(x) = x^3$, find the derivative of f using just Definition 4.2.1.

    Solution: We have

$$f'(a) = \lim_{x \to a} \frac{x^3 - a^3}{x - a} = \lim_{x \to a} \frac{(x - a)(x^2 + xa + a^2)}{x - a}$$
$$= \lim_{x \to a} (x^2 + xa + a^2) = 3a^2.$$

Thus, $f'(a) = 3a^2$.

Example 4.2.4. If $f(x) = \sqrt{x}$, find $f'(x)$ for $x > 0$ using just Definition 4.2.1.

    Solution: We have

$$f'(x) = \lim_{h \to 0} \frac{\sqrt{x + h} - \sqrt{x}}{h} = \lim_{h \to 0} \frac{x + h - x}{h(\sqrt{x + h} + \sqrt{x})}$$
$$= \lim_{h \to 0} \frac{1}{\sqrt{x + h} + \sqrt{x}} = \frac{1}{2\sqrt{x}}.$$

Thus, $f'(x) = \dfrac{1}{2\sqrt{x}}$.

Differentiation Theorems

We will use what we know about limits to prove the main theorems concerning differentiation. Some of these are proved in the typical calculus course and some are not.

Theorem 4.2.5. *If f is differentiable at a, then f is continuous at a.*

Proof. If f is defined in an open interval containing a and x, and if $x \neq a$, then

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

We take the limit of both sides as $x \to a$. If f is differentiable at a, then $\lim\limits_{x \to a} \dfrac{f(x) - f(a)}{x - a}$ exists and is finite. Since $\lim_{x \to a}(x - a) = 0$, this implies that $\lim_{x \to a} f(x) = f(a)$. Thus, f is continuous at a. $\qquad\qquad\square$

Theorem 4.2.6. *Let f and g be functions defined on an open interval I containing a and suppose f and g are both differentiable at a and c is a constant. Then cf, $f + g$, fg are differentiable at a, as is f/g provided $g(a) \neq 0$, and*

(a) $(cf)'(a) = cf'(a);$

(b) $(f + g)'(a) = f'(a) + g'(a);$

(c) $(fg)'(a) = f'(a)g(a) + f(a)g'(a);$

(d) $\left(\dfrac{f}{g}\right)'(a) = \dfrac{f'(a)g(a) - f(a)g'(a)}{g^2(a)}.$

Proof. We will prove (c) and (d) and leave (a) and (b) to the exercises.

To prove (c), we write

$$\frac{f(x)g(x) - f(a)g(a)}{x - a} = \frac{f(x) - f(a)}{x - a}g(x) + f(a)\frac{g(x) - g(a)}{x - a} \qquad (4.2.2)$$

By the previous theorem, $\lim_{x \to a} g(x) = g(a)$, and so the Main Limit Theorem implies that the limit of the right side of (4.2.2) as $x \to a$ exists and is equal to $f'(a)g(a) + f(a)g'(a)$. Thus, the limit of the left side of this equality as $x \to a$ exists as well. Hence, $(fg)'(a)$ exists and is equal to $f'(a)g(a) + f(a)g'(a)$.

To prove part (d), we first prove that $1/g$ is differentiable at a and

$$\left(\frac{1}{g}\right)'(a) = -\frac{g'(a)}{g^2(a)}.$$

In fact

$$\frac{1/g(x) - 1/g(a)}{x - a} = \frac{g(a) - g(x)}{g(a)g(x)(x - a)} = \frac{g(a) - g(x)}{x - a}\frac{1}{g(a)g(x)}.$$

If we take the limit of both sides and use the Main Limit Theorem, the conclusion is that $(1/g)'(a)$ exists and is equal to $-\dfrac{g'(a)}{g^2(a)}$, as claimed.

Now part (d) of the theorem follows from the computation

$$\left(\frac{f}{g}\right)'(a) = \left(f\frac{1}{g}\right)'(a) = f'(a)\frac{1}{g(a)} - f(a)\frac{g'(a)}{g^2(a)}$$
$$= \frac{f'(a)g(a) - f(a)g'(a)}{g^2(a)}.$$

$\square$

The Chain Rule

Theorem 4.2.7. *Suppose g is defined in an open interval I containing a and f is defined in an open interval containing $g(I)$. If g is differentiable at a and f is differentiable at $g(a)$, then $f \circ g$ is differentiable at a and*

$$(f \circ g)'(a) = f'(g(a))g'(a).$$

Proof. We let $b = g(a)$ and we define a function h by

$$h(y) = \begin{cases} \dfrac{f(y) - f(b)}{y - b} & \text{if} \quad y \neq b \\ f'(y) & \text{if} \quad y = b. \end{cases}$$

Then, since

$$\lim_{y \to b} \frac{f(y) - f(b)}{y - b} = f'(b),$$

the function h is continuous at $b = g(a)$. Furthermore,

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

Since h is continuous at $b = g(a)$ and g is continuous at a, we conclude that $h(g(x))$ is continuous at $x = a$. Thus, if we take the limit of both sides of the above identity, we conclude that

$$(f \circ g)'(a) = \lim_{x \to a}\frac{f(g(x)) - f(g(a))}{x - a} = h(g(a))\lim_{x \to a}\frac{g(x) - g(a)}{x - a} = f'(g(a))g'(a).$$

$\square$

Example 4.2.8. Find $(\sin \sqrt{x})'$ using the Chain Rule.

  Solution: The derivative of $\sin$ is $\cos$ and the derivative of $\sqrt{x}$ is $\dfrac{1}{2\sqrt{x}}$. Thus, by the Chain Rule,

$$(\sin \sqrt{x})' = (\cos \sqrt{x})\frac{1}{2\sqrt{x}} = \frac{\cos \sqrt{x}}{2\sqrt{x}}.$$

Derivative of an Inverse Function

If f is continuous and strictly monotone on an interval I, then it has a continuous inverse function g, defined on $J = f(I)$, such that $g(J) = I$ (Theorem 3.2.6). If I is an open interval and a is a point of I, then J is also an open interval and $b = f(a) \in J$ (Exercise 4.2.5).

Theorem 4.2.9. *If f is strictly monotone on an open interval I containing a, f is differentiable at a, and $f'(a) \neq 0$, then the inverse function g of f is differentiable at $b = f(a)$ and*

$$g'(b) = \frac{1}{f'(a)} = \frac{1}{f'(g(b))}.$$

Proof. For $y \in J$, we set $x = g(y) \in I$. Then $f(x) = y$. We also have $b = f(a)$ and $a = g(b)$. Then

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

If we denote by h the function of x on the right, then, since f is strictly monotone on I, h is defined everywhere on I except at $x = a$. Since $\lim_{x \to a} h(x) = \dfrac{1}{f'(a)}$, the function h will be defined and continuous at a if we give it the value $\dfrac{1}{f'(a)}$ at $x = a$. Then

$$\frac{g(y) - g(b)}{y - b} = h(g(y)).$$

If we pass to the limit as $y \to b$, then, by Theorem 4.1.12, the expression on the right has limit $h(g(b)) = \dfrac{1}{f'(g(b))}$, since g is continuous at b. This implies the expression on the left has the same limit, which means that $g'(b)$ exists and equals $\dfrac{1}{f'(g(b))}$.   □

Example 4.2.10. Find the derivative of $\sin^{-1}(x)$.

   Solution: The function $\sin x$, when restricted to the domain $[-\pi/2, \pi/2]$ is strictly increasing. Its inverse function $\sin^{-1}(x)$ is also increasing and has domain $[-1,1]$ – the image of $[-\pi/2, \pi/2]$ under $\sin$. Thus, $\sin^{-1}$ has a non-negative derivative on $(-1,1)$ and by Theorem 4.2.9, it is given by

$$(\sin^{-1} x)' = \frac{1}{\cos(\sin^{-1} x)} = \frac{1}{\sqrt{1 - \sin^2(\sin^{-1} x)}} = \frac{1}{\sqrt{1 - x^2}},$$

since $\sin(\sin^{-1} x) = x$.

Exercise Set 4.2

1. Using just the definition of the derivative, show that the derivative of $1/x$ is $-1/x^2$.

2. Using just the definition of the derivative, find $(x^2 + 3x)'$.

3. Show how to derive the expression for the derivative of $\tan x$ if you know the derivatives of $\sin x$ and $\cos x$.

4. Using theorems from this section, find the derivative of $\tan\left(\dfrac{x}{x^2 + 1}\right)$.

5. We know that the image of a closed interval under a continuous function is a closed interval or a point (Theorem 3.2.4). Show that the image of an open interval under a continuous, strictly monotone function is an open interval.

6. If $f \circ g \circ h(x) = f(g(h(x)))$ is the composition of three functions, find an expression for its derivative. You may use the Chain Rule.

7. Using Theorem 4.2.9, derive the expression for the derivative of $\sqrt{x}$.

8. Using Theorem 4.2.9, derive the expression for the derivative of $\tan^{-1} x$.

9. Prove that if f is defined on an open interval I and has a positive derivative at a point $a \in I$, then there is an open interval J, containing a and contained in I, such that $f(x) < f(a) < f(y)$ whenever $x, y \in J$ and $x < a < y$. Hint: see Exercise 4.1.11.

10. If f is a monotone function on an interval and g is its inverse function, then

$$f \circ g(y) = y$$

for every y in the domain J of g. Use the Chain Rule on this identity to derive the expression for the derivative of the inverse function g. This argument is not a substitute for the proof in Theorem 4.2.9. Why?

11. Is the function defined by

$$f(x) = \begin{cases} x \sin 1/x & \text{if} \quad x \neq 0 \\ 0 & \text{if} \quad x = 0 \end{cases}$$

differentiable at 0? How about the function

$$f(x) = \begin{cases} x^2 \sin 1/x & \text{if} \quad x \neq 0 \\ 0 & \text{if} \quad x = 0 \end{cases}?$$

12. Is the function defined by

$$f(x) = \begin{cases} x^2 & \text{if} \quad x > 0 \\ 0 & \text{if} \quad x \leq 0 \end{cases}$$

differentiable at 0?

4.3   The Mean Value Theorem

Critical Points

The proof of the Mean Value Theorem rests on the fact that a continuous function on a closed bounded interval $[a, b]$ takes on its maximum and minimum values only at critical points. A *critical point* for f on $[a, b]$ is a point $c \in [a, b]$ which satisfies one of the following:

1. c is an endpoint (a or b);

2. c is a stationary point, meaning $c \in (a, b)$ and $f'(c) = 0$; or

3. c is a singular point, meaning $c \in (a, b)$ and $f'(c)$ does not exist.

Theorem 4.3.1. *If f is a continuous function on a closed bounded interval $[a, b]$ and $c \in [a, b]$ is a point at which f assumes a maximum or a minumum value on $[a, b]$, then c is a critical point for f on $[a, b]$.*

Proof. Assume f has a maximum at c. The proof in the case where it has a minimum is the same, except that the inequalities reverse.

   We will prove that if c is not an endpoint or a singular point, then it must be a stationary point. This implies that it has to be one of the three.

If c is not an endpoint and not a singular point, then $a < c < b$ and f has a derivative at c. Since $f(x) \le f(c)$ for all $x \in [a, b]$, we have

$$\frac{f(x) - f(c)}{x - c} \begin{cases} \le 0 & \text{for } x > c, \\ \ge 0 & \text{for } x < c. \end{cases}$$

It follows from Exercise 4.1.12 that

$$\lim_{x \to c^+} \frac{f(x) - f(c)}{x - c} \le 0 \quad \text{and} \quad \lim_{x \to c^-} \frac{f(x) - f(c)}{x - c} \ge 0.$$

Since these two one-sided limits must be equal if the limit itself exists, we conclude that the limit must be 0. That is, $f'(c) = 0$. Hence c is a stationary point. $\qquad \square$

The Mean Value Theorem

The Mean Value Theorem is one of the most heavily used tools of calculus. It says that if f is continuous on $[a, b]$ and differentiable on (a, b), then for at least one point between a and b the graph of f has tangent line parallel to the line joining $(a, f(a))$ to $(b, f(b)$; this may happen at several points (see Figure 4.1). More precisely,

Theorem 4.3.2. *If a function f is continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b), then there is at least one point $c \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}. \tag{4.3.1}$$

Proof. The function whose graph is the line joining $(a, f(a))$ to $(b, f(b))$ is

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

If we subtract this from f the result is the function s, where

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

The function s is also continuous on $[a, b]$ and differentiable on (a, b). By Theorem 3.2.1, s assumes both a maximum value and a minimum value on $[a, b]$. However,

$$s(a) = s(b) = 0,$$

and so s is either identically zero or it assumes a non-zero maximum or a non-zero minimum on (a, b). In each of these cases, s has a critical point in (a, b). Let c be such a critical point. Since s is differentiable on (a, b), c must be a point at which s' is 0. Thus,

$$s'(c) = f'(c) - \frac{f(b) - f(a)}{b - a} = 0,$$

which implies that c satisfies (4.3.1). $\qquad \square$
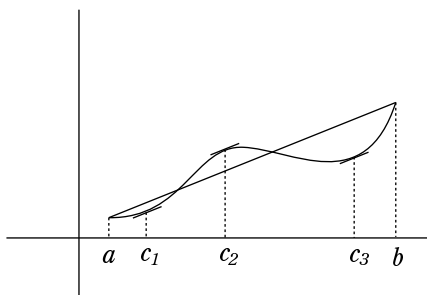
Figure 4.1: Three choices for the c in the Mean Value Theorem.

The Mean Value Theorem has a wide variety of applications. Many of the frequently used facts that we take for granted in calculus are direct consequences of this theorem. It is also used to prove many new facts that go beyond standard calculus material.

Functions with Vanishing Derivative

Theorem 4.3.3. *If f is a differentiable function on an open interval (a, b) and f' is identically 0 on (a, b), then f is a constant.*

Proof. let x, y be any two points of (a, b) with $x < y$. Then the Mean Value Theorem implies that there is a number c between x and y such that

$$f'(c) = \frac{f(y) - f(x)}{y - x}.$$

Since $f'(c) = 0$, this implies that $f(x) - f(y) = 0$, or $f(x) = f(y)$. Thus, f has the same value at any two points of (a, b) and this means that it is constant. ☐

Corollary 4.3.4. *If f and g are differentiable functions on (a, b) and $f'(x) = g'(x)$ for all $x \in (a, b)$, then there is a constant c such that $f(x) = g(x) + c$ on (a, b).*

Proof. We apply the previous theorem to $f - g$. ☐

Another way to say this corollary is: If a function h has an antiderivative on (a, b), then any two of its antiderivatives differ by a constant. We use this fact all the time in integration theory.

Monotone Functions

Theorem 4.3.5. *If f is a function which is continuous on a closed interval $[a, b]$ and differentiable on the open interval (a, b), then f is strictly increasing on $[a, b]$ if $f'(x) > 0$ for all $x \in (a, b)$, while f is strictly decreasing on $[a, b]$ if $f'(x) < 0$ for all $x \in (a, b)$.*

Proof. If x and y are any two points of $[a, b]$ with $x < y$, then the Mean Value Theorem tells us there is a $c \in (x, y) \subset (a, b)$ at which

$$f'(c) = \frac{f(y) - f(x)}{y - x}.$$

Since the denominator is positive, this means that $f'(c)$ and $f(y) - f(x)$ have the same sign. This implies that f is strictly increasing (resp. decreasing) on $[a, b]$ if $f'(c)$ is positive (resp. negative) for all $c \in (a, b)$. $\square$

This is the basis for the familiar graphing technique which uses the sign of the derivative of f to determine intervals on which f is increasing or decreasing.

The converse of Theorem 4.3.5 is not true, since a function which is strictly increasing on an interval (a, b) can have a derivative that is 0 at some points of (a, b) (for example, $f(x) = x^3$ is strictly increasing on $(-\infty, +\infty)$, but its derivative is 0 at 0). However, the following related result is an "if and only if" Theorem. Its proof is left to the exercises.

Theorem 4.3.6. *Let f be a continuous function on $[a, b]$ which is differentiable on (a, b). Then f is non-decreasing on $[a, b]$ if and only if $f'(x) \geq 0$ for all $x \in (a, b)$, while if f is non-increasing on $[a, b]$ if and only if $f'(x) \leq 0$ for all $x \in (a, b)$.*

Example 4.3.7. Find the intervals on which the function $f(x) = x^3 - 3x + 5$ is increasing, decreasing.

Solution: The derivative of f is $f'(x) = 3x^2 - 3 = 3(x - 1)(x + 1)$. This function is positive for $x > 1$ and $x < -1$ and is negative for $-1 < x < 1$. Thus, by Theorem 4.3.5, f is increasing on $(-\infty, -1]$ and $[1, +\infty)$ and it is decreasing on $[-1, 1]$.

Example 4.3.8. Prove that $\sin x < x$ for all $x > 0$.

Solution: Let $f(x) = x - \sin x$. Then $f(0) = 0$ and $f'(x) = 1 - \cos x \geq 0$ for all x. In fact, $f'(x) > 0$ except at multiples of 2π. By Theorem 4.3.5, f is increasing on $[0, 2\pi]$. Since it is 0 at $x = 0$, it must be positive on $(0, 2\pi]$. Thus, $\sin x < x$ for $x \in (0, 2\pi]$. It is obvious that $\sin x < x$ for $x > 2\pi$ (since $\sin x \leq 1$ for all x).

Uniform Continuity

We know that a continuous function on a closed, bounded interval I is uniformly continuous. If the interval I is not closed or not bounded, then continuous functions on I need not be uniformly continuous. However, we have the following application of the Mean Value Theorem:

Theorem 4.3.9. *If f is a differentiable function on a (possibly infinite) open interval (a, b), and if f' is bounded on (a, b), then f is uniformly continuous on (a, b).*

Proof. Let M be an upper bound for $|f'|$ on (a, b). Then $|f'(x)| \leq M$ for all $x \in (a, b)$. By the Mean Value Theorem, if $x, y \in (a, b)$, then there is a c between x and y such that

$$\frac{f(x) - f(y)}{x - y} = f'(c).$$

If we take the absolute value of both sides and multiply by $|x - y|$ and , this yields

$$|f(x) - f(y)| = |f'(c)||x - y| \leq M|x - y|.$$

Thus, given $\epsilon > 0$, if we choose $\delta = \epsilon/M$, then

$$|f(x) - f(y)| \leq \epsilon \quad \text{whenever} \quad |x - y| < \delta.$$

This proves that f is uniformly continuous on (a, b). $\square$

Exercise Set 4.3

1. If f is a continuous function on $[-1, 1]$ which is differentiable on $(-1, 1)$ and satisfies $f(-1) = 0$, $f(0) = 0$, and $f(1) = 1$, then show that f' takes on the values $0, 1/2$, and 1 on $[-1, 1]$.

2. Prove that $|\sin x - \sin y| \leq |x - y|$ for all $x, y \in \mathbb{R}$.

3. If $r > 0$ prove that $\ln y - \ln x \leq \dfrac{y - x}{r}$ if $r \leq x \leq y$.

4. Suppose f is a continuous function on $[0, \infty)$ which is differentiable on $(0, \infty)$. If $f(0) = 0$ and $|f'(x)| \leq M$ for all $x \in (0, \infty)$, then prove that $|f(x)| \leq Mx$ on $[0, \infty)$.

5. Prove that if f is a differentiable function on $(0, \infty)$ and f and f' both have finite limits at ∞, then $\lim_{x \to \infty} f'(x) = 0$. Hint: apply the Mean Value Theorem to f for large values of a and b.

6. If $f(x) = 2x^3 + 3x^2 - 12x + 5$, find the intervals on which f is increasing and those on which it is decreasing.

7. Prove that $\ln x \leq x - 1$ for all $x > 0$. Hint: analyze where $x - 1 - \ln x$ is increasing and where it is decreasing.

8. Find where $e^{-x} x^e$ is increasing and where it is decreasing. Which is bigger e^π or π^e?

9. Prove Theorem 4.3.6.

10. Suppose f is a differentiable function on an interval (a, b) and that f' takes on both positive and negative values on (a, b). Prove that f' must take on the value 0 as well. Hint: show that if $f'(x) > 0$ and $f'(y) < 0$ for points x, y with $a < x < y < b$, then the maximum of f on $[x, y]$ occurs at some point strictly between x and y; the same argument will show that if $f'(x) < 0$ and $f'(y) > 0$, then the minimum of f on $[x, y]$ occurs at a point strictly between x and y.

11. Use the result of the previous exercise to show that, if f is differentiable on (a, b) and f' takes on two values c and d on (a, b), then it take on every value between c and d. This is the Intermediate Value Theorem for Derivatives. Note that we do not assume f' is continuous on $[a, b]$.

12. Let f be differentiable on $\mathbb{R}$. Prove that, if there is an $r < 1$ such that $|f'(x)| \leq r$ for all $x \in \mathbb{R}$, then $|f(x) - f(y) \leq r|x - y|$ for all $x, y \in \mathbb{R}$. A function with this property is called a *contraction mapping*.

13. Let f satisfy the conditions of the previous exercise. Show there is a fixed point for f – that is, an $x \in \mathbb{R}$ such that $f(x) = x$. Hint: construct a sequence $\{x_n\}$ inductively by setting $x_1 = 0$ and $x_{n+1} = f(x_n)$. Show that this sequence is Cauchy and that it converges to a fixed point for f.

14. Prove that if f is increasing on $[a, b]$ and on $[b, c]$, then f is also increasing on $[a, c]$.

15. The following is a partial converse to Theorem 4.3.9: Prove that if f is differentiable on a, possibly infinite, interval (a, b) and if $\lim\limits_{x \to b} f'(x) = \infty$, then f is not uniformly continuous on (a, b). The same conclusion holds if $\lim\limits_{x \to a} f'(x) = \infty$.

16. Show that $\ln x$ is uniformly continuous on $[1, \infty)$, but not on $(0, 1]$.

4.4   L'Hôpital's Rule

In this section we prove the familiar L'Hôpital's Rule – a tool from calculus, useful in calculating limits of indeterminate forms. It has two forms, depending on whether the indeterminate form is of type $0/0$ or of type ∞/∞. The proof uses the following generalization of the Mean Value Theorem.

Cauchy's Mean Value Theorem

Theorem 4.4.1. *Let f and g be functions which are continuous on a closed, bounded interval $[a, b]$ and differentiable on (a, b). Assume that $g'(x) \neq 0$ for all $x \in (a, b)$. Then there exists $c \in (a, b)$ such that*

$$\frac{f(b) - f(a)}{g(b) - g(a)} = \frac{f'(c)}{g'(c)}. \tag{4.4.1}$$

Proof. We begin by observing that g is strictly monotone on $[a, b]$. This follows from the fact that g' is never 0 on (a, b). If it is never 0, then it cannot take on both positive and negative values on (a, b) (Exercise 4.3.10). Thus, it is always positive or always negative, and this implies that it is strictly monotone on $[a, b]$. In particular, $g(b) \neq g(a)$.

The proof now follows the same strategy as the proof of the ordinary Mean Value Theorem (Theorem 4.3.2). The only difference is that $x - a$ and $b - a$

are replaced by $g(x) - g(a)$ and $g(b) - g(a)$ in the definition of the function s. Thus, we set

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

Note that s is continuous on $[a, b]$ and differentiable on (a, b). By Theorem 3.2.1, s assumes both a maximum value and a minimum value on $[a, b]$. However,

$$s(a) = s(b) = 0,$$

and so s is either identically zero or it assumes a non-zero maximum or a non-zero minimum on (a, b). In any of these cases, s has a critical point in (a, b). Let c be such a critical point. Since s is differentiable on (a, b), c must be a point at which s' is 0. Thus,

$$s'(c) = f'(c) - \frac{f(b) - f(a)}{g(b) - g(a)}g'(c) = 0,$$

which implies that c satisfies (4.4.1). $\square$

Example 4.4.2. Prove that $|\cos x - 1| \leq \dfrac{x^2}{2}$ for all x.

   Solution: We use Cauchy's Mean Value Theorem with $f(x) = \cos x$ and $g(x) = x^2$. It implies that there is c between 0 and x such that

$$\frac{\cos x - 1}{x^2} = \frac{\cos x - \cos 0}{x^2 - 0^2} = \frac{-\sin c}{2c}.$$

Since $|\sin c| \leq |c|$ by Exercise 4.3.2, this implies that

$$\left| \frac{\cos x - 1}{x^2} \right| \leq \frac{1}{2},$$

which implies that $|\cos x - 1| \leq \dfrac{x^2}{2}$.

L'Hôpital's Rule

The problem of finding

$$\lim_{x \to 1} \frac{\ln x}{x^2 - 1}$$

cannot be attacked by using the part of the Main Limit Theorem which deals with limits of quotients, because the limit of the denominator is 0. In fact, both numerator and denominator have limit 0. A limit problem of this type is called a 0/0 form.

   Similarly, the problem of finding

$$\lim_{x \to \infty} \frac{e^x}{x^2}$$

cannot be attacked using the limit of quotients part of the Main Limit Theorem. This time the problem is that both numerator and denominator have limit $+\infty$. A limit problem of this type is called an ∞/∞ form.

Problems of this type can often be solved by using the following theorem.

Theorem 4.4.3. (L'Hôpital's Rule) *Let f and g be differentiable functions on a (possibly infinite) interval (a, b) and let u stand for a^+ or b^-. Suppose, $g(x)$ and $g'(x)$ are non-zero on all of (a, b) and*

1. $\lim_{x \to u} f(x) = 0 = \lim_{x \to u} g(x)$, or

2. $\lim_{x \to u} f(x) = \infty = \lim_{x \to u} g(x)$.

Then

$$\lim_{x \to u} \frac{f(x)}{g(x)} = \lim_{x \to u} \frac{f'(x)}{g'(x)}, \tag{4.4.2}$$

provided the limit on the right exists.

Proof. We will present the proof in the case where $u = a^+$ and the limit on the right in (4.4.2) is a finite number L. The case where this limit is infinite can be reduced to the finite case (Exercise 4.4.16). The proof in the case $u = b^-$ is entirely analogous.

If $x, y \in (a, b)$, then Cauchy's Mean Value Theorem tells us that there is a c between x and y such that

$$f(x) - f(y) = (g(x) - g(y))\frac{f'(c)}{g'(c)},$$

or

$$\frac{f(x)}{g(x)} = \frac{f(y)}{g(x)} + \left(1 - \frac{g(y)}{g(x)}\right)\frac{f'(c)}{g'(c)}$$

On subtracting L and performing some algebra, this becomes

$$\frac{f(x)}{g(x)} - L = \frac{f(y)}{g(x)} + \left(1 - \frac{g(y)}{g(x)}\right)\left(\frac{f'(c)}{g'(c)} - L\right) - L\frac{g(y)}{g(x)}.$$

On applying the triangle inequality, this yields

$$\left|\frac{f(x)}{g(x)} - L\right| \le \left|\frac{f(y)}{g(x)}\right| + \left(1 + \left|\frac{g(y)}{g(x)}\right|\right)\left|\frac{f'(c)}{g'(c)} - L\right| + \left|L\frac{g(y)}{g(x)}\right|. \tag{4.4.3}$$

Given $\epsilon > 0$, we will show how to make each of the terms on the right in this inequality be less than $\epsilon/3$ by choosing x sufficiently close to a.

At this point the proof splits into two cases, depending on whether hypothesis (1) or (2) holds. If (1) holds, then since $\lim_{x \to a^+} f'(x)/g'(x) = L$, Definition 4.1.6 tells us there is an $m \in (a, b)$ so that

$$\left|\frac{f'(c)}{g'(c)} - L\right| < \epsilon/6 \tag{4.4.4}$$

whenever $a < c < m$. This condition will be satisfied if x is any number with $a < x < m$ and y any number with $a < y < x$ (since c is between x and y). Now, given any x, we can choose a y (depending on x) so that $a < y < x$ and

$$\left| \frac{f(y)}{g(x)} \right| < \frac{\epsilon}{3}, \quad \text{and} \tag{4.4.5}$$

$$\left| \frac{g(y)}{g(x)} \right| < \min\left(1, \frac{\epsilon}{3|L|}\right). \tag{4.4.6}$$

This is possible because $\lim_{y \to a^+} f(y) = 0 = \lim_{y \to a^+} g(y)$ holds by hypothesis (1). Taken together, inequalities (4.4.3) through (4.4.6) imply that

$$\left| \frac{f(x)}{g(x)} - L \right| < \epsilon \quad \text{whenever} \quad a < x < m.$$

This implies that $\lim\limits_{x \to a^+} \dfrac{f(x)}{g(x)} = L$ and completes the proof in the case where (1) holds.

In the case where hypothesis (2) holds, the proof is almost the same. We still use (4.4.3), but the order in which x, y, and m are chosen changes and x and y reverse positions in the inverval (a, b). We first choose y such that (4.4.4) holds whenever $a < c < y$. This is possible because $\lim_{c \to a^+} f'(c)/g'(c) = L$.

We then choose $m \in (a, y)$ in such a way that (4.4.5) and (4.4.6) hold whenever $a < x < m$. This is possible because $\lim_{x \to a^+} g(x) = \infty$ holds by hypothesis (2). Because $m < y$, such a choice of x will force $x < y$ and, hence, $c < y$ (again, since c is between x and y).

As before, inequalities (4.4.3) through (4.4.6) imply that

$$\left| \frac{f(x)}{g(x)} - L \right| < \epsilon \quad \text{whenever} \quad a < x < m.$$

This implies that $\lim\limits_{x \to a^+} \dfrac{f(x)}{g(x)} = L$ and completes the proof in the case where (2) holds. $\square$

Example 4.4.4. Find $\lim\limits_{x \to 1} \dfrac{\ln x}{x^2 - 1}$.

Solution: This is a 0/0 form since $\lim_{x \to 1} \ln x = 0 = \lim_{x \to 1}(x^2 - 1)$. If we differentiate numerator and denominator, and take the limit of the resulting fraction, we get

$$\lim_{x \to 1} \frac{1/x}{2x} = \frac{1}{2}.$$

We conclude that

$$\lim_{x \to 1} \frac{\ln x}{x^2 - 1} = \frac{1}{2}$$

as well.

Example 4.4.5. Find $\lim\limits_{x\to\infty} \dfrac{x^2}{e^x}$.

Solution: This is an ∞/∞ form since $\lim_{x\to\infty} e^x = \infty = \lim_{x\to\infty} x^2$. If we differentiate numerator and denominator, and take the limit of the resulting fraction, we get

$$\lim_{x\to\infty} \frac{2x}{e^x}.$$

This is still an ∞/∞ form. If we again differentiate numerator and denominator and pass to the limit, we get

$$\lim_{x\to\infty} \frac{2}{e^x} = 0.$$

We conclude from L'Hôpital's Rule that

$$\lim_{x\to\infty} \frac{2x}{e^x} = 0,$$

and, hence, that

$$\lim_{x\to\infty} \frac{x^2}{e^x} = 0.$$

Example 4.4.6. Find $\lim_{n\to\infty}(1 + r/n)^n$.

Solution: This is the limit of a sequence. However, we may compute this limit by replacing the integer valued variable n with the real valued variable x. If we find that $\lim_{x\to\infty}(1+r/x)^x$ has a limit, then $\lim_{n\to\infty}(1+r/n)^n$ must have the same limit.

We set $f(x) = (1 + r/x)^x$, $g(x) = \ln(f(x)) = x\ln(1 + r/x)$, and $y = 1/x$. Then

$$\lim_{x\to\infty} g(x) = \lim_{y\to 0} g(1/y) = \lim_{y\to 0} \frac{\ln(1 + ry)}{y}.$$

This is a $0/0$ form and L'Hôpital's Rule implies that this limit is

$$\lim_{y\to 0} \frac{r}{1 + ry} = r.$$

Then

$$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} e^{g(x)} = e^r.$$

by Theorem 4.1.12.

Exercise Set 4.4

1. Prove that if $r > 0$ and $x > 1$, then $\ln x \le \dfrac{x^r - 1}{r}$. Hint: use Cauchy's form of the Mean Value Theorem with $f(x) = \ln x$ and $g(x) = x^r$.

2. Prove that $|\sin x - x| \le \dfrac{1}{6}|x|^3$.

3. Prove that $1 + x^2 \le e^{x^2}$ for all $x \in \mathbb{R}$.

4. If f is a function which is differentiable on an open interval I containing 0 and if $f(0) = 0$, then prove that there is a c between 0 and x at which

$$f(x) = \frac{f'(c)}{c^{n-1}} \frac{x^n}{n}.$$

Hint: apply the Cauchy Mean Value Theorem to $f(x)$ and $g(x) = x^n$.

5. Use the previous exercise and induction to prove that if f is n-times differentiable on an open interval I containing 0 and if the kth derivative, $f^{(k)}$ of f is 0 at 0 for $k = 0, 1, \cdots, n-1$, then there is a c between 0 and x at which

$$f(x) = f^{(n)}(c)\frac{x^n}{n!}.$$

Find each of the following limits if they exist:

6. $\lim\limits_{x \to \infty} \dfrac{\ln x}{x^r}$ where $r > 0$.

7. $\lim\limits_{x \to 0} x \ln x$.

8. $\lim\limits_{x \to 0} \dfrac{\sin x - x}{x^3}$.

9. $\lim\limits_{x \to 0} \dfrac{1 + \cos x}{x^2}$.

10. $\lim\limits_{x \to 0} x^x$.

11. $\lim\limits_{x \to \infty} x^{1/x}$.

12. $\lim\limits_{x \to \infty} (\sqrt{x+1} - \sqrt{x})$.

13. $\lim\limits_{n \to \infty} \dfrac{\ln n}{\sqrt{n}}$

14. Let f be a differentiable function on $(0, \infty)$. Prove that if $\lim\limits_{x \to \infty} f(x) = \infty$ and $\lim\limits_{x \to \infty} f'(x) = L$, then

$$\lim_{x \to \infty} \frac{e^{f(x)}}{\int_0^x e^{f(t)} \, dt} = L.$$

15. let f be a differentiable function on an interval of the form $(a, +\infty)$. Prove that if there is a number $r \neq 0$ such that $\lim_{x \to \infty} (rf'(x) + f(x)) = L$ is finite, then $\lim_{x \to \infty} f'(x) = 0$ and $\lim_{x \to \infty} f(x) = L$. Hint: apply L'Hôpital's Rule to $\dfrac{e^{x/r} f(x)}{e^{x/r}}$.

16. Finish the proof of Theorem 4.4.3 by showing that if the theorem is true whenever $\lim_{x \to u} f'(x)/g'(x)$ is finite, then it is also true whenever this limit is infinite.

Chapter 5

The Integral

In this chapter we define the Riemann integral and develop its most important properties. We also prove the Fundamental Theorem of Calculus and discuss improper integrals.

5.1   Definition of the Integral

If $[a, b]$ is a closed, bounded interval, then a *partition* P of $[a, b]$ is a finite, ordered set of points

$$P = \{a = x_0 < x_1 < \cdots < x_n = b\}$$

of $[a, b]$, beginning with a and ending with b. Such a set of points has the effect of dividing $[a, b]$ into a collection of n subintervals

$$[x_0, x_1], [x_1, x_2], \cdots, [x_{n-1}, x_n].$$

Given a partition P, as above, of $[a, b]$ and a bounded function f, defined on $[a, b]$, a *Riemann Sum* for f and P on $[a, b]$ is a sum of the form

$$\sum_{k=1}^{n} f(\bar{x}_k)(x_k - x_{k-1}) \tag{5.1.1}$$

where, for each k, $\bar{x}_k$ is some point in the interval $[x_{k-1}, x_k]$. For each k, the term $f(\bar{x}_k)(x_k - x_{k-1})$ represents the area (or minus the area, if $f(\bar{x}_k) < 0$) of a rectangle with width $x_k - x_{k-1}$ and with height $|f(\bar{x}_k)|$ (see Figure 5.1).

In calculus, the Riemann Integral of f is defined as a limit of Riemann sums, although how this limit is defined and how one shows that it actually exists for a reasonable class of functions are things that are usually left for a more advanced course. This is that course.

Here we will give a precise definition of the integral and prove that it exists for a large class of functions on $[a, b]$. In particular, we will prove that the integral of every continuous function on $[a, b]$ exists.
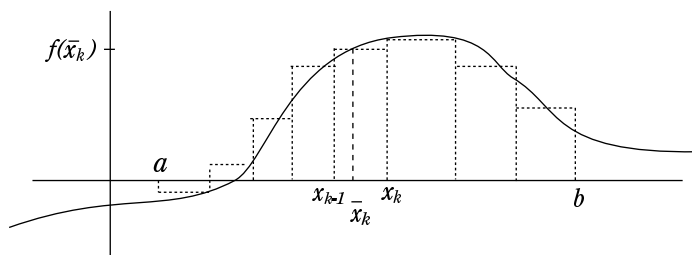
Figure 5.1: A Riemann Sum.

Upper and Lower Sums

Given a partition $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ of $[a, b]$ and a bounded function f on $[a, b]$, we can write down two sums which have every Riemann sum for this partition and this function trapped in between them. These are the upper and lower sums for P and f:

Definition 5.1.1. Given a partition P and function f, as above, for $k = 1, \cdots, n$, we set

$$M_k = \sup\{f(x) : x \in [x_{k-1}, x_k]\} \quad \text{and} \quad m_k = \inf\{f(x) : x \in [x_{k-1}, x_k]\}.$$

Then the *upper sum* for f and P is

$$U(f, P) = \sum_{k=1}^{n} M_k(x_k - x_{k-1}), \tag{5.1.2}$$

while the *lower sum* for f and P is

$$L(f, P) = \sum_{k=1}^{n} m_k(x_k - x_{k-1}). \tag{5.1.3}$$

Now, for any choice of $\bar{x}_k \in [x_{k-1}, x_k]$, we have

$$m_k \leq f(\bar{x}_k) \leq M_k.$$

This inequality remains true if we multiply through by the positive number $(x_k - x_{k-1})$. On summing the resulting inequalities, we conclude that

$$L(f, P) \leq \sum_{k=1}^{n} f(\bar{x}_k)(x_k - x_{k-1}) \leq U(f, P). \tag{5.1.4}$$

Thus, the upper sum $U(f, P)$ is an upper bound for all Riemann sums for f and P and the lower sum is a lower bound for all these sums. In fact, it is not hard to prove that $U(f, P)$ is the least upper bound for all Riemann sums for f and P, while $L(f, P)$ is the greatest lower bound of this set (Exercise 5.1.6).

Example 5.1.2. Find the upper sum and lower sum for the function $f(x) = x^2$ and the partition $P = \{0 < 1/4 < 1/2 < 3/4 < 1\}$ of the interval $[0,1]$.

Solution: The function f is increasing on $[0,1]$ and so its sup on each subinterval is achieved at the right endpoint of the interval and its inf is achieved at the left endpoint. Thus,

$$L(f,P)$$
$$= 0(1/4 - 0) + 1/16(1/2 - 1/4) + 1/4(3/4 - 1/2) + 9/16(1 - 3/4) = \frac{7}{32}$$

while

$$U(f,P)$$
$$= 1/16(1/4 - 0) + 1/4(1/2 - 1/4) + 9/16(3/4 - 1/2) + 1(1 - 3/4) = \frac{15}{32}.$$

Refinement of Partitions

It is useful to think of a partition of $[a,b]$ as simply a finite subset of $[a,b]$ that contains a and b. The elements of this finite set are then given labels $x_0, x_1, \cdots, x_n$ which are consistent with the order in which these elements occur in $[a,b]$. Thus, $a = x_0 < x_1 < \cdots < x_n = b$. To think of partitions as subsets of $[a,b]$ allows us to use set theoretic relations and operations such as "$\subset$" and "$\cup$" on them.

Definition 5.1.3. Let P and Q be partitions of a closed bounded interval $[a,b]$. Then we say that Q is a *refinement* of P if $P \subset Q$.

For example, the partition $0 < 1/4 < 1/3 < 1/2 < 2/3 < 3/4 < 1$ is a refinement of the partition $0 < 1/4 < 1/2 < 3/4 < 1$.

Theorem 5.1.4. *Let f be a bounded function on a closed bounded interval $[a,b]$. If Q and P are partitions of $[a,b]$ and Q is a refinement of P, then*

$$L(f,P) \leq L(f,Q) \leq U(f,Q) \leq U(f,P). \tag{5.1.5}$$

Proof. We will prove this in the case where Q is obtained from P by adding just one additional point to P. The general case then follows from this using an induction argument on the number of additional points needed to get from P to Q (Exercise 5.1.7).

Suppose $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ and Q is obtained by adding one point y to P. Suppose this new point lies between x_{j-1} and x_j. Then, in passing from P to Q, the subinterval $[x_{j-1}, x_j]$ is cut into the two subintervals $[x_{j-1}, y]$ and $[y, x_j]$, while all other subintervals $[x_{k-1}, x_k]$ $(k \neq j)$ remain the same. Thus, in the formulas (5.1.2) and (5.1.1) for the upper and lower sums, the terms for which $k \neq j$ are unchanged when we pass from P to Q. To prove the theorem, we just need to analyze what happens to the jth terms in (5.1.2) and (5.1.1) when P is replaced by Q.

With m_j and M_j as in Definition 5.1.1 for the partition P, we set

$$m'_j = \inf\{f(x) : x \in [x_{j-1}, y]\}, \quad M'_j = \sup\{f(x) : x \in [x_{j-1}, y]\},$$
$$m''_j = \inf\{f(x) : x \in [y, x_j]\}, \quad M''_j = \sup\{f(x) : x \in [y, x_j]\}.$$

Then $m_j = \min\{m'_j, m''_j\}$ and $M_j = \max\{M'_j, M''_j\}$, and so

$$m_j(x_j - x_{j-1}) = m_j(y - x_{j-1}) + m_j(x_j - y)$$
$$\leq m'_j(y - x_{j-1}) + m''_j(x_j - y),$$

while

$$M'_j(y - x_{j-1}) + M''_j(x_j - y)$$
$$\leq M_j(y - x_{j-1}) + M_j(x_j - y) = M_j(x_j - x_{j-1}).$$

Now (5.1.5) follows from this and the fact that the other terms in the sums defining $U(f, P)$ and $L(f, P)$ are unchanged when P is replaced by Q. □

Note that any two partitions P and Q of an interval $[a, b]$ have a common refinement. In fact, the set theoretic union $P \cup Q$ is a common refinement of P and Q. This, together with the preceding result leads to the following theorem, which says that every lower sum is less than or equal to every upper sum.

Theorem 5.1.5. *If P and Q are any two partitions of a closed bounded interval $[a, b]$ and f is a bounded function on $[a, b]$, then*

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

Proof. We simply apply the previous theorem to P and its refinement $P \cup Q$ and to Q and its refinement $P \cup Q$. This yields

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

□

The Integral

Given a closed bounded interval $[a, b]$ and a bounded function f on $[a, b]$, we set

$$\overline{\int}_a^b f \, dx = \inf\{U(f, Q) : Q \text{ a partition of } [a, b]\},$$
$$\underline{\int}_a^b f \, dx = \sup\{L(f, Q) : Q \text{ a partition of } [a, b]\}.$$

We will call these the *upper integral* and *lower integral*, respectively, of f on $[a, b]$. Theorem 5.1.5 says that every lower sum for f is less than or equal to every upper sum for f. Thus, each upper sum $U(f, P)$ is an upper bound for

the set of all lower sums. Hence, it is at least as large as the least upper bound of this set; that is

$$\underline{\int}_a^b f \, dx \le U(f, P) \quad \text{for all partitions} \quad P \text{ of } [a, b].$$

This, in turn, means that $\underline{\int}_a^b f \, dx$ is a lower bound for the set of all upper sums and, hence, is less than or equal to the greatest lower bound of this set. That is,

$$\underline{\int}_a^b f \, dx \le \overline{\int}_a^b f \, dx.$$

Definition 5.1.6. Suppose f is a bounded function on a closed bounded interval $[a, b]$. If the upper and lower integrals of f on $[a, b]$ are equal, we will say that f is *integrable* on $[a, b]$. In this case the common value of $\underline{\int}_a^b f \, dx$ and $\overline{\int}_a^b f \, dx$ will be denoted by

$$\int_a^b f(x) \, dx.$$

and called the *Riemann Integral* of f on $[a, b]$.

Theorem 5.1.7. *The Riemann Integral of f on $[a, b]$ exists if and only if, for each $\epsilon > 0$, there is a partition P of $[a, b]$ such that*

$$U(f, P) - L(f, P) < \epsilon. \tag{5.1.6}$$

Proof. Suppose the integral exists. Then

$$\sup_P L(f, P) = \underline{\int}_a^b f \, dx = \overline{\int}_a^b f \, dx = \inf_P U(f, P),$$

where P ranges over all partitions of $[a, b]$. Thus, given $\epsilon > 0$, the number $\underline{\int}_a^b f \, dx - \epsilon/2$ is not an upper bound for the set of all $L(f, P)$ and the number $\overline{\int}_a^b f \, dx + \epsilon/2$ is not a lower bound for the set of all $U(f, P)$. This means there are partitions Q_1 and Q_2 of $[a, b]$ such that

$$\underline{\int}_a^b f \, dx - \epsilon/2 < L(f, Q_1) \le U(f, Q_2) < \overline{\int}_a^b f \, dx + \epsilon/2.$$

If P is a common refinement of Q_1 and Q_2, then Theorem 5.1.4 implies that

$$\underline{\int}_a^b f \, dx - \epsilon/2 < L(f, Q_1) \le L(f, P) \le U(f, P) \le U(f, Q_2) < \overline{\int}_a^b f \, dx + \epsilon/2.$$

Since $\underline{\int}_a^b f \, dx = \overline{\int}_a^b f \, dx$, this implies that (5.1.6) holds.

Conversely, suppose that for each $\epsilon > 0$ there is a partition P such that (5.1.6) holds. Then

$$L(f,P) \leq \underline{\int_a^b} f\,dx \leq \overline{\int_a^b} f\,dx \leq U(f,P)$$

implies that

$$\overline{\int_a^b} f\,dx - \underline{\int_a^b} f\,dx \leq U(f,P) - L(f,P) < \epsilon.$$

This means that $0 \leq \overline{\int_a^b} f\,dx) - \underline{\int_a^b} f\,dx < \epsilon$ for every positive ϵ, which is possible only if $\overline{\int_a^b} f\,dx) - \underline{\int_a^b} f\,dx = 0$. Thus, $\overline{\int_a^b} f\,dx) = \underline{\int_a^b} f\,dx$.   □

The above theorem leads to a sequential characterization of the Riemann Integral which will be highly useful in proving theorems about the integral.

Theorem 5.1.8. *The Riemann Integral exists if and only if there is a sequence* $\{P_n\}$ *of partitions of* $[a,b]$ *such that*

$$\lim(U(f,P_n) - L(f,P_n)) = 0. \tag{5.1.7}$$

In this case,

$$\int_a^b f(x)\,dx = \lim S_n(f)$$

where, for each n, $S_n(f)$ *may be chosen to be* $U(f,P_n)$, $L(f,P_n)$ *or any Riemann sum (5.1.1) for* f *and the partition* P_n.

Proof. If, for every $\epsilon > 0$, we can find a partition P of $[a,b]$ such that (5.1.6) holds, then, in particular, for each $n \in \mathbb{N}$ we can find a partition P_n such that

$$U(f,P_n) - L(f,P_n) < 1/n.$$

Then $\lim(U(f,P_n) - L(f,P_n)) = 0$.
    Conversely, if there is a sequence of partitions $\{P_n\}$ with

$$\lim(U(f,P_n) - L(f,P_n)) = 0,$$

then, given $\epsilon > 0$, there is an N such that

$$U(f,P_n) - L(f,P_n) < \epsilon \quad \text{whenever} \quad n > N.$$

By the previous theorem, this implies that the Riemann integral exists.
    Now given a sequence $\{P_n\}$ satisfying (5.1.7), we know that

$$L(f,P_n) \leq \int_a^b f(x)\,dx \leq U(f,P_n)$$

for each n. It follows that the sequences $\{L(f, P_n)\}$ and $\{U(f, P_n)\}$ both converge to $\int_a^b f(x)\,dx$. However, by (5.1.4), we also have

$$L(f, P_n) \leq S_n(f) \leq U(f, P_n)$$

if $S_n(f)$ is any Riemann sum for f and the partition P_n or is $U(f, P_n)$ or $L(f, P_n)$. By the squeeze principle (Theorem 2.3.3) , we conclude

$$\int_a^b f(x)\,dx = \lim S_n(f).$$

$\square$

Example 5.1.9. Prove that the Riemann Integral of $f(x) = x^2$ on $[0, 1]$ exists and is equal to $1/3$.

   Solution: The function is increasing and so its sup on any interval is achieved at the right endpoint and its inf is achieved at the left endpoint. For each $n \in \mathbb{N}$ we define a partition P_n of $[0, 1]$ by

$$P_n = \{0 < 1/n < 2/n < \cdots < n/n = 1\}.$$

This divides $[0, 1]$ into n subintervals, each of which has length $1/n$. The corresponding upper sum is then

$$U(f, P_n) = \sum_{k=1}^{n} \left(\frac{k}{n}\right)^2 \frac{1}{n} = \frac{1}{n^3} \sum_{k=1}^{n} k^2,$$

while the lower sum is

$$L(f, P_n) = \sum_{k=1}^{n} \left(\frac{k-1}{n}\right)^2 \frac{1}{n} = \frac{1}{n^3} \sum_{k=0}^{n-1} k^2.$$

The difference is

$$U(f, P_n) - L(f, P_n) = \frac{n^2}{n^3} = \frac{1}{n}.$$

This sequence certainly has limit 0 and so, by Theorem 5.1.8, the Riemann Integral exists. To find what it is, we need a formula for the sum $\sum_{k=1}^{n} k^2$. Such a formula exists. In fact, it can be proved by induction (Exercise 5.1.3) that

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

Thus,

$$U(f, P_n) = \frac{n(n+1)(2n+1)}{6n^3} = \frac{(1+1/n)(2+1/n)}{6}.$$

This expression has limit $1/3$ as $n \to \infty$ and so $\int_0^1 x^3\,dx = 1/3$.

Exercise Set 5.1

1. Find the upper sum $U(f, P)$ and lower sum $L(f, P)$ if $f(x) = 1/x$ on $[1, 2]$ and P is the partition of $[1, 2]$ into four subintervals of equal length.

2. Prove that $\int_0^1 x \, dx$ exists by computing $U(f, P_n)$ and $L(f, P_n)$ for the function $f(x) = x$ and a partition P_n of $[0, 1]$ into n equal subintervals. Then show that condition (5.1.7) of Theorem 5.1.8 is satisfied. Calculate the integral by taking the limit of the upper sums. Hint: use Exercise 1.2.10.

3. Prove by induction that

$$\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{6}.$$

4. Prove that $\int_0^a x^2 \, dx = \frac{a^3}{3}$ by expressing this integral as a limit of Riemann sums and finding the limit.

5. Let f be the function on $[0, 1]$ which is 0 at every rational number and is 1 at every irrational number. Is this function integrable on $[0, 1]$. Prove that your answer is correct by using the definition of the integral.

6. Prove that the upper sum $U(f, P)$ for a partition of $[a, b]$ and a bounded function f on $[a, b]$ is the least upper bound of the set of all Riemann sums for f and P.

7. Finish the proof of Theorem 5.1.4 by showing that if the theorem is true when only one element is added to P to obtain Q, then it is also true no matter how many elements need to be added to P to obtain Q.

8. Suppose m and M are lower and upper bounds for f on $[a, b]$; that is $m \le f(x) \le M$ for all $x \in [a, b]$. Prove that

$$m(b-a) \le \underline{\int_a^b} f(x) \, dx \le \overline{\int_a^b} f(x) \, dx \le M(b-a).$$

What conclusion about $\int_a^b f(x) \, dx$ do you draw from this if the integral exists?

9. If k is a constant and $[a, b]$ a bounded interval, prove that k is integrable on $[a, b]$ and

$$\int_a^b k \, dx = k(b-a).$$

10. If f is a bounded function on $[a, b]$ and $P = \{x_0 < x_1 < \cdots < x_n\}$ a partition of $[a, b]$, show that

$$U(f, P) - L(f, P) = \sum_{k=1}^{n} (M_k - m_k)(x_k - x_{k-1}),$$

where M_k is the sup and m_k the inf of f on $[x_{k-1}, x_k]$. What does this simplify to if P is a partition of $[a, b]$ into n equal subintervals?

11. Suppose f is any non-decreasing function on a bounded interval $[a, b]$. If P_n is the partition of $[a, b]$ into n equal subintervals, show that

$$U(f, P_n) - L(f, P_n) = (f(b) - f(a)) \frac{b - a}{n}.$$

What do you conclude about the integrability of f?

5.2   Existence and Properties of the Integral

We first show that the integral exists for a large class of functions, a class which includes all the functions of interest to us in this course. We then show that the integral has the properties claimed for it in calculus courses.

Existence Theorems

Theorem 5.2.1. *If f is a monotone function on a closed bounded interval $[a, b]$, then f is integrable on $[a, b]$.*

Proof. This was essentially stated as an exercise (Exercise 5.1.11) in the previous section. In this exercise, it is claimed that, if f is a non-decreasing function on $[a, b]$ and P_n is the partition of $[a, b]$ into n equal subintervals, then

$$U(f, P_n) - L(f, P_n) = (f(b) - f(a)) \frac{b - a}{n}. \tag{5.2.1}$$

This implies that

$$\lim(U(f, P_n) - L(f, P_n)) = 0$$

and, by Theorem 5.1.8, this implies that the Riemann Integral of f on $[a, b]$ exists.

In the case where f is non-increasing, the same proof works. The only difference is that $f(b) - f(a)$ is replaced by $f(a) - f(b)$ in (5.2.1). □

Theorem 5.2.2. *If f is a continuous function on a closed, bounded interval $[a, b]$, then f is integrable on $[a, b]$.*

Proof. Since f is continuous on the closed, bounded interval $[a, b]$, it is uniformly continuous on $[a, b]$ by Theorem 3.3.4. Thus, given $\epsilon > 0$ there is a $\delta > 0$ such that

$$|f(x) - f(y)| < \frac{\epsilon}{b - a} \quad \text{whenever} \quad |x - y| < \delta.$$

Then, if $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ is any partition of $[a, b]$ with the property that the interval $[x_{k-1}, x_k]$ has length less than δ for each k, then the maximum value M_k of f on this interval and the minimum value m_k of f on this interval differ by less than $\epsilon/(b - a)$. By Exercise 5.1.10, this implies that

$$U(f, P) - L(f, P) = \sum_{k=1}^{n} (M_k - m_k)(x_k - x_{k-1}) < \frac{\epsilon}{b - a} \sum_{k=1}^{n} (x_k - x_{k-1}) = \epsilon,$$

since $\sum_{k=1}^{n} (x_k - x_{k-1}) = b - a$. It follows from Theorem 5.1.7 that f is integrable on $[a, b]$. □

Linearity of the Integral

In the remainder of this section we adopt the following notation, introduced in Section 1.5 for the sup and inf of a function f on an interval I:

$$\sup_I f = \sup\{f(x) : x \in I\} \quad \text{and} \quad \inf_I f = \inf\{f(x) : x \in I\}.$$

The integral is a *linear transformation* from the space of integrable functions on $[a, b]$ to the real numbers. This just means that the following familiar theorem is true.

Theorem 5.2.3. *If f and g are integrable functions on a closed, bounded interval $[a, b]$ and c is a constant, then*

(a) cf is integrable and $\int_a^b cf(x)\, dx = c \int_a^b f(x)\, dx$;

(b) $f + g$ is integrable and $\int_a^b (f(x) + g(x))\, dx = \int_a^b f(x)\, dx + \int_a^b g(x)\, dx$.

Proof. We begin by investigating the upper and lower sums for a partition $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ and the functions cf and $f + g$. We let $I_k = [x_{k-1}, x_k]$ denote the kth subinterval determined by this partition.

If $c \geq 0$, then Theorem 1.5.10(a) tells us that

$$\sup_{I_k} cf = c \sup_{I_k} f \quad \text{and} \quad \inf_{I_k} cf = c \inf_{I_k} f$$

for $k = 1, \cdots, n$. This implies that

$$U(cf, P) = cU(f, P) \quad \text{and} \quad L(cf, P) = cL(f, P) \quad \text{if} \quad c \geq 0. \tag{5.2.2}$$

On the other hand, by Theorem 1.5.10(b),

$$\sup_{I_k}(-f) = -\inf_{I_k} f \quad \text{and} \quad \inf_{I_k}(-f) = -\sup_{I_k} f$$

for each k. This implies that

$$U(-f, P) = -L(f, P) \quad \text{and} \quad L(-f, P) = -U(f, P). \tag{5.2.3}$$

For the sum $f + g$, we have

$$\inf_{I_k} f + \inf_{I_k} g \le \inf_{I_k}(f + g) \le \sup_{I_k}(f + g) \le \sup_{I_k} f + \sup_{I_k} g$$

for each k, by 1.5.10(c). These inequalities imply that

$$L(f, P) + L(g, P) \le L(f + g, P) \le U(f + g, P) \le U(f, P) + U(g, P). \tag{5.2.4}$$

With these results in hand, the proof of the theorem becomes a relatively simple matter. We use Theorem 5.1.8. Since f is integrable, there is a sequence $\{P_n\}$ of partitions of $[a, b]$ such that

$$\lim(U(f, P_n) - L(f, P_n)) = 0. \tag{5.2.5}$$

If $c \ge 0$, then (5.2.2) implies that

$$\lim(U(cf, P_n) - L(cf, P_n) = \lim c(U(f, P_n) - L(f, P_n)) = 0$$

which implies that cf is integrable. It also follows from (5.2.2) that

$$\int_a^b cf(x)\, dx = \lim U(cf, P_n) = c \lim U(f, P_n) = c \int_a^b f(x)\, dx.$$

Similarly, using (5.2.3) yields

$$\lim(U(-f, P_n) - L(-f, P_n)) = \lim(-L(f, P_n) + U(f, P_n)) = 0,$$

which implies that $-f$ is integrable. It also follows from (5.2.3) that

$$\int_a^b -f(x)\, dx = \lim U(-f, P_n) = -\lim L(f, P_n) = -\int_a^b f(x)\, dx.$$

Combining these results proves part (a) of the theorem.

Since, g is also integrable, there is a sequence of partitions $\{Q_n\}$ such that (5.2.5) holds with f replaced by g and P_n by Q_n. In fact, we may replace $\{P_n\}$ and $\{Q_n\}$ by the sequence of common refinements $\{P_n \cup Q_n\}$ and get a sequence of partitions that works for both f and g. Since this is so, we may as well assume that $\{P_n\}$ was chosen in the first place to be a sequence of partitions such that (5.2.5) holds and

$$\lim(U(g, P_n) - L(g, P_n)) = 0. \tag{5.2.6}$$

also holds. Then 5.2.4 implies that

$$0 \le U(f+g, P_n) - L(f+g, P_n) \le U(f, P_n) - L(f, P_n) + U(g, P_n) - L(g, P_n).$$

Since the expression on the right has limit 0, so does $U(f+g, P_n) - L(f+g, P_n)$. Hence, $f + g$ is integrable. Also, on passing to the limit as P ranges through the sequence of partitions P_n, inequality (5.2.4) implies that

$$\int_a^b (f(x) + g(x)) \, dx = \int_a^b f(x) \, dx + \int_a^b g(x) \, dx.$$

This completes the proof of part (b) of the theorem.   □

The Order Preserving Property

The integral is order preserving:

Theorem 5.2.4. *If f and g are integrable functions on $[a, b]$ and $f(x) \le g(x)$ for all $x \in [a, b]$, then*

$$\int_a^b f(x) \, dx \le \int_a^b g(x) \, dx.$$

Proof. We first prove that if h is an integrable function which is non-negative on $[a, b]$, then

$$\int_a^b h(x) \, dx \ge 0.$$

In fact, this is obvious. If h is non-negative, then its inf and sup on any subinterval in any partition are also non-negative. This implies that the upper sums $U(h, P)$ and lower sums $L(h, P)$ are non-negative for any partition P. Since the integral is greater than or equal to every lower sum, it is non-negative.

To finish the proof, we apply the result of the previous paragraph to the function $h = g - f$ which is non-negative on $[a, b]$ if $f(x) \le g(x)$ for $x \in [a, b]$. Using linearity (Theorem 5.2.3) we conclude that

$$\int_a^b g(x) \, dx - \int_a^b f(x) \, dx = \int_a^b (g(x) - f(x)) \, dx \ge 0.$$

This proves the theorem.   □

This has the following useful corollary. Its proof is left to the exercises.

Corollary 5.2.5. *Let f be an integrable function on the closed bounded interval $I = [a, b]$ and set $M = \sup_I f$, and $m = \inf_I f$. Then*

$$m(b-a) \le \int_a^b f(x) \, dx \le M(b-a)$$

Theorem 5.2.6. *If f is integrable on $[a, b]$, then $|f|$ is also integrable on $[a, b]$ and*

$$\left| \int_a^b f(x)\, dx \right| \leq \int_a^b |f(x)|\, dx$$

.

Proof. Let f be integrable on $[a, b]$. Suppose we can show that $|f|$ is also integrable on $[a, b]$. To derive the above inequality is then quite easy. The inequalties $-|f(x)| \leq f(x) \leq |f(x)|$, together with Theorem 5.2.4, imply that

$$-\int_a^b |f(x)|\, dx \leq \int_a^b f(x)\, dx \leq \int_a^b |f(x)|\, dx$$

and this implies that

$$\left| \int_a^b f(x)\, dx \right| \leq \int_a^b |f(x)|\, dx.$$

To complete the proof, we must show that the integrability of f on $[a, b]$ implies the integrability of $|f|$.

Let I be an arbitrary subinterval of $[a, b]$. Then, by the triangle inequality,

$$|f(x)| - |f(y)| \leq |f(x) - f(y)|$$

for all $x, y \in I$. It follows from this (Exercise 5.2.8) that

$$\sup_I |f| - \inf_I |f| \leq \sup_I f - \inf_I f.$$

If we apply this as I ranges over each subinterval in a partition P, the result for the upper and lower sums is

$$U(|f|, P) - L(|f|, P) \leq U(f, P) - L(f, P).$$

It now follows from Theorem 5.1.7 that $|f|$ is integrable on $[a, b]$ if f is integrable on $[a, b]$. ☐

Interval Additivity

Note that, in the following theorem, we do not assume that f is integrable.

Theorem 5.2.7. *Suppose $a \leq b \leq c$ and f is a bounded function defined on $[a, c]$. Then the upper and lower integrals of f satisfy*

$$\underline{\int_a^c} f(x)\, dx = \underline{\int_a^b} f(x)\, dx + \underline{\int_b^c} f(x)\, dx, \quad \overline{\int_a^c} f(x)\, dx = \overline{\int_a^b} f(x)\, dx + \overline{\int_b^c} f(x)\, dx.$$

Proof. We will prove the result for the lower integral. The proof for the upper integral is analogous.

Let $P = \{a = x_o \leq x_1 \leq \cdots \leq x_n = c\}$ be a partition of $[a, c]$ which has the point b as its mth partition point. Then P determines partitions

$$P' = \{a = x_0 < x_1 < \cdots < x_m = b\} \quad \text{of} \quad [a, b] \quad \text{and}$$
$$P'' = \{b = x_m < x_{m+1} < \cdots < x_n = c\} \quad \text{of} \quad [b, c].$$

In this case,

$$L(P', f) + L(P'', f) = L(P, f). \tag{5.2.7}$$

Each pair consisting of a partition P' of $[a, b]$ and a partition P'' of $[c, d]$ fit together to form a partition P of $[a, c]$ of this type.

Now let Q be any partition of $[a, c]$. Then the union of Q with the singeton set $\{b\}$ forms a refinement P of Q which is of the above type. Then

$$L(f, Q) \leq L(f, P) \leq \underline{\int}_a^c f(x)\, dx.$$

But $\underline{\int}_a^c f(x)\, dx$ is the sup of all numbers of the form $L(Q, f)$ for Q a partition and $L(f, P) = L(f, P') + L(L, P'')$. of $[a, c]$, It follows from 1.5.7(c) that

$$\underline{\int}_a^b f(x)\, dx + \underline{\int}_b^c f(x)\, dx = \sup_{P'} L(P', f) + \sup_{P''} L(P'', f)$$

$$= \sup\{L(P', f) + L(P'', f)\} = \underline{\int}_a^c f(x)\, dx.$$

where P' and P'' range over arbitrary partitions of $[a, b]$ and $[b, c]$. This proves the theorem for lower integrals. The proof for upper integrals is essentially the same.   □

This theorem has as a corollary the interval additivity property for the integral. The details of how this corollary follows from the above theorem are left to the exercises.

Corollary 5.2.8. *With f and $a \leq b \leq c$ as in the previous theorem, f is integrable on $[a, c]$ if and only if it is integrable on $[a, b]$ and on $[b, c]$. In this case,*

$$\int_a^c f(x)\, dx = \int_a^b f(x)\, dx + \int_b^c f(x)\, dx.$$

Theorem of the Mean for Integrals

If f is an integrable function on a bounded interval $[a, b]$, then the *mean* or *average* of f on $[a, b]$ is the number

$$\frac{1}{b - a} \int_a^b f(x)\, dx.$$

The following theorem is an easy consequence of the Intermediate Value Theorem. We leave its proof to the exercises.

Theorem 5.2.9. *If f is a continuous function on a closed bounded interval $[a, b]$, then there is a point $c \in [a, b]$ such that*

$$f(c) = \frac{1}{b-a} \int_a^b f(x) \, dx.$$

Exercise Set 5.2

1. Show that if a function f on a bounded interval can be written in the form $g - h$ for functions g and h which are non-decreasing on $[a, b]$, then f is integrable on $[a, b]$.

2. If f is a bounded function defined on a closed bounded interval $[a, b]$ and if f is integrable on each interval $[a, r]$ with $a < r < b$, then prove that f is integrable on $[a, b]$ and

$$\int_a^b f(x) \, dx = \lim_{r \to b} \int_a^r f(x) \, dx.$$

   Observe that the analogous result holds if $[a, r]$ is replaced by $[r, b]$ in the hypothesis and in the integral on the right, and the limit is taken as $r \to a$. Hint: use Theorem 5.2.7 and Exercise 5.1.8.

3. Suppose f is a bounded function on a bounded interval $[a, b]$ and there is a partition $\{a = x_0 < x_1 < \cdots < x_n = b\}$ of $[a, b]$ such that f is continuous on each subinterval (x_{k-1}, x_k). Prove that such a function is integrable on $[a, b]$.

4. Prove Corollary 5.2.5.

5. Prove Corollary 5.2.8.

6. Prove that $1 \le \int_{-1}^1 \frac{1}{1 + x^{2n}} \, dx \le 2$ for all $n \in \mathbb{N}$.

7. Prove that $\int_{-1}^1 \frac{x^2}{1 + x^{2n}} \, dx \le 2/3$ for all $n \in \mathbb{N}$.

8. If f is a bounded function defined on an interval I, then prove that

$$\sup_I |f| - \inf_I |f| \le \sup_I f - \inf_I f$$

   by using Theorem 1.5.10(d) and the triangle inequality $|f(x)| - |f(y)| \le |f(x) - f(y)|$.

9. Prove that if f is integrable on $[a, b]$ then so is f^2. Hint: if $|f(x)| \leq M$ for all $x \in [a, b]$, then show that

$$|f^2(x) - f^2(y)| \leq 2M|f(x) - f(y)|.$$

for all $x, y \in [a, b]$. Use this to estimate $U(f^2, P) - L(f^2, P)$ in terms of $U(f, P) - L(f, P)$ for a given partition P.

10. Prove that if f and g are integrable on $[a, b]$, then so is fg. Hint: write fg as the difference of two squares of functions you know are integrable and then use the previous exercise.

11. Give an example of a function f such that $|f|$ is integrable on $[0, 1]$ but f is not integrable on $[0, 1]$.

12. Prove Theorem 5.2.9.

13. Let $\{f_n\}$ be a sequence of integrable functions defined on a closed bounded interval $[a, b]$. If $\{f_n\}$ converges uniformly on $[a, b]$ to a function f, prove that f is integrable and

$$\int_a^b f(x)\, dx = \lim \int_a^b f_n(x)\, dx.$$

14. Is the function which is $\sin 1/x$ for $x \neq 0$ and 0 for $x = 0$ integrable on $[0, 1]$? Justify your answer.

5.3   The Fundamental Theorems of Calculus

There are two fundamental theorems of calculus. Both relate differentiation to integration. In most calculus courses, the Second Fundamental Theorem is usually proved first and then used to prove the First Fundamental Theorem. Unfortunately, this results in a First Fundamental Theorem that is weaker than it could be. To prove the best possible theorems, one should give independent proofs of the two theorems. This is what we shall do.

First Fundamental Theorem

The following theorem concerns the integral of f' on $[a, b]$ where f is a function which we assume is differentiable on (a, b) but not necessarily at a or b. The reason the integral still makes sense is that, for a function that is integrable on $[a, b]$, changing its value at one point (or at finitely many points) does not affect its integrability or its integral (Exercise 5.3.9). Thus, a function which is missing values at a and/or b can be assigned values there arbitrarily and the integrability and value of the integral will not depend on how this is done.

Theorem 5.3.1. *Let $[a, b]$ be a closed bounded interval and f a function which is continuous on $[a, b]$ and differentiable on (a, b) with f' integrable on $[a, b]$. Then*

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

Proof. Let $P = \{a = x_0 < x_1 < \cdots < x_n = b\}$ be a partition of $[a, b]$. We apply the Mean Value Theorem to f on each of the intervals $[x_{k-1}, x_k]$. This tells us there is a point $c_k \in (x_{k-1}, x_k)$ such that

$$f'(c_k)(x_k - x_{k-1}) = f(x_k) - f(x_{k-1}).$$

If we sum this over $k = 1, \cdots, n$, the result is

$$\sum_{k=1}^n f'(c_k)(x_k - x_{k-1}) = f(b) - f(a).$$

The sum on the left is a Riemann sum for f' and the partition P and so, by (5.1.4), it lies between the lower and upper sums for f' and P. Thus,

$$L(f', P) \le f(b) - f(a) \le U(f', P). \tag{5.3.1}$$

Since f' is integrable on $[a, b]$, there is a sequence of partitions $\{P_n\}$ for which the corresponding sequences of upper and lower sums for f' both converge to $\int_a^b f'(x)\, dx$. However, in view of (5.3.1) the only number both sequences can converge to is $f(b) - f(a)$. $\square$

The above theorem is somewhat stronger than the one usually stated in calculus, because we only assume that the derivative f' is integrable on $[a, b]$, not that it is continuous. Are there functions which are differentiable with an integrable derivative which is not continuous?

Example 5.3.2. Find a function f which is differentiable on an interval, with an integrable derivative which is not continuous.
   Solution: Let $f(x) = x^2 \sin 1/x$ if $x \ne 0$ and set $f(0) = 0$. Then, f is differentiable on all of $\mathbb{R}$ and its derivative is

$$f'(x) = 2x \sin 1/x - \cos 1/x \quad \text{if} \quad x \ne 0$$

and is 0 at $x = 0$. This follows from the Chain Rule and the Product Rule for derivatives everywhere except at $x = 0$. At $x = 0$ we calculate the derivative using the definition of derivative:

$$f'(0) = \lim_{x \to 0} \frac{x^2 \sin 1/x}{x} = \lim_{x \to 0} x \sin 1/x = 0.$$

Now the function $f'(x)$ is integrable on any closed bounded interval (see Exercise 5.2.2), but it is not continuous at 0. Thus, f is a function to which the previous theorem applies, but it does not have a continuous derivative.

Second Fundamental Theorem

So far we have defined the integral $\int_a^b f(x)\,dx$ only in the case where $a < b$. We remedy this by defining the integral to be 0 if $a = b$ and declaring

$$\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx \quad \text{if} \quad b < a.$$

This ensures that the integral in the following theorem makes sense whether x is to the left or the right of a.

Theorem 5.3.3. *Let f be a function which is integrable on a closed bounded interval $[b, c]$. For $a, x \in [b, c]$ define a function $F(x)$ by*

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

Then F is continuous on $[b, c]$. At each point x of (b, c) where f is continuous the function F is differentiable and

$$F'(x) = f(x).$$

Proof. The definition of F makes sense, because it follows from Theorem 5.2.7 that a function integrable on an interval is also integrable on every subinterval.

Since f is integrable on $[b, c]$ it is bounded on $[b, c]$. Thus, there is an M such that

$$|f(t)| \le M \quad \text{for all} \quad t \in [b, c].$$

If $x, y \in [b, c]$ then

$$F(y) - F(x) = \int_a^y f(t)\,dt - \int_a^x f(t)\,dt = \int_x^y f(t)\,dt. \tag{5.3.2}$$

(see Exercise 5.3.11). Then by Exercise 5.3.12 ,

$$|F(y) - F(x)| = \left| \int_x^y f(t)\,dt \right| \le M|y - x|.$$

Thus, given $\epsilon > 0$, if we choose $\delta = \epsilon/M$, then

$$|F(y) - F(x)| < \epsilon \quad \text{whenever} \quad |y - x| < \delta.$$

This shows that F is uniformly continuous on $[b, c]$.

Now suppose $x \in (b, c)$ is a point at which f is continuous. If y is also in (b, c), then

$$\int_x^y f(x)\,dt = f(x)(y - x)$$

since $f(x)$ is a constant as far as the variable of integration t is concerned. This and (5.3.2) imply that

$$
\begin{aligned}
\frac{F(y) - F(x)}{y - x} - f(x) &= \frac{1}{y - x} \left(\int_x^y f(t)\, dt - \int_x^y f(x)\, dt \right) \\
&= \frac{1}{y - x} \int_x^y (f(t) - f(x))\, dt.
\end{aligned}
\tag{5.3.3}
$$

Since f is continuous at x, given $\epsilon > 0$, we may choose $\delta > 0$ such that

$$
|f(t) - f(x)| < \epsilon \quad \text{whenever} \quad |x - t| < \delta.
$$

Then, for y with $|y - x| < \delta$, it will be true that $|x - t| < \delta$ for every t between x and y. Thus, for such a choice of y, we have

$$
\left| \frac{1}{y - x} \int_x^y (f(t) - f(x))\, dt \right| \leq \frac{1}{|y - x|} \epsilon |y - x| = \epsilon
$$

In view of (5.3.3), this implies that

$$
\lim_{y \to x} \frac{F(y) - F(x)}{y - x} = f(x).
$$

Thus, F is differentiable at x and $F'(x) = f(x)$. $\qquad\qquad\qquad$ $\square$

Example 5.3.4. Find $\dfrac{d}{dx} \displaystyle\int_0^{\sin x} e^{-t^2}\, dt$.

$\quad$ **Solution:** This is a composite function. If $F(u) = \displaystyle\int_0^u e^{-t^2}\, dt$, then the function we are asked to differentiate is $F(\sin x)$. By the Chain Rule, the derivative of this composite function is

$$
F'(\sin x) \cos x.
$$

By the previous theorem, $F'(u) = e^{-u^2}$. Thus,

$$
\frac{d}{dx} \int_0^{\sin x} e^{-t^2}\, dt = F'(\sin x) \cos x = e^{-\sin^2 x} \cos x.
$$

Example 5.3.5. Find $\dfrac{d}{dx} \displaystyle\int_x^{2x} \sin t^2\, dt$.
$\quad$ **Solution:** We begin by writing

$$
G(x) = \int_x^{2x} \sin t^2\, dt = \int_0^{2x} \sin t^2\, dt - \int_0^x \sin t^2\, dt.
$$

Then using the previous theorem and the Chain Rule yields

$$
G'(x) = 2 \sin 4x^2 - \sin x^2.
$$

Substitution

We will not rehash all the integration techniques that are taught in the typical calculus class. However, two of these techniques are of such great theoretical importance, that it is worth discussing them again. The techniques in question are substitution and integration by parts. Each of these follows from the Fundamental Theorems and an important theorem from differential calculus – the chain rule in the case of substitution and the product rule in the case of integration by parts. We begin with substitution.

Theorem 5.3.6. *Let g be a differentiable function on an open interval I with g' integrable on I and let $J = g(I)$. Let f be continuous on J. Then for any pair $a, b \in I$,*

$$\int_a^b f(g(t))g'(t)\, dt = \int_{g(a)}^{g(b)} f(u)\, du. \tag{5.3.4}$$

Proof. The composite function $f \circ g$ is continuous on I since g is continuous on I and f is continuous on J. By Exercise 5.2.10, this implies that $f(g(t))g'(t)$ is an integrable function of t on I. We set

$$F(v) = \int_{g(a)}^v f(u)\, du.$$

Then $F'(v) = f(v)$ by the Second Fundamental Theorem, and so, by the Chain Rule,

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

Thus, $F \circ g$ is a differentiable function on I with an integrable derivative $f(g(x))g'(x)$. By the First Fundamental Theorem,

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

By the definition of F, $F(g(a)) = 0$ and $F(g(b)) = \displaystyle\int_{g(a)}^{g(b)} f(u)\, du$. Thus,

$$\int_a^b f(g(x))g'(x)\, dx = \int_{g(a)}^{g(b)} f(u)\, du,$$

as claimed.   □

Note that the above theorem states formally what happens when we make the substitution $u = g(t)$ in the integral on the left in (5.3.4).

Integration by Parts

The integration by parts formula is a direct consequence of the Fundamental Theorems and the product rule for differentiation.

Theorem 5.3.7. *Suppose f and g are continuous functions on a closed bounded interval $[a, b]$ and suppose that f and g are differentiable on (a, b) with derivatives that are integrable on $[a, b]$. Then fg' and $f'g$ are integrable on $[a, b]$ and*

$$\int_a^b f(x)g'(x)\,dx = f(b)g(b) - f(a)g(a) - \int_a^b g(x)f'(x)\,dx. \qquad (5.3.5)$$

Proof. We have f and g integrable because they are continuous on $[a, b]$, while f' and g' are integrable by hypothesis. By Exercise 5.2.10, fg' and gf' are both integrable.

The product fg is differentiable on (a, b) and

$$(fg)' = fg' + gf'.$$

Thus, $(fg)'$ is also integrable and, by the First Fundamental Theorem,

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

Formula (5.3.5) follows immediately from this.   □

Example 5.3.8. Suppose f is a continuous function on $[-\pi, \pi]$ which is differentiable on $(-\pi, \pi)$ with an integrable derivative. Also suppose $f(-\pi) = f(\pi)$. Prove that, for each $n \in \mathbb{N}$,

$$\begin{aligned}
\int_{-\pi}^{\pi} f'(x) \sin nx\,dx &= -n \int_{-\pi}^{\pi} f(x) \cos nx\,dx \\
\int_{-\pi}^{\pi} f'(x) \cos nx\,dx &= n \int_{-\pi}^{\pi} f(x) \sin nx\,dx.
\end{aligned} \qquad (5.3.6)$$

Solution: These are the equations relating the Fourier coefficients of the derivative of a function f to the Fourier coefficients of f itself.

The first equation is proved using the integration by parts formula (5.3.5) for $f(x)$ and $g(x) = \sin x$. Since $\sin(-n\pi) = \sin(n\pi) = 0$, the terms $f(b)g(b) - f(a)g(a)$ are 0. The first equation then follows directly from (5.3.5).

The second equation follows from (5.3.5) for $f(x)$ and $g(x) = \cos x$. However, this time the terms $f(b)g(b) - f(a)g(a)$ contribute 0 because cos is an even function and $f(-\pi) = f(\pi)$.

Exercise Set 5.3

1. Find $\int_{4/\pi}^{2/\pi} (2x \sin 1/x - \cos 1/x)\,dx$. Hint: see Example 5.3.2.

2. Find $\dfrac{d}{dx} \int_1^x \cos 1/t\,dt$ for $x > 0$.

3. Find $\dfrac{d}{dx} \int_0^{2x} \sin t^2\,dt$.

4. Find $\dfrac{d}{dx} \displaystyle\int_{1/x}^{x} e^{-t^2} \, dt$.

5. If $f(x) = -1/x$ then $f'(x) = 1/x^2$. Thus, Theorem 5.3.1 seems to imply that

$$\int_{-1}^{1} 1/x^2 \, dx = f(1) - f(-1) = -1 - 1 = -2.$$

However, $1/x^2$ is a positive function, and so its integral over $[-1, 1]$ should be positive. What is wrong?

6. If f is a differentiable function on $[a, b]$ and f' is integrable on $[a, b]$, then find

$$\int_{a}^{b} f(x) f'(x) \, dx.$$

7. Let f be a continuous function on the interval $[0, 1]$. Express

$$\int_{0}^{\pi/2} f(\sin \theta) \cos \theta \, d\theta$$

as an integral involving only the function f.

8. Find $\displaystyle\int_{0}^{x} t^n \ln t \, dt$ where n is an arbitrary integer.

9. Prove that if f is integrable on $[a, b]$ and $c \in [a, b]$, then changing the value of f at c does not change the fact that f is integrable or the value of its integral on $[a, b]$.

10. The function $f(x) = x/|x|$ has derivative 0 everywhere but at $x = 0$. Its derivative $f'(x) = 0$ is integrable on $[-1, 1]$ and has integral 0. However $f(1) - f(-1) = 1 - (-1) = 2$. This seems to contradict Theorem 5.3.1. Explain why it does not.

11. The interval additivity property (Theorem 5.2.7) is stated for three points a, b, c satisfying $a < b < c$. Show that it actually holds regardless of how the points a, b, and c are ordered. Hint: you will need to consider various cases.

12. Suppose f is integrable on an interval containing a and b and $|f(x)| \le M$ on I. Prove that

$$\left| \int_{a}^{b} f(x) \, dx \right| \le M |b - a|.$$

Note that we do not assume that $a < b$.

5.4 Logs, Exponentials, Improper Integrals

The following development of the log and exponential functions is the one presented in most calculus classes these days. It is such a beautiful application of the Second Fundamental Theorem that we felt obligated to include it here.

The Natural Logarithm

One consequence of the Second Fundamental Theorem is that every function f which is continuous on an open interval I has an anti-derivative on I. In fact, if a is any point of I, then

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

is an anti-derivative for f on I (that is, $F'(x) = f(x)$ on I).

Now $\dfrac{x^{n+1}}{n+1}$ is an antiderivative for x^n for all integers n with the exception of $n = -1$. However, since x^{-1} is continuous on $(0, +\infty)$ and on $(-\infty, 0)$, it has an antiderivative on each of these intervals. There is no mystery about what the antiderivatives are. On $(0, +\infty)$ the function

$$\int_1^x \frac{1}{t}\,dt$$

is an antiderivative for $1/x$. Obviously, this function is important enough to deserve a name.

Definition 5.4.1. We define the natural logarithm to be the function ln, defined for $x \in (0, +\infty)$ by

$$\ln x = \int_1^x \frac{1}{t}\,dt.$$

This is the unique antiderivative for $1/x$ on $(0, +\infty)$ which has the value 0 when $x = 1$.

On $(-\infty, 0)$ an antiderivative for $1/x$ is given by

$$\int_{-1}^x \frac{1}{t}\,dt.$$

Note that the x that appears in this integral is negative, and so $-x = |x|$. If we make the substitution $s = -t$, then Theorem 5.3.6 implies that

$$\int_{-1}^x \frac{1}{t}\,dt = \int_1^{-x} \frac{1}{s}\,ds = \ln(-x) = \ln|x|.$$

Thus, $\ln|x|$ is an antiderivative for $1/x$ on both $(0, +\infty)$ and $(-\infty, 0)$.

The next two theorems show that ln has the key properties that we expect of a logarithm.

Theorem 5.4.2. *For all $a, b \in (0, +\infty)$, $\ln ab = \ln a + \ln b$.*

Proof. By the Chain Rule, the derivative of $\ln ax$ is $\dfrac{1}{ax}a = \dfrac{1}{x}$. Thus, $\ln ax$ and $\ln x$ have the same derivative on the interval $(0, +\infty)$. By Corollary 4.3.4

$$\ln ax = \ln x + c$$

for some constant c. The constant may be evaluated by setting $x = 1$. Since $\ln 1 = 0$, this tells us that $c = \ln a$. Thus,

$$\ln ax = \ln x + \ln a.$$

This gives $\ln ab = \ln a + \ln b$ when we set $x = b$. $\square$

Theorem 5.4.3. *If $a > 0$ and r is any rational number, then $\ln a^r = r \ln a$.*

Proof. The proof of this is similar to the proof of the previous theorem. The key is to compute the derivative of the function $\ln x^r$. We leave the details to Exercise 5.4.1. $\square$

Theorem 5.4.4. *The natural logarithm is strictly increasing on $(0, +\infty)$. Also,*

$$\lim_{x \to \infty} \ln x = +\infty \quad and \quad \lim_{x \to 0} \ln x = -\infty.$$

Proof. The function $\ln x$ is strictly increasing on $(0, +\infty)$ because its derivative is positive on this interval.

Since $\ln 1 = 0$ and $\ln$ is increasing, $\ln 2$ is positive. Given any number M, choose an integer m such that $m \ln 2 > M$ and set $N = 2^m$. Then

$$\ln x > \ln 2^m = m \ln 2 > M \quad \text{whenever} \quad x > N.$$

This implies that $\lim_{x \to \infty} \ln x = +\infty$. The fact that $\lim_{x \to 0} \ln x = -\infty$ follows easily from $\lim_{x \to \infty} \ln x = +\infty$ and properties of $\ln$. The details are left to the exercises. $\square$

The Exponential Function

The function $\ln$ is strictly increasing on $(0, +\infty)$ and, therefore, it has an inverse function. The image of $(0, +\infty)$ under $\ln$ is an open interval by Exercise 4.2.5. By Theorem 5.4.4 this open interval must be the interval $(-\infty, \infty)$. Therefore, the inverse function for $\ln$ has domain $(-\infty, \infty)$ and image $(0, \infty)$.

Definition 5.4.5. We define the exponential function to be the function with domain $(-\infty, \infty)$ which is the inverse function of $\ln$. We will denote it by $\exp x$.

The theorems we proved about $\ln$ immediately translate into theorems about $\exp$.

Theorem 5.4.6. *The function $\exp$ is its own derivative – that is, $\exp'(x) = \exp(x)$.*

Proof. By Theorem 4.2.9 we have

$$\exp'(x) = \frac{1}{\ln'(\exp(x))} = \frac{1}{1/\exp(x)} = \exp(x).$$

$\square$

Theorem 5.4.7. *The exponential function satisfies*

(a) $\exp(a + b) = \exp a \exp b$ *for all* $a, b \in \mathbb{R}$;

(b) $\exp(ra) = (\exp(a))^r$ *for all* $a \in R$ *and* $r \in \mathbb{Q}$.

Proof. Let $x = \exp a$ and $y = \exp b$, so that $a = \ln x$ and $b = \ln y$. Then

$$\exp(a + b) = \exp(\ln x + \ln y) = \exp(\ln xy) = xy = \exp a \exp b$$

by Theorem 5.4.2. This proves (a). The proof of (b) is similar and is left to the exercises. $\square$

We define the number e to be $\exp 1$, so that $\ln e = 1$. It follows from (b) of the above theorem that, if r is a rational number, then

$$e^r = (\exp 1)^r = \exp r. \qquad (5.4.1)$$

Now at this point, a^r is defined for every positive a and rational r. We have not yet defined a^x if x is a real number which is not rational. However, $\exp x$ is defined for every real x. Since (5.4.1) tells us that $e^r = \exp r$ if r is rational, it makes sense to *define* e^x for any real x to be $\exp x$.

More generally, if a is any positive real number, then

$$a^r = (\exp \ln a)^r = \exp(r \ln a),$$

and so it makes sense to define a^x for any real x to be $\exp(x \ln a)$. The following definition formalizes this discussion.

Definition 5.4.8. If x is any real number and a is a positive real number, we define a^x by

$$a^x = \exp(x \ln a).$$

In particular,

$$e^x = \exp x.$$

With this definition of a^x, the laws of exponents

$$a^{x+y} = a^x a^y \quad \text{and} \quad a^{xy} = (a^x)^y$$

are satisfied. The proofs are left to the exercises.

The General Logarithm

We define the logarithm to the base a, $\log_a$, to be the inverse function of the function a^x. The following theorem gives a simple description of it in terms of the natural logarithm $\ln x$. The proof is left to the exercises.

Theorem 5.4.9. *For each $a > 0$, we have $\log_a x = \dfrac{\ln x}{\ln a}$.*

Improper Integrals

So far, we have defined the integral $\displaystyle\int_a^b f(x)\,dx$ only for bounded intervals $[a, b]$ and bounded functions f on $[a, b]$. Thus, our definition does not allow for integrals such as

$$\int_0^\infty \frac{1}{1+x^2}\,dx \quad \text{or} \quad \int_0^1 \frac{1}{\sqrt{x}}\,dx.$$

It turns out that a perfectly good meaning can be attached to each of these integrals. To do so requires extending our definition of the integral.

We first consider an integral of the form $\displaystyle\int_a^\infty f(x)\,dx$ where a is finite. We assume that f is integrable on each interval of the form $[a, s]$ for $a \leq s < \infty$. Then we set

$$\int_a^\infty f(x)\,dx = \lim_{s \to \infty} \int_a^s f(x)\,dx,$$

provided this limit exists and is finite. In this case, we say that the *improper integral* $\displaystyle\int_a^\infty f(x)\,dx$ *converges.*

Integrals of the form $\displaystyle\int_{-\infty}^b f(x)\,dx$ are treated similarly. Assuming f is integrable on each interval of the form $[r, b]$ with $-\infty < r \leq b$, we set

$$\int_{-\infty}^b f(x)\,dx = \lim_{r \to -\infty} \int_r^b f(x)\,dx,$$

provided this limit exists and is finite. In this case, we say that the *improper integral* $\displaystyle\int_{-\infty}^b f(x)\,dx$ *converges.*

For an integral of the form $\displaystyle\int_{-\infty}^\infty f(x)\,dx$, we simply break the integral up into a sum of improper integrals involving only one infinite limit of integration. That is, we write

$$\int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^0 f(x)\,dx + \int_0^\infty f(x)\,dx$$

If the two improper integrals on the right converge, we then say the improper integral on the left converges – it converges to the sum on the right.

Example 5.4.10. Find $\displaystyle\int_{-\infty}^{\infty} \frac{1}{1+x^2}$ or show that it fails to converge.

Solution: We write

$$\int_{-\infty}^{\infty} \frac{1}{1+x^2}\,dx = \int_{-\infty}^{0} \frac{1}{1+x^2}\,dx + \int_{0}^{\infty} \frac{1}{1+x^2}\,dx.$$

Then, since $\mathrm{arctan}'(x) = \dfrac{1}{1+x^2}$, the First Fundamental Theorem implies that

$$\int_{-\infty}^{0} \frac{1}{1+x^2}\,dx = \lim_{r\to-\infty} \int_{r}^{0} \frac{1}{1+x^2}\,dx$$
$$= \lim_{r\to-\infty} (\arctan 0 - \arctan r) = \pi/2,$$

and

$$\int_{0}^{\infty} \frac{1}{1+x^2}\,dx = \lim_{s\to\infty} \int_{0}^{s} \frac{1}{1+x^2}\,dx$$
$$= \lim_{s\to\infty} (\arctan s - \arctan 0) = \pi/2,$$

Thus, $\displaystyle\int_{-\infty}^{\infty} \frac{1}{1+x^2}$ converges to π.

Functions With Singularities

If a function f is integrable on $[r, b]$ for every r with $a < r \le b$, but unbounded on the interval $(a, b]$, then it is not integrable on $[a, b]$. It is said to have a *singularity* at a. Still, its improper integral over $[a, b]$ may exist in the sense that

$$\lim_{r\to a^+} \int_{r}^{b} f(x)\,dx$$

may exist and be finite. In this case we say that the improper integral $\displaystyle\int_{a}^{b} f(x)\,dx$ *converges*. Its value, of course, is the indicated limit.

Similarly, a function f may be integrable on $[a, s]$ for every s with $a \le s < b$, but not bounded on $[a, b)$. In this case, its improper integral over $[a, b]$ is

$$\lim_{s\to b^-} \int_{a}^{s} f(x)\,dx$$

provided this limit converges.

It may be that the singular point for f is an interior point c of the interval over which we wish to integrate f. That is, it may be that $a < c < b$ and f is integrable on closed subintervals of $[a, b]$ that don't contain c, but f blows up at c. In this case, we write

$$\int_{a}^{b} f(x)\,dx = \int_{a}^{c} f(x)\,dx + \int_{c}^{b} f(x)\,dx.$$

If the two improper integrals on the right converge, then we say the improper integral on the left converges and it converges to the sum on the right.

Example 5.4.11. Find $\displaystyle\int_{-1}^{1} x^{-1/3}\,dx$.

   Solution: Here the integrand blows up at 0. An antiderivative for $x^{-1/3}$ is $\frac{3}{2}x^{2/3}$. Thus,

$$\int_{-1}^{0} x^{-1/3}\,dx = \lim_{s\to 0^-} \frac{3}{2}(s^{2/3} - (-1)^{2/3}) = -\frac{3}{2},$$

while

$$\int_{0}^{1} x^{-1/3}\,dx = \lim_{r\to 0^+} \frac{3}{2}((1)^{2/3} - (r)^{2/3}) = \frac{3}{2}.$$

Thus,

$$\int_{-1}^{1} x^{-1/3}\,dx = \int_{-1}^{0} x^{-1/3}\,dx + \int_{0}^{1} x^{-1/3}\,dx$$

converges to $-\frac{3}{2} + \frac{3}{2} = 0$.

   The following is a theorem which can be used to conclude that an improper integral converges without actually carrying out the integration.

Theorem 5.4.12. *Let $\displaystyle\int_{a}^{b} f(x)\,dx$ be an improper integral – improper due to the fact that $a = -\infty$ or $b = \infty$ or f has a singularity at a or f has a singularity at b. If g is a non-negative function such that $|f(x)| \leq g(x)$ for all $x \in (a,b)$ and if*

$$\int_{a}^{b} g(x)\,dx$$

converges, then

$$\int_{a}^{b} f(x)\,dx$$

also converges.

Proof. We will prove this in the case where the bad point is b – either $b = \infty$ or f blows up at b. The case where a is the bad point is entirely analogous.
   Let $h(x) = f(x) + |f(x)|$. Then $0 \leq h(x) \leq 2g(x)$ for all $x \in (a,b)$. So

$$H(s) = \int_{a}^{s} h(x)\,dx \quad \text{and} \quad \int_{a}^{s} g(x)\,dx$$

are non-decreasing functions of s (Exercise 5.4.14) and

$$H(s) \leq 2\int_{a}^{s} g(x)\,dx \leq 2\int_{a}^{b} g(x)\,dx.$$

The integral on the right is finite by hypothesis. It follows that the non-decreasing function $H(s)$ is bounded above. By Exercise 4.1.13, $\lim_{s\to b^-} H(s)$ converges, Hence, the improper integral $\int_a^b h(x)\,dx$ converges.

The same argument, with h replaced by $|f(x)|$ shows that $\int_a^b |f(x)|\,dx$ converges. Since $f = h - |f|$, it follows that $\int_a^b f(x)\,dx$ also converges. $\qquad\square$

Example 5.4.13. Determine whether $\int_{-\infty}^{\infty} e^{-x^2}\,dx$ converges.

Solution: Since $e^{-x^2} \leq \dfrac{1}{1+x^2}$ (by Exercise 4.4.3) and each of

$$\int_{-\infty}^0 \frac{1}{1+x^2}\,dx \quad \text{and} \quad \int_0^{\infty} \frac{1}{1+x^2}\,dx$$

converges by Example 5.4.10, the same is true of the corresponding integrals for e^{-x^2}. It follows that $\int_{-\infty}^{\infty} e^{-x^2}\,dx$ converges.

Cauchy Principal Value

Note that we break an improper integral of the form

$$\int_{-\infty}^{\infty} f(x)\,dx \tag{5.4.2}$$

up into the sum of $\int_{-\infty}^0 f(x)\,dx$ and $\int_0^{\infty} f(x)\,dx$ and then require that each of these improper integrals converges before we are willing to say that (5.4.2) converges. This ensures that

$$\lim_{a,b\to\infty} \int_{-a}^b f(x)\,dx$$

exists and is the same number, independently of how a and b approach ∞. This is a strong requirement. In many situations, the improper integral in this sense will fail to converge even though the limit may exist if (a,b) is constrained to lie along some line in the plane. Of special interest is the case when a and b are constrained to be equal. This leads to

$$\lim_{a\to\infty} \int_{-a}^a f(x)\,dx.$$

If this limit exists then we say that the *Cauchy Principal Value* of the improper integral (5.4.2) exists. Similarly, the *Cauchy Principal Value* of an integral over an inteval $[a,b]$ on which f has a singularity at an interior point c is

$$\lim_{r\to 0} \left[\int_a^{c-r} f(x)\,dx + \int_{c+r}^b f(x)\,dx \right]$$

if this limit exists. The existence of the Cauchy principal value is much weaker than ordinary convergence for an improper integral.

Example 5.4.14. Show that the improper integral

$$\int_{-\infty}^{\infty} \frac{x}{1+x^2}\, dx,$$

does not converge but it does have a Cauchy principal value.

   Solution: We have

$$\int_{-\infty}^{\infty} \frac{x}{1+x^2}\, dx = \lim_{a\to\infty} \int_{-a}^{0} \frac{x}{1+x^2}\, dx + \lim_{b\to\infty} \int_{0}^{b} \frac{x}{1+x^2}\, dx$$

The first of the above limits is $\lim_{a\to\infty} -1/2\ln(1+a^2) = -\infty$ while the second is $\lim_{b\to\infty} 1/2\ln(1+b^2) = \infty$. Neither of these converges and so the improper integral does not converge. However, the Cauchy principal value is

$$\lim_{a\to\infty} \int_{-a}^{a} \frac{x}{1+x^2}\, dx = \lim_{a\to\infty} 1/2(\ln a - \ln a) = 0.$$

Exercise Set 5.4

1. Supply the details for the proof of Theorem 5.4.3.

2. Prove that $\ln\left(\dfrac{a}{b}\right) = \ln a - \ln b$ for all $a, b \in (0, +\infty)$.

3. Finish the proof of Theorem 5.4.4 by showing that $\lim_{x\to 0} \ln x = -\infty$. Hint: this follows easily from $\lim_{x\to\infty} \ln x = +\infty$ and properties of ln.

4. Prove Part (b) of Theorem 5.4.7.

5. Using Definition 5.4.8 and the properties of exp prove the laws of exponents:
$$a^{x+y} = a^x a^y \quad \text{and} \quad a^{xy} = (a^x)^y.$$

6. Compute the derivative of a^x for each $a > 0$.

7. Find an antiderivative for a^x for each $a > 0$.

8. Prove Theorem 5.4.9.

9. For which values of $p > 0$ does the improper integral $\displaystyle\int_{1}^{\infty} \frac{1}{x^p}\, dx$ converge. Justify your answer.

10. For which values of $p > 0$ does the improper integral $\displaystyle\int_{0}^{1} \frac{1}{x^p}\, dx$ converge. Justify your answer.

11. Show that $\displaystyle\int_{-\infty}^{\infty} \frac{\sin x}{1 + x^2}$ converges. Can you tell what it converges to?

12. Does the improper integral $\displaystyle\int_{0}^{1} \ln x \, dx$ converge? If so, what does it converge to?

13. Prove that the improper integral $\displaystyle\int_{-\infty}^{\infty} \frac{x^{1/3}}{\sqrt{1 + x^2}} \, dx$ does not converge, but it has Cauchy principal value 0.

14. Prove that if f is an integrable function on every interval $[a, s)$ with $s < b$ and if $f(x) \geq 0$ on $[a, b]$, then the function

$$F(s) = \int_{a}^{s} f(x) \, dx$$

is a non-decreasing function on $[a, b)$.

Chapter 6

Infinite Series

Infinite series play a fundamental role in mathematics. They are used to approximate complicated or uncomputable quantities or functions by simpler quantities or functions. They are widely used by engineers and scientists in real world applications of mathematics.

6.1 Convergence of Infinite Series

An infinite series of numbers is a formal sum

$$\sum_{k=1}^{\infty} a_k = a_1 + a_2 + a_3 + \cdots + a_k + \cdots \tag{6.1.1}$$

of an infinite sequence of numbers a_k called the *terms* of the series. We say *formal sum*, because the actual sum may or may not exist. What does it mean for the actual sum to exist? To answer this, we proceed in much the same way that we did in defining improper integrals. We cut off the sum after some finite number n of terms and then take the limit as $n \to \infty$. That is, we set

$$s_n = \sum_{k=1}^{n} a_k = a_1 + a_2 + a_3 + \cdots + a_n. \tag{6.1.2}$$

The number s_n is called the nth *partial sum* of the series.

Definition 6.1.1. The series (6.1.1) is said to converge to the number s if $\lim s_n = s$. In this case we write

$$\sum_{k=1}^{\infty} a_k = s.$$

The number s is called the *sum* of the series. If the sequence $\{s_n\}$ diverges, then we say the series (6.1.1) diverges.

It is important to keep firmly in mind the difference between a sequence and a series. A *series* is a formal sum of a *sequence* of numbers. Each series

$$a_1 + a_2 + a_3 + \cdots + a_k + \cdots$$

has two sequences associated to it: the sequence of terms $\{a_k\}$ and the sequence of partial sums $\{s_n\}$, where $s_n = a_1 + a_2 + \cdots + a_n$.

A series (6.1.1) converges if and only if its sequence of partial sums converges. What about the sequence of terms $\{a_n\}$? What is the relationship between convergence of the series and convergence of its sequence of terms? The following theorem gives a partial answer.

Theorem 6.1.2. (Term Test) *If a series* $a_1 + a_2 + a_3 + \cdots + a_k + \cdots$ *converges, then* $\lim a_n = 0$.

Proof. If the series converges to s, then $\lim s_n = s$, where $\{s_n\}$ is the sequence of partial sums (6.1.2). However, $a_n = s_n - s_{n-1}$ if $n > 1$, and so

$$\lim a_n = \lim s_n - \lim s_{n-1} = s - s = 0.$$

$\square$

The above theorem is called the term test because it provides a test that the terms of a series must pass if the series converges. If the series fails this test – that is, if $\lim a_n$ either fails to exist or is not 0 if it does exist, then the series diverges. However, this test can never be used to prove that a series converges, since it does *not* say that if $\lim a_n = 0$ then the series converges. In fact, the series

$$1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{k} + \cdots$$

has a sequence of terms $\{1/k\}$ which converges to 0, but the series itself does not converge. This series is called the *harmonic series*. To see that it diverges, group the terms in the following way:

$$(1) + \left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \cdots .$$

Each group in parentheses is a sum of 2^n terms each of which is at least as big as $1/2^{n+1}$. Thus, each group in parentheses sums to a number greater than or equal to $1/2$. It follows that the 2^nth partial sum of the harmonic series is at least $n/2$. Thus, the sequence of partial sums has limit $+\infty$, and so the series diverges.

Example 6.1.3. Does the series $\displaystyle\sum_{k=1}^{\infty} \frac{k}{2k+1}$ converge?

Solution: No. Its sequence of terms is $\left\{\dfrac{k}{2k+1}\right\}$ and this sequence has limit $1/2$ as $k \to \infty$. Since the sequence of terms does not converge to 0, the series fails the term test, and so it diverges.

Example 6.1.4. Does the term test tell us whether $\sum_{k=1}^{\infty} \dfrac{k}{k^2 + 1}$ converges?

   Solution: If we apply the term test, the result is

$$\lim \frac{k}{k^2 + 1} = \lim \frac{1/k}{1 + 1/k^2} = 0.$$

The fact that this limit is 0 tells us nothing. The series may or may not converge (in fact, in Example 6.1.14 we will prove that it diverges).

Remark 6.1.5. Although, in our discussion so far, we have assumed that the index of summation k for a series runs from 1 to ∞, there is really no reason to start the summation at $k = 1$. It could just as easily start at $k = 0$, $k = 2$, or $k = 100$. Our discussion of convergence for series is not effected by where the summation begins, since the only effect on the partial sums s_n of changing the starting point will be to add the same constant to each of them.

Geometric Series

The simplest meaningful series is also one of the most useful. This is the *geometric* series

$$\sum_{k=0}^{\infty} ar^k = a + ar + ar^2 + \cdots + ar^k + \cdots. \tag{6.1.3}$$

Here a and r are any two real numbers. The number a is the initial term of the series, while the number r is called the *ratio* for the geometric series, since, for $k > 1$, it is the ratio of the kth term ar^k to the previous term ar^{k-1}. It is the fact that this ratio is independent of k that characterizes the geometric series.

Theorem 6.1.6. *If $a \neq 0$, the geometric series (6.1.3) converges to $\dfrac{a}{1 - r}$ if $|r| < 1$ and diverges if $|r| \geq 1$.*

Proof. The series fails the term test if $|r| \geq 1$, since $\lim ar^k \neq 0$ in this case. Thus, the geometric series diverges if $|r| \geq 1$.
   Assume $|r| < 1$. If $s_n = a + ar + ar^2 + \cdots + ar^n$ is the nth partial sum of the series, then

$$rs_n = ar + ar^2 + ar^3 + \cdots + ar^{n+1}$$

and so

$$(1 - r)s_n = s_n - rs_n = a - ar^{n+1}.$$

Thus, since $r \neq 1$, we may divide by $1 - r$ to obtain

$$s_n = \frac{a - ar^{n+1}}{1 - r}.$$

This sequence converges to $\dfrac{a}{1 - r}$ since $\lim r^{n+1} = 0$. $\qquad\qquad\square$

Example 6.1.7. Does the series $1/2 + 1/4 + 1/8 + \cdots + 1/2^n + \cdots$ converge? If so what does it converge to?

    Solution: This is a geometric series with ratio $r = 1/2$ and initial term $a = 1/2$. Thus, it converges to $\dfrac{1/2}{1 - 1/2} = 1$, by the previous theorem.

Series with Non-Negative Terms

Let $a_1 + a_2 + \cdots + a_k + \cdots$ be a series with $a_k \geq 0$ for all k. Then, its sequence $\{s_n\}$ of partial sums satisfies

$$s_{n+1} = s_n + a_{n+1} \geq s_n.$$

That is, it is a non-decreasing sequence. If such a sequence is bounded above, then it converges by Theorem 2.4.1. If it is not bounded above, then it has limit $+\infty$. This proves the following theorem.

Theorem 6.1.8. *An infinite series of non-negative terms converges if and only if its sequence of partial sums is bounded above.*

Comparison Test

The comparison test stated in most calculus texts follows easily from the preceding theorem (see Exercise 6.1.11). With a little more work, the following, more general, version of the comparison test can also be proved this way. We give a different proof, based on Cauchy's criterion for convergence.

Theorem 6.1.9. (Comparison Test) *Suppose $a_1 + a_2 + \cdots + a_k + \cdots$ and $b_1 + b_2 + \cdots + b_k + \cdots$ are series, with $b_k \geq 0$ for all k, and suppose there are positive constants K and M such that*

$$|a_k| \leq Mb_k \quad \text{for all} \quad n \geq K. \tag{6.1.4}$$

Then if $b_1 + b_2 + \cdots + b_k + \cdots$ converges, so does $a_1 + a_2 + \cdots + a_k + \cdots$.

Proof. Let $s_n = \displaystyle\sum_{k=1}^{n} a_k$ and $t_n = \displaystyle\sum_{k=1}^{n} b_k$ be the nth partial sums for the two series. If the series with terms b_k converges, then the sequence $\{t_n\}$ converges and, hence, is Cauchy. This implies that , given $\epsilon > 0$, there is an N such that

$$\sum_{k=n+1}^{m} b_k = |t_m - t_n| \leq \frac{\epsilon}{M} \quad \text{whenever} \quad m \geq n > N.$$

Then (6.1.4) implies that

$$|s_m - s_n| = \left| \sum_{k=n+1}^{m} a_k \right| \leq \sum_{k=n+1}^{m} |a_k| \leq M \sum_{k=n+1}^{m} b_k < \epsilon$$

whenever $m \geq n > \max(N, K)$. This implies that $\{s_n\}$ is a Cauchy sequence and, hence, converges. It follows that the series $\sum_{k=1}^{\infty} a_k$ converges. $\square$

Suppose $\sum_{k=1}^{\infty} a_k$ is an arbitrary series. If we set $b_k = |a_k|$, then the condition $|a_k| \leq M b_k$ of the previous theorem is satisfied with $M = 1$ and $K = 1$. This observation yields the following corollary.

Corollary 6.1.10. *If $\sum_{k=1}^{\infty} |a_k|$ converges, then so does $\sum_{k=1}^{\infty} a_k$.*

This leads to the following definition.

Definition 6.1.11. A series $\sum_{k=1}^{\infty} a_k$ is said to *converge absolutely* if the series $\sum_{k=1}^{\infty} |a_k|$ converges.

Thus, Corollary 6.1.10 asserts that if a series converges absolutely, then it converges.

Example 6.1.12. Does the series $\sum_{k=1}^{\infty} \dfrac{k}{2^k}$ converge? Why?

Solution: Since $\lim \dfrac{k}{2^{k/2}} = 0$ (l'Hôpital's Rule), there is an N such that

$$\frac{k}{2^{k/2}} < 1 \quad \text{whenever} \quad k > N.$$

Then

$$\frac{k}{2^k} < \frac{1}{2^{k/2}} = \frac{1}{(\sqrt{2})^k} \quad \text{whenever} \quad k > N.$$

Since the series $\sum_{k=1}^{\infty} \dfrac{1}{(\sqrt{2})^k}$ is a convergent geometric series, the series $\sum_{k=1}^{\infty} \dfrac{k}{2^k}$ converges by the comparison test.

Example 6.1.13. Does the series $\sum_{k=1}^{\infty} (-1)^k \dfrac{k}{2^k}$ converge? Why?

Solution: By the previous exercise, the series $\sum_{k=1}^{\infty} \dfrac{k}{2^k}$ converges and this means that $\sum_{k=1}^{\infty} (-1)^k \dfrac{k}{2^k}$ converges absolutely and, hence, converges by Corollary 6.1.10.

The comparison test can also be used to prove that a series diverges.

Example 6.1.14. Prove that the series $\sum_{k=1}^{\infty} \dfrac{k}{k^2+1}$ diverges.

Solution: We compare with the harmonic series. Since $k^2 + 1 \leq 2k^2$ for $k \in \mathbb{N}$, we have

$$\frac{1}{k} \leq 2\frac{k}{k^2+1} \quad \text{for all} \quad k \in \mathbb{N}.$$

If the series $\sum_{k=1}^{\infty} \dfrac{k}{k^2+1}$ converges, then so does $\sum_{k=1}^{\infty} \dfrac{1}{k}$ by the comparison test.

However, the harmonic series diverges. Therefore $\sum_{k=1}^{\infty} \dfrac{k}{k^2+1}$ also diverges.

Exercise Set 6.1

In each of the following six exercises, determine whether the indicated series converges. Justify your answer.

1. $\sum_{k=2}^{\infty} \dfrac{k-1}{2k+1}.$

2. $\sum_{k=1}^{\infty} \dfrac{1}{2^k+k-1}.$

3. $\sum_{k=0}^{\infty} \dfrac{2^{k+1}}{3^k}.$

4. $\sum_{k=1}^{\infty} \dfrac{k^2-3k+1}{3k^2+k-2}.$

5. $\sum_{k=1}^{\infty} \dfrac{k^2}{4^k}.$

6. $\sum_{k=1}^{\infty} \dfrac{k}{k^2-k+2}.$

In each of the next four exercises, determine whether the indicated series converges absolutely. Justify your answer.

7. $\sum_{k=0}^{\infty} (-2/3)^k.$

8. $\sum_{k=1}^{\infty} \dfrac{(-1)^{k+1}}{\sqrt{k}}.$

9. $\displaystyle\sum_{k=1}^{\infty} \frac{\sin k}{2^k}$.

10. $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^k}{\ln(1+k)}$.

11. Prove the following weak version of the comparison test using Theorem 6.1.8: If $a_1 + a_2 + \cdots + a_k + \cdots$ and $b_1 + b_2 + \cdots + b_k + \cdots$ are series of non-negative terms with $a_k \leq b_k$ for all k, then if $b_1 + b_2 + \cdots + b_k + \cdots$ converges, so does $a_1 + a_2 + \cdots + a_k + \cdots$.

12. Consider the decimal expansion $.d_1 d_2 d_3 d_4 \cdots$ of a real number between 0 and 1, where $\{d_k\}$ is a sequence of integers between 0 and 9. This decimal expansion represents the sum of a certain infinite series. What series is it and why does it converge?

13. Show that every real number in the interval $[0, 1]$ has a decimal expansion as described in the previous exercise.

6.2 Tests for Convergence

In this section we will develop the standard tests for convergence of infinite series. Most of these are based on Theorem 6.1.8 or Theorem 6.1.9.

Integral Test

Theorem 6.2.1. *Suppose f is a positive, non-increasing function on $[1, \infty)$ and $a_k = f(k)$ for each $k \in \mathbb{N}$. Then the series $\displaystyle\sum_{k=1}^{\infty} a_k$ converges if and only if the improper integral $\displaystyle\int_1^{\infty} f(x)\,dx$ converges.*

Proof. Consider the function $g(x)$ on $[1, \infty)$ which, for each $k \in \mathbb{N}$, is constant on the interval $[k, k+1)$ and equal to $f(k)$ at k. That is,

$$g(x) = f(k) = a_k \quad \text{if} \quad k \leq x < k+1, \ k \in \mathbb{N}.$$

This is a piecewise continuous function, hence integrable on any finite interval $[1, b)$. Also, since f is non-increasing, it follows that

$$g(x+1) \leq f(x) \leq g(x) \quad \text{for all} \quad x \in [1, \infty).$$

(see Figure 6.1). On integrating from 1 to n, this yields

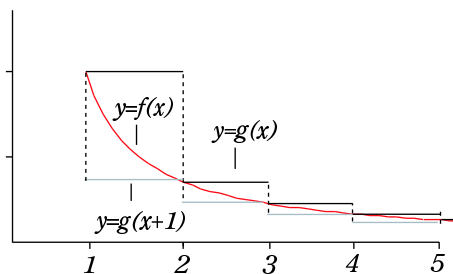$$\int_1^n g(x+1)\,dx \leq \int_1^n f(x)\,dx \leq \int_1^n g(x)\,dx.$$

Figure 6.1: Setup for Proof of the Integral Test.

However, by Exercise 6.2.9,

$$\int_1^n g(x+1)\,dx = \sum_{k=2}^n a_k \quad \text{and} \quad \int_1^n g(x)\,dx = \sum_{k=1}^{n-1} a_k. \qquad (6.2.1)$$

If $s_n = \sum_{k=1}^n a_k$, then this implies that

$$s_n - a_1 \leq \int_1^n f(x)\,dx \leq s_{n-1}.$$

It follows that the sequence of partial sums $\{s_n\}$ is bounded above if and only if the increasing function of b, $\int_1^b f(x)\,dx$, is bounded above. A non-decreasing sequence converges if and only if it is bounded above and a non-decreasing function on $[1,\infty)$ has a finite limit at ∞ if and only if it is bounded above. Thus, the series converges if and only if the improper integral converges. $\square$

Example 6.2.2. A p-series is a series of the form $\sum_{k=1}^\infty \dfrac{1}{k^p}$, where $p > 0$. Prove that a p-series converges if and only if $p > 1$.

Solution We apply the integral test for the function $f(x) = 1/x^p$. Note that this is a positive, decreasing function on $[1,\infty)$ and $f(k) = 1/k^p$ for $k \in \mathbb{N}$. If $p \neq 1$ we have

$$\int_1^b \frac{1}{x^p}\,dx = \frac{b^{1-p} - 1}{1 - p}.$$

As $b \to \infty$, this has limit $\dfrac{1}{p-1}$ if $p > 1$ and $+\infty$ if $p < 1$. Thus, the p-series converges for $p > 1$ and diverges for $p < 1$ by the Integral Test.

For $p = 1$, the p-series is the harmonic series and we already know it diverges. However, it is instructive to see how this follows from the Integral Test.

In the case $p = 1$, the function f is $f(x) = 1/x$. We have

$$\int_1^b \frac{1}{x}\, dx = \ln b,$$

and this has limit $+\infty$ as $b \to \infty$. Thus, applying the Integral Test gives another proof that the harmonic series $\sum_{k=1}^{\infty} \frac{1}{k}$ diverges.

Example 6.2.3. Does the series $\sum_{k=1}^{\infty} \frac{3\sqrt{k}}{2k^2 - 1}$ converge or diverge. Justify your answer.

Solution: For large k, $\frac{3\sqrt{k}}{2k^2 - 1}$ is close to $\frac{3}{k^{3/2}}$. This suggests we do a comparison with the p-series $\sum_{k=1}^{\infty} \frac{1}{k^{3/2}}$.

We have $2k^2 - 1 \geq k^2$ for all $k \geq 1$ and so

$$\frac{3\sqrt{k}}{2k^2 - 1} \leq \frac{3\sqrt{k}}{k^2} = \frac{3}{k^{3/2}}.$$

Since the p-series with $p = 3/2$ converges, so does our series, by the comparison test.

Root Test

This test is particularly important in the study of power series.

Theorem 6.2.4. *Given an infinite series $\sum_{k=1}^{\infty} a_k$, let*

$$\rho = \limsup |a_k|^{1/k}.$$

Then the series converges absolutely if $\rho < 1$ and diverges if $\rho > 1$.

Proof. Recall that

$$\limsup |a_k|^{1/k} = \lim t_n \quad \text{where} \quad t_n = \sup\{|a_k|^{1/k} : k \geq n\}.$$

Also recall that $\{t_n\}$ is a non-increasing sequence. Thus, if $\rho > 1$, then

$$t_n = \sup\{|a_k|^{1/k} : k \geq n\} > 1 \quad \text{for all} \quad n \in \mathbb{N}.$$

This means that, for every $n \in \mathbb{N}$, there is an $k \geq n$ such that $|a_k|^{1/k} > 1$. Then $|a_k| > 1$ also. It follows that the sequence of terms $\{a_k\}$ does not have limit 0. Hence, the series fails the term test and must diverge in this case.

If $\rho < 1$, we can choose r such that $\rho < r < 1$. Then there is an N such that

$$t_n < r \quad \text{whenever} \quad n > N$$

and this implies that

$$|a_k|^{1/k} < r \quad \text{whenever} \quad k > N.$$

This, in turn, implies that

$$|a_k| < r^k \quad \text{whenever} \quad k > N.$$

Thus, the series $\sum_{k=1}^{\infty} |a_k|$ converges in this case, by comparison with the geometric series with ratio $r < 1$. Therefore, the original series converges absolutely. $\square$

Note that the root test tells us nothing about convergence if the number ρ turns out to be 1.

Example 6.2.5. Does the series $\sum_{k=1}^{\infty} k(9/10)^k$ converge? Why?
   Solution: We apply the root test. In this case, the lim sup of Theorem 6.2.4 is actually a limit, since the limit exists. In fact,

$$\rho = \lim k^{1/k}(9/10) = (9/10) \lim k^{1/k} = 9/10 < 1,$$

since $\lim k^{1/k} = 1$ by Exercise 2.3.12. By the root test, the series converges.

Ratio Test

Theorem 6.2.6. *Given a series* $\sum_{k=1}^{\infty} a_k$, *let*

$$r = \lim \frac{|a_{k+1}|}{|a_k|} \tag{6.2.2}$$

provided this limit exists.   Then the series converges absolutely if $r < 1$ and diverges if $r > 1$.

Proof. Observe first that, for the limit defining r to exist, the numbers a_k must eventually be all non-zero – otherwise, the ratio $|a_{k+1}|/|a_k|$ would be undefined or $+\infty$ for infinitely many k.
   If $r > 1$, then there is an N such that

$$|a_k| > 0 \quad \text{and} \quad \frac{|a_{k+1}|}{|a_k|} > 1 \quad \text{for all} \quad k \geq N.$$

Then, for $k > N$

$$|a_k| = \frac{|a_k|}{|a_{k-1}|}\frac{|a_{k-1}|}{|a_{k-2}|} \cdots \frac{|a_{N+2}|}{|a_{N+1}|}\frac{|a_{N+1}|}{|a_N|}|a_N| > |a_N|.$$

This implies the sequence of terms $\{a_k\}$ fails to have limit 0, and the sequence diverges by the term test.

If $r < 1$ we choose a t such that $r < t < 1$. Since (6.2.2) holds, there is an N such that

$$\frac{|a_{k+1}|}{|a_k|} < t \quad \text{whenever} \quad n \geq N.$$

Then, for $k > N$,

$$|a_k| = \frac{|a_k|}{|a_{k-1}|} \frac{|a_{k-1}|}{|a_{k-2}|} \cdots \frac{|a_{N+2}|}{|a_{N+1}|} \frac{|a_{N+1}|}{|a_N|} |a_N| < t^{k-N} |a_N|.$$

Thus, $|a_k| < t^k \dfrac{|a_N|}{t^N}$ whenever $k > N$. By comparison with the geometric series with ratio t, the series converges. $\qquad\square$

The ratio test tends to work well on series where the terms a_k involve products of an increasing number of factors – things like factorials. These are generally more difficult to attack with the root test than with the ratio test.

Example 6.2.7. Does the series $\displaystyle\sum_{k=1}^{\infty} \frac{k!}{k^k}$ converge? Why?

Solution: We apply the ratio test.

$$r = \lim \frac{(k+1)!}{(k+1)^{k+1}} \div \frac{k!}{k^k} = \lim \frac{(k+1)! k^k}{(k+1)^{k+1} k!}$$

$$= \lim \left(\frac{k}{k+1} \right)^k = \lim \frac{1}{(1 + 1/k)^k} = \frac{1}{e} < 1.$$

Hence, the series converges by the ratio test.

For many series, the ratio test and the root test work equally well. However, the ratio test is not applicable in many situations where the root test works well.

Example 6.2.8. Prove that the series $1/3 + 1/2^2 + 1/3^3 + 1/2^4 + 1/3^5 + \cdots$ converges.

Solution: This one can easily be done using the comparison test. However, it is instructive to see how attempts to use the ratio test and root test work out. The ratio test doesn't work, because the successive ratios are

$$3/4, 4/27, 27/16, 16/243, 243/64 \cdots ,$$

and this sequence of numbers has no limit.

On the other hand, the root test yields that ρ is the lim sup of the sequence

$$1/3, 1/2, 1/3, 1/2, 1/3, \cdots .$$

That is, $\rho = 1/2$. Therefore, the series converges by the root test.

Exercise Set 6.2

In each of the following eight exercises, determine whether the indicated series
converges. Justify your answer by indicating what test to use and then carrying
out the details of the application of that test.

1. $\displaystyle\sum_{k=2}^{\infty} \frac{1}{k \ln k}$.

2. $\displaystyle\sum_{k=1}^{\infty} \frac{\ln k}{k^2}$

3. $\displaystyle\sum_{k=1}^{\infty} \frac{k2^k}{3^k}$.

4. $\displaystyle\sum_{k=0}^{\infty} \frac{5^k}{k!}$.

5. $\displaystyle\sum_{k=1}^{\infty} \frac{k}{(3+(-1)^k)^k}$.

6. $\displaystyle\sum_{k=1}^{\infty} \frac{k!}{4^k}$.

7. $\displaystyle\sum_{k=1}^{\infty} \frac{\sqrt{k}}{k^2 - k + 2}$.

8. $\displaystyle\sum_{k=1}^{\infty} k\,e^{-\sqrt{k}}$.

9. Verify the integral formulas (6.2.1) used in the proof of the Integral Test.

10. Prove that if $\displaystyle\sum_{k=1}^{\infty} a_k$ and $\displaystyle\sum_{k=1}^{\infty} b_k$ are convergent series and c is a constant,
    then $\displaystyle\sum_{k=1}^{\infty} ca_k$ and $\displaystyle\sum_{k=1}^{\infty} (a_k + b_k)$ are also convergent. Furthermore,

$$\sum_{k=1}^{\infty} ca_k = c\sum_{k=1}^{\infty} a_k, \quad \text{and}$$

$$\sum_{k=1}^{\infty} (a_k + b_k) = \sum_{k=1}^{\infty} a_k + \sum_{k=1}^{\infty} b_k.$$

11. Prove that if $\sum_{k=1}^{\infty} a_k$ conveges absolutely and $\{b_k\}$ is a bounded sequence, then $\sum_{k=1}^{\infty} a_k b_k$ also converges absolutely.

12. Prove that if $\sum_{k=1}^{\infty} a_k$ and $\sum_{k=1}^{\infty} b_k$ are series and $a_k = b_k$ except for finitely many values of k, then the two series either both converge or they both diverge.

6.3   Absolute and Conditional Convergence

By Corollary 6.1.10, if a series converges absolutely, then it converges. The converse is not true. As we shall see, it is possible for a series to converge even though the corresponding series of absolute values does not converge.

Definition 6.3.1. A series which converges, but does not converge absolutely is said to converge *conditionally*.

Thus, a conditionally convergent series is one which converges, but its corresponding series of absolute values does not converge. For examples of conditionally convergent series, we turn to alternating series.

Alternating Series

An alternating series is one in which the terms alternate in sign – each positive term is followed by a negative term and vice-verse. Under reasonable additional conditions, such a series will converge.

Theorem 6.3.2. (Alternating Series Test) *Let $\{a_k\}$ be a non-increasing sequence of non-negative numbers which converges to 0. Then the series*

$$\sum_{k=1}^{\infty} (-1)^{k+1} a_k = a_1 - a_2 + a_3 - a_4 + \cdots$$

converges. In fact, if s_n is the nth partial sum of this series and $s = \lim s_n$, then

$$|s - s_n| \leq a_{n+1} \quad \text{for all} \quad n.$$

Proof. Since $\{a_k\}$ is a non-increasing sequence of non-negative numbers, we have $a_k - a_{k+1} \geq 0$ for all k. For n odd, this means

$$s_{n+1} \leq s_{n+1} + a_{n+2} = s_{n+2} = s_n - (a_{n+1} - a_{n+2}) \leq s_n.$$

That is,

$$s_{n+1} \leq s_{n+2} \leq s_n \quad \text{for odd} \quad n.$$

Similarly,

$$s_n \leq s_{n+2} \leq s_{n+1} \quad \text{for even} \quad n.$$

Thus, $s_2 \leq s_3 \leq s_1$ and, after that, each term of the sequence $\{s_n\}$ lies between the previous two terms. It follows that

$$s_2 \leq s_4 \leq s_6 \leq \cdots \leq s_{2n} \leq s_{2n+1} \leq \cdots s_5 \leq s_3 \leq s_1.$$

Hence, the subsequence of $\{s_n\}$ consisting of terms with odd index n forms a non-increasing sequence which is bounded below, while the subsequence of terms with even index n forms a non-decreasing sequence which is bounded above. These two monotone, bounded sequences converge, and they must converge to the same limit s because

$$|s_{n+1} - s_n| = a_{n+1}$$

and the sequence $\{a_n\}$ converges to 0. Since s is between s_n and s_{n+1} for each n, this also shows that

$$|s - s_n| \leq a_{n+1},$$

as claimed. $\square$

An alternating p-series is a series of the form

$$1 - \frac{1}{2^p} + \frac{1}{3^p} - \cdots + (-1)^{k-1}\frac{1}{k^p} + \cdots .$$

where $p > 0$.

Example 6.3.3. Show that each alternating p-series with $0 < p \leq 1$ converges conditionally.

Solution: The alternating p-series satisfies the conditions of the alternating series test, since $\{1/k^p\}$ is a decreasing sequence which converges to 0. Thus, the alternating p-series converges for all $p > 0$. However, the ordinary p-series $\sum_{k=1}^{\infty} \frac{1}{k^p}$ diverges if $p \leq 1$ (Example 6.2.2). Thus, the alternating p-series converges conditionally for $0 < p \leq 1$.

In particular, the alternating harmonic series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + (-1)^{k-1}\frac{1}{k} + \cdots$$

converges conditionally.

Absolute verses Conditional Convergence

Absolute convergence is a much stronger condition than conditional convergence. The importance of the concept of absolute convergence stems from the fact that, if the terms of an absolutely convergent series are rearranged to form a new series, then the new series converges to the same number as the original series (Theorem 6.3.5 below). This is not true of conditionally convergent

series – in fact, it fails spectacularly. A conditionally convergent series can be rearranged so as to diverge to ∞ or $-\infty$ or to converge to any given number (Theorem 6.3.4 below).

By a *rearrangement* of a series $\sum_{k=1}^{\infty} a_k$ we mean a series of the form $\sum_{j=1}^{\infty} a_{k(j)}$, where $k(j)$ is a one-to-one function from $\mathbb{N}$ onto $\mathbb{N}$. In other words, the rearranged series has exactly the same terms as the original series, but arranged in a different order.

Theorem 6.3.4. *A conditionally convergent series has, for each extended real number L, a rearrangement that converges to L.*

Proof. If $\sum_{k=1}^{\infty} a_k$ is a conditionally convergent series, then by Exercise 6.3.7, the series of positive terms of this series diverges, as does the series of negative terms. Since the series of positive terms diverges, its sequence of partial sums is unbounded and, hence, has limit ∞. Similarly, for the series of negative terms, the partial sums have limit $-\infty$.

We will prove the theorem in the case where L is a real number. The cases where L is ∞ or $-\infty$ are left to the exercises.

Given a number L, we will define a sequence $\{b_j\}$ inductively in the following way: We let b_1 be the first positive term in $\{a_k\}$ if $0 < L$ and the first non-positive term in $\{a_k\}$ if $L \leq 0$. Suppose $b_1, b_2, \cdots, b_n$ have been chosen. We set

$$s_n = \sum_{j=1}^{n} b_j$$

and choose b_{n+1} according to the following rule: If $s_n < L$ we choose b_{n+1} to be the first positive term in $\{a_k\}$ that has not already been used. If $L \leq s_n$ we choose b_{n+1} to be the first non-positive term in $\{a_k\}$ that has not already been used. This defines the sequence $\{b_j\}$ inductively. The series $\sum_{j=1}^{\infty} b_j$ defined in this way has the following properties:

(1) Each successive partial sum s_n is either as close or closer to L than its predecessor s_{n-1}, or one of them is less than L and the other is greater than or equal to L. In the latter case, the distance from s_n to L is less than $|s_n - s_{n-1}| = |b_n|$. We call n a *crossing* integer in this case.

(2) There are infinitely many crossing integers. Our description of $\sum_{j=1}^{\infty} b_j$ involves adding successive positive terms until we reach or exceed L and then adding successive non-positive terms until we fall below L. Since the series of positive terms and the series of negative terms both diverge, no matter where a given partial sum lies we will always be able to add enough of the remaining positive terms to reach or exceed L or add enough of the remaining non-positive terms to fall below L. Thus, crossing L will occur infinitely often.

(3) All the terms of $\{a_k\}$ are used in constructing the sequence $\{b_j\}$, since at each step we are selecting the first positive term not already chosen or the first non-positive term not already chosen and both cases occur infinitely often. Thus, each a_k will be chosen eventually. Also, at each stage we only choose from the terms not already chosen, and so each a_k will be used just once. This means that the sequence $\{b_j\}$ is a rearrangement of the sequence $\{a_k\}$.

(4) Since $\sum_{k=1}^{\infty} a_k$ converges, we have $\lim a_k = 0$, and this implies $\lim b_j = 0$ also. This is proved as follows: If $\epsilon > 0$, there is an N such that $|a_k| < \epsilon$ whenever $k > N$. However, if we choose M to be an integer such that, by stage M in our construction all the terms $a_1, a_2, \cdots, a_N$ have been chosen, then $j > M$ implies that b_j is not one of these terms and, hence, is a term a_k with $k > N$. This, in turn, implies that $|b_j| < \epsilon$.

Now (1) and (2) and (4) imply that $\lim s_n = L$. That is, the crossing integers define a subsequence of $\{s_n\}$ (by (2)) that is converging to L (by (1) and (4)) and, between two successive crossing integers, the sequence $\{s_n\}$ stays at least as close to L as it was at the first crossing integer of the pair (by (1)).

Thus, $\sum_{k=1}^{\infty} b_k$ is a rearrangement of $\sum_{k=1}^{\infty} a_k$ which converges to L. $\qquad\square$

The above theorem illustrates that a conditionally convergent series is a rather unstable object, since its sum is dependent on the order in which the terms are added. On the other hand, an absolutely convergent series is quite stable in the sense that the sum is always the same regardless of the order in which the terms are summed. That is the content of the next theorem.

Theorem 6.3.5. *Each rearrangement of an absolutely convergent series converges to the same number as the original series.*

Proof. Let $\sum_{k=1}^{\infty} a_k$ be an absolutely convergent series which converges to the number s. Since this series is absolutely convergent, the series $\sum_{k=1}^{\infty} |a_k|$ also converges to a number t. The difference between t and the nth partial sum of this series is

$$\sum_{k=n+1}^{\infty} |a_k|.$$

Because the partial sums converge to t, given $\epsilon > 0$, there is an N such that

$$\sum_{k=n+1}^{\infty} |a_k| < \epsilon/2 \quad \text{for all} \quad n > N. \tag{6.3.1}$$

We fix one such n, and we also choose it to be large enough so that

$$\left| s - \sum_{k=1}^{n} a_k \right| < \epsilon/2. \tag{6.3.2}$$

Now suppose $\sum_{j=1}^{\infty} b_j$ is a rearrangement of $\sum_{k=1}^{\infty} a_k$. Then $b_j = a_{k(j)}$ for some one-to-one function $k(j)$ of $\mathbb{N}$ onto $\mathbb{N}$. Let J be the largest value of j for which $k(j) \leq n$. Then the terms $a_1, a_2, \cdots, a_n$ of the original series all appear as terms in the partial sum $\sum_{j=1}^{m} b_j$ as long as $m \geq J$. For such an m, the expression

$$\sum_{j=1}^{m} b_j - \sum_{k=1}^{n} a_k$$

is a finite sum of terms a_k with $k > n$. By (6.3.1) and the triangle inequality, such a sum must have absolute value less than $\epsilon/2$. This, combined with (6.3.2), implies that

$$\left| s - \sum_{j=1}^{m} b_j \right| < \epsilon \quad \text{whenever} \quad m \geq J.$$

Hence, the series $\sum_{j=1}^{\infty} b_j$ also converges to s. $\qquad \square$

Products of Series

In calculus we are taught how to multiply two power series. The formula for doing this relies on the following result, which requires that the two series be absolutely convergent (see Exercise 6.3.12).

Theorem 6.3.6. *Let* $\sum_{k=0}^{\infty} a_k$ *and* $\sum_{j=0}^{\infty} b_j$ *be two absolutely convergent series. Then*

$$\left(\sum_{k=0}^{\infty} a_k \right) \left(\sum_{j=0}^{\infty} b_j \right) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} a_k b_{n-k}, \qquad (6.3.3)$$

where the series on the right also converges absolutely.

Proof. Consider the set $S = \{a_k b_j : j, k \in \mathbb{N}\}$. The numbers in this set can be displayed in an infinite array or *matrix* as follows:

$$\begin{array}{cccccc}
a_0 b_0 & a_1 b_0 & a_2 b_0 & \cdots & a_n b_0 & \cdots \\
a_0 b_1 & a_1 b_1 & a_2 b_1 & \cdots & a_n b_1 & \cdots \\
a_0 b_2 & a_1 b_2 & a_2 b_2 & \cdots & a_n b_2 & \cdots \\
\vdots & \vdots & \vdots & \cdots & \vdots & \cdots \\
a_0 b_n & a_1 b_n & a_2 b_n & \cdots & a_n b_n & \cdots \\
\vdots & \vdots & \vdots & \cdots & \vdots & \cdots
\end{array} \qquad (6.3.4)$$

The sum of the absolute values of the members of any finite subset of this set is less than

$$M = \left(\sum_{k=0}^{\infty} |a_k| \right) \left(\sum_{j=0}^{\infty} |b_j| \right) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} |a_k b_j|.$$

Since M is finite, this means that, given any series formed by summing the elements of S in some order, the corresponding series of absolute values will have partial sums bounded above by M. Such a series must converge. Thus, each series formed by summing the elements of S in some order will be absolutely convergent, and all such series will converge to the same number by the previous theorem.

One series formed by summing the elements of S is

$$a_0 b_0 + a_0 b_1 + a_1 b_1 + a_1 b_0 + a_0 b_2 + a_1 b_2 + a_2 b_2 + a_2 b_1 + a_2 b_0 + \cdots .$$

That is, in the array (6.3.4), for succesive values of n, we sum from left to right along the nth row to the main diagonal and then along nth column from the main diagonal back to the top row. The n^2 partial sum of this sequence is

$$\left(\sum_{k=0}^{n} a_k \right) \left(\sum_{j=0}^{n} b_j \right) = \sum_{j=0}^{n} \sum_{k=0}^{n} a_k b_j.$$

This sequence of numbers converges to the left side of equation (6.3.3).

Another way of summing the numbers in the set S yields the series

$$\sum_{n=0}^{\infty} \sum_{k=0}^{n} a_k b_{n-k}.$$

This is obtained by summing the array (6.3.4) along diagonals of the form $k + j = n$ for successive values of n. The resulting sum is the right side of equation (6.3.3). Since these two series must sum to the same number by the previous theorem, Equation (6.3.3) is true. $\qquad\square$

Exercise Set 6.3

In each of the next five exercises, determine whether the given series converges absolutely, converges conditionally, or diverges. Justify your answer.

1. $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^k}{k^{1/3}}$.

2. $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^2}$.

3. $\displaystyle\sum_{k=2}^{\infty} \frac{(-1)^k}{\ln(k)}$.

4. $\displaystyle\sum_{k=1}^{\infty} (-1)^{k-1} \frac{2^k}{2^k + k^2}$.

5. $\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k^{2+(-1)^k}}$.

6. Give an example of two convergent series $\displaystyle\sum_{k=1}^{\infty} a_k$ and $\displaystyle\sum_{k=1}^{\infty} b_k$ such that the series $\displaystyle\sum_{k=1}^{\infty} a_k b_k$ diverges.

7. If $\displaystyle\sum_{k=1}^{\infty} a_k$ is a series, we set $a_k^+ = a_k$ if $a_k > 0$, $a_k^+ = 0$ if $a_k \le 0$ and $a_k^- = a_k$ if $a_k < 0$, $a_k^- = 0$ if $a_k > 0$. Prove that if the series is conditionally convergent, then both $\displaystyle\sum_{k=1}^{\infty} a_k^-$ and $\displaystyle\sum_{k=1}^{\infty} a_k^+$ diverge.

8. Approximate the sum of the alternating harmonic series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots + (-1)^{n-1}\frac{1}{n} + \cdots$$

to within .01 by computing an appropriate partial sum. You will need a calculator or computer.

9. For the alternating harmonic series of the preceding exercise, follow the procedure used in the proof of Theorem 6.3.4 to rearrange the series so that it converges to $\sqrt{2}$. Carry out this procedure until the partial sum of your new series is within .02 of $\sqrt{2}$. You will need a calculator or a computer.

10. Show how to modify the proof of Theorem 6.3.4 to cover the cases $L = \infty$ and $L = -\infty$.

11. The geometric series $\displaystyle\sum_{k=0}^{\infty} 2^{-k}$ converges to 2. Use the product formula of Theorem 6.3.6 to show that the series $\displaystyle\sum_{k=0}^{\infty} (k+1)2^{-k}$ converges to 4.

12. Show that the product formula in Theorem 6.3.6 may fail to be true if the series involved are not absolutely convergent. Hint: consider the case where both series are $\displaystyle\sum_{k=0}^{\infty} \frac{(-1)^k}{\sqrt{k+1}}$.

6.4   Power Series

One of the most useful and widely used techniques of modern mathematics is that of expressing a complicated function as the sum of a series of simple functions. Examples include power series, Fourier series, and various eigenfunction expansions for differential equations. All involve series whose terms are functions rather than numbers.

Series of Functions

Consider a series of the form

$$\sum_{k=1}^{\infty} f_k(x) = f_1(x) + f_2(x) + f_3(x) + \cdots + f_k(x) + \cdots, \qquad (6.4.1)$$

where I is an interval in $\mathbb{R}$ and each of the functions $f_k(x)$ is a function defined on I. For each fixed value of $x \in I$, this is just an ordinary series of numbers and it may or may not converge. The series may converge for some values of x and not for others. On the subset of I for which the series does converge, it defines a new function

$$g(x) = \sum_{k=1}^{\infty} f_k(x).$$

This function is the limit of the sequence of functions

$$g_n(x) = \sum_{k=1}^{n} f_k(x)$$

obtained by taking the partial sums of the series.

There are many questions one can ask about this situation: if the functions $f_k(x)$ are continuous or differentiable, is the same thing true of the function g that the series converges to? Can we integrate the function g over a subinterval of I by integrating the series term by term? When can we differentiate g by differentiating the series term by term? We can give satisfactory answers to a couple of these questions right away.

Definition 6.4.1. We say a series of functions (6.4.1) converges *uniformly* to g on I if its sequence of partial sums $\{g_n\}$ converges uniformly to g.

Theorem 6.4.2. *If each f_k is a continuous function on I and the series (6.4.1) converges uniformly to g on I, then g is also continuous on I.*

Proof. If the series (6.4.1) converges uniformly to g on I, then its sequence of partial sums $\{g_n\}$ converges uniformly to g on I. Each g_n is a finite sum of functions f_k which are continuous on I and, hence, is also continuous on I. Since the limit of a uniformly convergent sequence of continuous functions is continuous (Theorem 3.4.4) , we conclude that g is continuous on I. $\qquad \square$

The proof of the next theorem is very similar – the theorem follows directly from the analogous result about integrating the uniform limit of a sequence of functions (Exercise 5.2.13). We leave the details to the exercises.

Theorem 6.4.3. *If each f_k is continuous on $[a, b]$ and the series (6.4.1) converges uniformly to g on $[a, b]$, then*

$$\int_a^b g(x)\, dx = \sum_{k=1}^{\infty} \int_a^b f_k(x)\, dx.$$

This means, in particular, that the series on the right converges.

Weierstrass M-test

The following is a test for uniform convergence of a series. It follows from an analogous test for uniform convergence of sequences (Theorem 3.4.6).

Theorem 6.4.4. (Weierstrass M-test) *A series of functions (6.4.1) on an interval I converges uniformly on I if there is a convergent series of positive terms*

$$\sum_{k=1}^{\infty} M_k$$

such that $|f_k(x)| \le M_k$ for all $x \in I$ and all $k \in \mathbb{N}$.

Proof. By the comparison test, at each x the series (6.4.1) converges to a number $g(x)$. If

$$g_n(x) = \sum_{k=1}^{n} f_k(x)$$

then

$$|g(x) - g_n(x)| = \left| \sum_{k=n+1}^{\infty} f_k(x) \right| \le \sum_{k=n+1}^{\infty} |f_k(x)|$$

$$\le \sum_{k=n+1}^{\infty} M_k = S - S_n$$

where S and S_n are the sum and nth partial sum of the series $\sum_{k=1}^{\infty} M_k$. Since this series converges, $\lim(S - S_n) = 0$. The theorem now follows from Theorem 3.4.6. ∎

Example 6.4.5. Analyze the Fourier Series

$$\sum_{k=1}^{\infty} \frac{\cos kx}{k^2},$$

using the preceding three theorems.

Solution: We have $\left| \dfrac{\cos kx}{k^2} \right| \leq \dfrac{1}{k^2}$ for all $x \in \mathbb{R}$. The series $\sum\limits_{k=1}^{\infty} \dfrac{1}{k^2}$ converges, since it is a p-series with $p > 1$. Thus, it follows from the Weierstrass M-test that the series $\sum\limits_{k=1}^{\infty} \dfrac{\cos kx}{k^2}$ converges uniformly on $\mathbb{R}$. The function g that it converges to is continuous on $\mathbb{R}$ by Theorem 6.4.2. On every bounded interval $[a, b]$, we have

$$\int_a^b g(x)\,dx = \sum_{k=1}^{\infty} \frac{1}{k^2} \int_a^b \cos kx\,dx = \sum_{k=1}^{\infty} \frac{1}{k^3}(\sin kb - \sin ka),$$

also by Theorem 6.4.2.

Power Series

A *power series* centered at a is a series of the form

$$\sum_{k=0}^{\infty} c_k(x - a)^k \tag{6.4.2}$$

This is a series with terms $c_k(x - a)^k$ which are very simple – they are simple monomials in $(x - a)$ and, hence, each is defined on all of $\mathbb{R}$, is continuous and, in fact, has derivatives of all orders. The partial sums of a power series are polynomials.

A power series may converge for some values of x and not for others. The next theorem tells us a great deal about this question.

Theorem 6.4.6. *Given a power series (6.4.2), let*

$$R = \frac{1}{\limsup |c_k|^{1/k}},$$

where we interpret R to be ∞ (resp. 0) if $\limsup |c_k|^{1/k}$ is 0 (resp.∞).

If $R > 0$, then the series (6.4.2) converges for each x with $|x - a| < R$ and diverges for each x with $|x - a| > R$. Furthermore, the series converges uniformly on every interval of the form $[a - r, a + r]$ with $0 < r < R$. If $R = 0$, then the series converges only when $x = a$.

Proof. We first suppose $R > 0$. Given any $r > 0$, we have

$$\limsup |c_k r^k|^{1/k} = r \limsup |c_k|^{1/k} = \frac{r}{R}. \tag{6.4.3}$$

Now suppose $|x - a| = r > R$. Then $|c_k(x - a)^k| = |c_k|r^k$ and the series (6.4.2) diverges, by (6.4.3) and the root test.

On the other hand, if $r < R$ and $|x - a| \leq r$, then $|c_k(x - a)^k| \leq |c_k|r^k$. In this case $\sum\limits_{k=1}^{\infty} |c_k|r^k$ converges, by the root test and (6.4.3). Then the Weierstrass

M-test implies that the series (6.4.2) converges uniformly on the closed interval $[a - r, a + r] = \{x : |x - a| \le r\}$.

The uniform convergence of (6.4.2) on $[a - r, a + r]$ for every $r < R$ implies that the series converges on $(a - R, a + R)$, since for every x in this interval, there is an $r < R$ such that x is also in the interval $[a - r, a + r]$.

If $R = 0$ – that is, if $\limsup |c_k|^{1/k} = \infty$ – then the only value of x that will lead to $\limsup |c_k(x - a)^k|^{1/k} < 1$ is $x = a$. Thus, the power series converges only at $x = a$ in this case.

$\square$

The above theorem tells us that the convergence set for a power series (6.4.2) is an interval of radius $R = (\limsup |c_k|^{1/k})^{-1}$, centered at a. The number R is called the *radius of convergence* of the series. Since the theorem says nothing when $|x - a| = R$, it does not tell us whether this interval is open, closed, or half-open, half-closed, Each of these possibilities occurs.

Example 6.4.7. Give examples where the three possibilities mentioned in the previous paragraph occur.

Solution The examples are

$$\text{(a)} \quad \sum_{k=0}^{\infty} x^k \quad \text{(b)} \quad \sum_{k=0}^{\infty} \frac{x^k}{k} \quad \text{(c)} \quad \sum_{k=0}^{\infty} \frac{x^k}{k^2}.$$

In each case, the radius of convergence R is 1, since

$$1 = \lim k^{1/k} = (\lim k^{1/k})^2 = \lim (k^2)^{1/k}.$$

When $x = \pm 1$, series (a) diverges by the term test, since its terms are all ± 1; thus, its interval of convergence is $(-1, 1)$.

Series (b) is the harmonic series when $x = 1$ and the alternating harmonic series when $x = -1$; thus, its interval of convergence is $[-1, 1)$.

Series (c) is the p-series with $p = 2$ at $x = 1$ and the alternating p-series with $p = 2$ when $x = -1$. Both series are convergent and so the interval of convergence for (c) is $[-1, 1]$.

Remark 6.4.8. Although the expression for the radius of convergence R, given in the previous theorem, is useful because it makes sense regardless of the series, it is often the case that the ratio test provides a more practical method for calculating the radius of convergence of a power series.

Example 6.4.9. Find the radius of convergence of the power series $\displaystyle\sum_{k=1}^{\infty} \frac{x^k}{k!}$.

Solution: We apply the ratio test. We have

$$\lim \left| \frac{x^{k+1}}{(k+1)!} \right| \div \left| \frac{x^k}{k!} \right| = \lim \frac{|x|}{k + 1} = 0$$

for all x. Thus, the series converges for all x and its radius of convergence is $+\infty$.

Integration of Power Series

Since a power series centered at a, with radius of convergence R, converges uniformly on each interval of the form $[a-r, a+r]$ with $0 < r < R$, our earlier theorems concerning continuity (Theorem 6.4.2) and term by term integration (Theorem 6.4.3) apply. They lead to the following theorem.

Theorem 6.4.10. *If* $f(x) = \displaystyle\sum_{k=0}^{\infty} c_k(x-a)^k$ *on* $(a-R, a+R)$, *where* R *is the radius of convergence of this series, then* f *is continuous on* $(a-R, a+R)$ *and*

$$\int_a^x f(t)\, dt = \sum_{k=0}^{\infty} \frac{c_k}{k+1}(x-a)^{k+1}, \qquad\qquad (6.4.4)$$

if $x \in (a-R, a+R)$. *The latter series also has radius of convergence* R.

Proof. The continuity of f is a direct consequence of Theorem 6.4.2, while the integral formula follows directly from Theorem 6.4.3 and the fact that

$$\int_a^x (t-a)^k\, dt = \frac{(x-a)^{k+1}}{k+1}$$

The statement about radius of convergence is proved as follows: If we factor $(x-a)$ out of the series in (6.4.4), the remaining factor is

$$\sum_{k=0}^{\infty} \frac{c_k}{k+1}(x-a)^k,$$

which clearly has the same convergence set and radius of convergence. By Theorem 6.4.6, its radius of convergence is the inverse of

$$\limsup \left(\frac{|c_k|}{k+1} \right)^{1/k} = \limsup |c_k|^{1/k} \lim \frac{1}{(k+1)^{1/k}} = \limsup |c_k|^{1/k},$$

which is the radius of convergence of the original series. Here, the first equality follows from Exercise 2.6.8, while the second equality follows from the fact that $\lim(1+k)^{1/k} = 1$ (a simple consequence of l'Hôpital's Rule). Thus, the series in (6.4.4) has the same radius of convergence as the original series. $\qquad\square$

Example 6.4.11. Find a power series in x which converges to $\ln(1+x)$ in an open interval centered at 0. What is the largest such open interval?

Solution: If $|x| < 1$, the geometric series $\displaystyle\sum_{k=0}^{\infty} x^k$ converges to $\dfrac{1}{1-x}$. If we replace x by $-t$ in this series, the result is

$$\frac{1}{1+t} = \sum_{k=0}^{\infty} (-t)^k \quad \text{for} \quad |t| < 1.$$

If we integrate with respect to t from 0 to x, then it follows from the previous theorem that

$$\ln(1 + x) = \sum_{k=0}^{\infty} (-1)^k \frac{x^{k+1}}{k+1} = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{x^k}{k}.$$

for $|x| < 1$. The radius of convergence of this series is $(\limsup(1/k)^{1/k})^{-1} = 1$ and so $(-1, 1)$ is the largest open interval on which this series converges to $\ln(1 + x)$.

Differentiation of Power Series

We may also differentiate power series term by term.

Theorem 6.4.12. *If* $f(x) = \sum_{k=0}^{\infty} c_k (x - a)^k$ *on* $(a - R, a + R)$, *where* R *is the radius of convergence of this series, then* f *is differentiable on* $(a - R, a + R)$ *and, on this interval,*

$$f'(x) = \sum_{k=1}^{\infty} k c_k (x - a)^{k-1}. \tag{6.4.5}$$

This series also has radius of convergence R.

Proof. We set

$$g(x) = \sum_{k=1}^{\infty} k c_k (x - a)^{k-1}.$$

This series has the same radius of convergence as the series

$$\sum_{k=1}^{\infty} k c_k (x - a)^k = (x - a) \sum_{k=1}^{\infty} k c_k (x - a)^{k-1},$$

and that is

$$(\limsup |k c_k|^{1/k})^{-1} = (\lim k^{1/k} \limsup |c_k|^{1/k})^{-1} = R,$$

since $\lim k^{1/k} = 1$.

To complete the proof, we just need to show that g is the derivative of f. However, by the previous theorem,

$$\int_a^x g(t)\,dt = \sum_{k=1}^{\infty} c_k (x - a)^k = f(x) - c_0.$$

By the Second Fundamental Theorem, $f'(x) = g(x)$. $\qquad\square$

Example 6.4.13. Find a power series in x which converges to $\dfrac{1}{(1-x)^2}$ on an open interval centered at 0. What is the largest open interval on which this power series expansion is valid?

  Solution: As in the last example, we begin with the power series expansion of $\dfrac{1}{1-x}$ on $(-1,1)$,

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k.$$

If we differentiate, using the previous theorem, the result is

$$\frac{1}{(1-x)^2} = \sum_{k=1}^{\infty} kx^{k-1} = \sum_{k=0}^{\infty} (k+1)x^k.$$

on $(-1,1)$. By the theorem, this series has radius of convergence 1. Thus, $(-1,1)$ is the largest open interval on which this expansion is valid.

Exercise Set 6.4

1. Prove that the function $f(x) = \sum_{k=1}^{\infty} \dfrac{x^k}{k^2}$ is continuous on the interval $[-1,1]$.

2. Prove that the function $f(x) = \sum_{k=1}^{\infty} \dfrac{\sin kx}{2^k}$ is continuous on the entire real line.

3. Let $\{f_k\}$ be a sequence of differentiable functions on (a,b) and suppose there is a point $c \in (a,b)$ such that the series $\sum_{k=1}^{\infty} f_k(c)$ converges. Suppose also that the sequence of derivatives $\{f_k'\}$ satisfies $|f_k'(x)| \le M_k$ on (a,b) and the series $\sum_{k=1}^{\infty} M_k$ converges. Then prove that the series defining

$$f(x) = \sum_{k=1}^{\infty} f_k(x) \quad \text{and} \quad g(x) = \sum_{k=1}^{\infty} f_k'(x)$$

converge on (a,b) and f is differentiable with derivative g on (a,b).

   In each of the next five exercises, find the radius of convergence of the indicated power series.

4. $\sum_{k=1}^{\infty} \dfrac{1}{k3^k} x^k.$

5. $\displaystyle\sum_{k=0}^{\infty} \frac{(-1)^{k-1}}{k+1}(x+2)^k.$

6. $\displaystyle\sum_{k=1}^{\infty} \frac{1}{k^{\sqrt{k}}}x^k.$

7. $\displaystyle\sum_{k=0}^{\infty} k!(x-5)^k.$

8. $\displaystyle\sum_{k=0}^{\infty} 2^k x^{2k}.$

9. Beginning with the geometric series which converges to $\dfrac{1}{1-x}$ on $(-1,1)$, find power series which converge to $\dfrac{1}{1+x^2}$ and to $\arctan x$ on this same interval.

10. Prove that if $f(x)$ is the sum of a power series centered at a and with radius of convergence R, then f is infinitely differentiable on $(a - R, a + R)$ – that is, its derivative of order m exists on this interval for all $m \in \mathbb{N}$.

11. Suppose functions g and h are defined by

$$g(x) = x + \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + \frac{x^{2n+1}}{(2n+1)!} + \cdots$$

$$h(x) = 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + \frac{x^{2n}}{(2n)!} + \cdots .$$

Find the interval of convergence for each of these functions.

12. Prove that the functions in the previous exercise satisfy $g' = h$ and $h' = g$.

13. Prove Theorem 6.4.3.

6.5   Taylor's Formula

Definition 6.5.1. Suppose f is a function defined in an open interval containing a. If there is a power series, centered at a, which converges to f in some open interval centered at a, then we will say that f is *analytic* at a. If f is analytic at every point of an open interval I, then we will say that f is analytic on I.

When can we expect that f is analytic at a? According to Exercise 6.4.10, if f is the sum of a power series in some interval centered at a, then f is infinitely differentiable in this interval (meaning its nth derivative exists for every $n \in \mathbb{N}$). Thus, in order for a function to be analytic at a it must be infinitely differentiable in some interval centered at a. However, this is not enough. In fact Exercise 6.5.13 shows that there is a function which is infinitely differentiable in an open interval centered at 0, but is not the sum of a power series centered at 0.

Power Series Coefficients

If a function is analytic at a – that is, it has a power series expansion centered at a, then it is easy to see what the coefficients of the power series expansion must be.

Theorem 6.5.2. *Suppose* $f(x) = \sum_{k=0}^{\infty} c_k(x-a)^k$, *where this series converges to* $f(x)$ *on an open interval containing a. Then* $c_n = \dfrac{f^{(n)}(a)}{n!}$ *for each n.*

Proof. We prove by induction that the nth derivative of f is

$$f^{(n)}(x) = \sum_{k=n}^{\infty} \frac{k!}{(k-n)!} c_k(x-a)^{k-n}. \tag{6.5.1}$$

When $n = 1$, this just says that

$$f'(x) = \sum_{k=1}^{\infty} k c_k(x-a)^{k-1},$$

which is true by Theorem 6.4.12.

If we assume that (6.5.1) is true for a given n, then by differentiating it and again using Theorem 6.4.12, we obtain

$$f^{(n+1)}(x) = \sum_{k=n}^{\infty} \frac{k!}{(k-n)!}(k-n)c_k(x-a)^{k-n-1}$$

$$= \sum_{k=n+1}^{\infty} \frac{k!}{(k-n-1)!} c_k(x-a)^{k-n-1}.$$

Since this is (6.5.1) with n replaced by $n+1$, the induction is complete and we conclude that (6.5.1) is true for all $n \in \mathbb{N}$.

If we set $x = a$ in (6.5.1), all terms vanish except for the first one (the one where $k = n$). Thus,

$$f^{(n)}(a) = n! c_n \quad \text{or} \quad c_n = \frac{f^{(n)}(a)}{n!}.$$

$\square$

Taylor's Formula

The previous theorem tells us that the only power series, centered at a, that could possibly converge to $f(x)$ in an interval centered at a is the power series

$$\sum_{k=0}^{\infty} \frac{f^k(a)}{k!}(x-a)^k. \tag{6.5.2}$$

This is called the *Taylor Series* for f at a. The nth partial sum of this series,

$$f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n$$

is called the nth *Taylor polynomial* for f at a. The function f is analytic at a if and only if the sequence of Taylor polynomials for f converges to f in some open interval centered at a. Taylor's Formula helps decide whether this is true by providing a formula for the remainder when f is approximated by its nth Taylor polynomial.

Theorem 6.5.3. (Taylor's Formula) *Let f be a function which has continuous derivatives up through order $n+1$ in an open interval I centered at a. Then, for each $x \in I$,*

$$f(x) = f(a) + f'(a)(x-a) + \cdots + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x), \qquad (6.5.3)$$

where

$$R_n(x) = \frac{f^{(n+1)}(c)}{(n+1)!}(x-a)^{(n+1)}, \qquad (6.5.4)$$

for some c between a and x.

Proof. This theorem is reminiscent of the Mean Value Theorem. In fact, in the case $n = 0$, it is the Mean Value Theorem. It is not surprising that its proof mimics the proof of the Mean Value Theorem.

We set

$$R_n(x) = f(x) - f(a) - f'(a)(x-a) - \cdots - \frac{f^{(n)}(a)}{n!}(x-a)^n,$$

so that (6.5.3) holds. We then define a function $s(t)$ on I by

$$s(t) = f(x) - f(t) - f'(t)(x-t) - \cdots - \frac{f^{(n)}(t)}{n!}(x-t)^n - R_n(x)\left(\frac{x-t}{x-a}\right)^{n+1}.$$

Then $s(a) = s(x) = 0$, and so there must be a critical point c for s somewhere strictly between a and x. Since s is differentiable on I, this critical point must be a point where s' is 0 – that is, $s'(c) = 0$. In the calculation of s', all the terms cancel except two at the very end, leaving

$$0 = s'(c) = -\frac{f^{(n+1)}(c)}{n!}(x-c)^n + (n+1)R_n(x)\frac{(x-c)^n}{(x-a)^{n+1}}.$$

Equation (6.5.4) follows from this when we solve for $R_n(x)$. $\qquad\qquad \square$

Example 6.5.4. Find the Taylor series expansion of e^x at 0 and tell for which values of x this expansion converges to e^x.

Solution: The function e^x is infinitely differentiable on $\mathbb{R}$ with kth derivative equal to e^x for all x. Thus, its kth derivative evaluated at 0 is 1 for all k. Taylor's Formula then tells us that

$$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + R_n(x),$$

where

$$R_n(x) = e^c \frac{x^{n+1}}{(n+1)!},$$

for some c between 0 and x.

For all values of x and c, $\lim e^c \dfrac{x^{n+1}}{(n+1)!} = 0$ (Exercise 6.5.1). This implies that the Taylor polynomials for e^x converge to e^x for all $x \in \mathbb{R}$ – that is, the Taylor series expansion

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots \qquad (6.5.5)$$

is valid for all $x \in \mathbb{R}$.

Example 6.5.5. Find the Taylor series expansion of $\sin x$ at 0 and tell for which values of x this expansion converges to $\sin x$.

Solution: The function $f(x) = \sin x$ is infinitely differentiable on $\mathbb{R}$ and its first 4 derivatives are

$$f'(x) = \cos x, \quad f''(x) = -\sin x, \quad f'''(x) = -\cos x, \quad f^{(4)}(x) = \sin x.$$

Since $f^{(4)} = f$, taking nth derivatives leads to $f^{(n+4)} = f^{(n)}$ for every non-negative integer n. Thus, at 0, the sin and its derivatives form the following repeating sequence with period 4:

$$0, 1, 0, -1, 0, 1, 0, -1, 0, \cdots .$$

Hence, Taylor's formula for $\sin x$ at $a = 0$ is

$$x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + R_{2n+2}(x),$$

where

$$R_{2n+2}(x) = \sin^{(2n+3)}(c) \frac{x^{2n+3}}{(2n+3)!} \quad \text{for some} \quad c.$$

The reason we use $R_{2n+2}(x)$ rather than $R_{2n+1}(x)$ for the remainder (they are actually equal, since the term of degree $2n+2$ is 0 in Taylor's Formula for $\sin x$) is that we get better estimates on the size of the remainder if we use $R_{2n+2}(x)$.

Since $|\sin^{(2n+3)}(c)| \leq 1$, we have

$$|R_{2n+2}(x)| \leq \frac{|x|^{2n+3}}{(2n+3)!}$$

which implies that the remainder has limit 0 for all x (see Exercise 6.5.1). Thus, the Taylor series expansion

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + \cdots$$

is valid for all $x \in \mathbb{R}$.

Example 6.5.6. Find an estimate for the error if $\sin x$ is approximated by $x - x^3/3!$ for x in the interval $[-\pi/4, \pi/4]$. By an estimate for the error, we mean an upper bound for the error which is as close to the actual error as possible without going to extraordinary effort.

  Solution: By the previous example, the difference between $\sin x$ and its third degree Taylor polynomial has absolute value less than or equal to

$$\frac{|x|^5}{5!} \le \frac{(\pi/4)^5}{5!} < .003 \quad \text{for} \quad -\pi/4 \le x \le \pi/4.$$

Lagrange's Form for the Remainder

The following integral formula for the remainder in Taylor's Formula sometimes leads to better estimates on the size of the remainder than does the usual form.

Theorem 6.5.7. *If f is a function with continuous derivatives up through order $n + 1$ on an open interval I containing a and x, then the remainder $R_n(x)$ in Taylors formula for f at a can be written as*

$$R_n(x) = \frac{1}{n!} \int_a^x (x - t)^n f^{(n+1)}(t)\, dt \qquad (6.5.6)$$

Proof. We prove (6.5.6) by induction on n with the base case being $n = 0$. In the case where $n = 0$, Taylor's formula is

$$f(x) = f(a) + R_0(x) \quad \text{so that} \quad R_0(x) = f(x) - f(a).$$

Equation (6.5.6) in this case is

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

which is just the Fundamental Theorem of Calculus. Thus, (6.5.6) holds when $n = 0$.

  For the induction step, we assume (6.5.6) holds for a given n and proceed to prove that it then holds for $n + 1$. If we apply integration by parts to the integral on the right side of (6.5.6), the result is

$$R_n(x) = \frac{f^{(n+1)}(a)}{(n+1)!}(x - a)^{n+1} + \frac{1}{(n+1)!} \int_a^x (x - t)^{n+1} f^{(n+2)}(t)\, dt.$$

Since, $R_{n+1}(x) = R_n(x) - \dfrac{f^{(n+1)}(a)}{(n+1)!}(x - a)^{n+1}$, this proves (6.5.6) holds with n replaced by $n + 1$, thus completing the induction step. $\qquad\square$

Example 6.5.8. Find a power series expansion for $f(x) = (1 + x)^p$ which is valid on $(-1, 1)$, where p is any real number.

   Solution: The derivatives of f are

$$p(1+x)^{p-1}, \ p(p-1)(1+x)^{p-2}, \ \cdots, \ p(p-1)\cdots(p-n+1)(1+x)^{p-n}\cdots .$$

The nth derivative evaluated at 0 is $p(p-1)\cdots(p-n+1)$. Thus, Taylor's formula for f is

$$(1+x)^p = 1 + px + \frac{p(p-1)}{2}x^2 + \cdots + \frac{p(p-1)\cdots(p-n+1)}{n!}x^n + R_n(x),$$

where
$$R_n(x) = \frac{p(p-1)\cdots(p-n)}{n!} \int_0^x \frac{(x-t)^n}{(1+t)^{n+1-p}}\, dt,$$

if we use Lagrange's form of the remainder. However, since t is between 0 and x, t and x have the same sign, and this implies that

$$\left| \frac{x-t}{t+1} \right| \le |x|. \tag{6.5.7}$$

(Exercise 6.5.9). From this, we conclude that

$$|R_n(x)| \le \frac{p(p-1)\cdots(p-n)}{n!}|x|^n \int_0^x (1+t)^{p-1}\, dt.$$

This is just the constant $\displaystyle\int_0^x (1+t)^{p-1}\, dt$ times the absolute value of the nth term in the power series

$$1 + px + \frac{p(p-1)}{2}x^2 + \cdots + \frac{p(p-1)\cdots(p-n+1)}{n!}x^n + \cdots , \tag{6.5.8}$$

which happens to be the Taylor series for $(1+x)^p$ at 0. If we can show that this series converges when $|x| < 1$, then the Term Test implies its sequence of terms converges to 0 and, by the above, this shows that the remainder $R_n(x)$ converges to 0 and, hence, that this series converges to $(1+x)^p$ when $|x| < 1$.

   We prove that (6.5.8) converges on $(-1, 1)$ by using the Ratio Test. For the absolute value of the ratio of term $n + 1$ to term n, we get

$$\frac{|p-n|}{n+1}|x|$$

which has limit $|x|$ as $n \to \infty$. Hence, the series (6.5.8) converges for $|x| < 1$ and it converges to $(1+x)^p$.

   Note that when p is a positive integer, the series (6.5.8) terminates at $n = p$, that is, all terms with $n > p$ are zero and Taylor's Formula for $(1 + x)^p$ at 0,

with $n \geq p$ is just

$$(1+x)^p = 1 + px + \frac{p(p-1)}{2}x^2 + \cdots + \frac{p(p-1)\cdots(p-p+1)}{p!}x^p$$

$$= \sum_{k=0}^{p} \frac{p!}{k!(p-k)!}x^k,$$

(6.5.9)

which is the Binomial Theorem (Theorem 1.2.12) with $a = 1$ and $b = x$. The Binomial Theorem for general a and b can be deduced from this (Exercise 6.5.14).

Exercise Set 6.5

1. Prove that $\lim \dfrac{x^n}{n!} = 0$ for all x.

2. Find the Taylor Series expansion for $\cos x$ at 0 and show that it converges for all x.

3. Use Taylor's Formula to estimate the error if $\cos x$ is approximated by $1 - \dfrac{x^2}{2}$ on the interval $[-.1, .1]$.

4. What is the smallest n for which we can be sure that

$$1 + 1 + \frac{1}{2} + \frac{1}{3!} + \cdots + \frac{1}{n!}$$

   is within .001 of e?

5. What is Taylor's Formula for the function $f(x) = \sqrt{1+x}$ with $a = 0$?

6. What is Taylor's Formula for the function $f(x) = x^3 - x^2 - 4x + 4$ with $a = 1$?

7. What is Taylor' Formula for $\ln(1+x)$ with $a = 0$. Compare with Example 6.4.11.

8. Use the binomial series with $p = -1/2$ to get a power series expansion for $\dfrac{1}{\sqrt{1-x}}$ valid on $(-1, 1)$. Use this to get power series expansions for first $\dfrac{1}{\sqrt{1-x^2}}$, and then $\arcsin x$ on this same interval.

9. Prove that if $x \in (-1, 1)$ and t is between 0 and x (so that t and x have the same sign and $|t| \leq |x| < 1$), then

$$\left| \frac{x-t}{t+1} \right| \leq |x|.$$

10. Verify the computation of s' given in the proof of Theorem 6.5.3.

11. Prove that if f is an infinitely differentiable function on $(a - r, a + r)$ and there is a constant K such that

$$|f^{(n)}(x)| \le K\frac{n!}{r^n}$$

for all $n \in N$ and all $x \in (a - r, a + r)$, then the Taylor Series for f at a converges to f on $(a - r, a + r)$.

12. Use l'Hôpital's Rule to show that $\displaystyle\lim_{x \to 0} \frac{e^{-1/x^2}}{x^n} = 0$ for every n.

13. If $g(x) = e^{-1/x^2}$ for $x \ne 0$ and $g(0) = 0$, show that g is infinitely differentiable on the entire real line, but all of its derivatives at 0 are equal to 0. Argue that this means that g cannot be analytic at 0. Hint: use the previous exercise to help compute the derivatives of g at 0.

14. Prove that the Binomial Formula (Theorem 1.2.12) for a general a and b follows from the Taylor Series expansion (6.5.9) of $(1 + x)^p$ for p a positive integer.

15. Give a new proof that $e^x e^y = e^{x+y}$ by using the Taylor series expansion for e^x (6.5.5) and the product formula of Theorem 6.3.6. You will also need to use the binomial formula.

Chapter 7

Convergence in Euclidean Space

With this chapter we begin our study of calculus in several variables. The first task is to define $\mathbb{R}^d$ – Euclidean space of dimension d. We will then study convergence of sequences of points in this space and introduce the concepts of open and closed sets. These are generalizations to $\mathbb{R}^d$ of the concepts of open and closed intervals in $\mathbb{R}$. In the final two sections we introduce the concepts of compact sets and connected sets. These are also generalizations to $\mathbb{R}^d$ of properties of intervals in $\mathbb{R}$. These ideas will be of fundamental importance when we study continuous functions on $\mathbb{R}^d$ in the next chapter.

In order to define and study convergence and continuity, we don't need to use all of the properties of $\mathbb{R}^d$ – only the ones derived from the concept of distance between points. A set together with a well behaved notion of distance between pairs of points is called a *metric space*. In the coming pages, we will give a more precise definition of metric space and point out how many of the definitions and theorems we develop in this chapter are valid, not only in $\mathbb{R}^d$, but in any metric space.

7.1 Euclidean Space

The space $\mathbb{R}^d$ is defined to be the set of all d-tuples of real numbers, where, by a d-tuple of real numbers, we mean an ordered set $(x_1, x_2, \cdots, x_d)$ of d real numbers. It is ordered because the numbers are listed in a certain order and, if this order is changed, then the new d-tuple is different from the old one (unless the change of order just interchanges identical numbers). For example, $(5, 0, 7)$ and $(0, 5, 7)$ are different 3-tuples and, hence, different points of $\mathbb{R}^3$.

The spaces $\mathbb{R}^2$ and $\mathbb{R}^3$ are familiar from calculus. The space R^2 is the set of all ordered pairs (x_1, x_2) of real numbers, while $\mathbb{R}^3$ is the set of ordered triples (x_1, x_2, x_3) of real numbers. Often points of $\mathbb{R}^2$ are denoted (x, y) rather than (x_1, x_2) and points of $\mathbb{R}^3$ are denoted (x, y, z) rather than (x_1, x_2, x_3).

The Vector Space $\mathbb{R}^d$

We will often refer to a point of $\mathbb{R}^d$ as a *vector* in $\mathbb{R}^d$, while a point of $\mathbb{R}$ will often be refered to as a *scalar*.

There are natural operations of addition of vectors in $\mathbb{R}^d$ and multiplication of vectors by scalars. That is, if $x = (x_1, x_2, \cdots, x_d)$ and $y = (y_1, y_2, \cdots, y_d)$ are vectors in $\mathbb{R}^d$, and a is a scalar, then we set

$$x + y = (x_1 + y_1, x_2 + y_2, \cdots, x_d + y_d)$$

and

$$ax = (ax_1, ax_2, \cdots, ax_d).$$

The zero vector (also called the *origin* of $\mathbb{R}^d$) is the vector

$$0 = (0, 0, \cdots, 0).$$

Note that we use the same symbol, 0, to stand for both the scalar 0 and the vector $0 \in \mathbb{R}^d$. This shouldn't cause any confusion, since it will always be obvious from the context which is meant.

Given a vector $x = (x_1, x_2, \cdots, x_d)$ in $\mathbb{R}^d$, the *components* of x are the numbers $x_1, x_2, \cdots, x_d$. The jth component is the number x_j. Two vectors are identical if and only if their jth components are identical for $j = 1, 2, \cdots, d$.

As noted in the next theorem, addition in $\mathbb{R}^d$ satisfies the associative and commutative laws and 0 has the appropriate properties. Also, scalar multiplication satisfies an associative law and two distributive laws.

Theorem 7.1.1. *Let u, v, w be points of $\mathbb{R}^d$ and a and b real numbers. Then*

(a) $u + (v + w) = (u + v) + w$;

(b) $u + v = v + u$;

(c) $0 + u = u$;

(d) $0u = 0$ and $1u = u$;

(e) $a(bu) = (ab)u$;

(f) $(a + b)u = au + bu$;

(g) $a(u + v) = au + av$.

Proof. Each statement asserts that two vectors are identical. Thus, each can be proved by proving that the jth components of the two vectors are identical for each j. In each case, this follows immediately from the definitions and the fact that $\mathbb{R}$ satisfies the field axioms **A1** - **A4**, **M1** - **M4**, and **D** (see Section 1.3). □

A set together with operations of addition and scalar multiplication (where the scalars belong to some field F), satisfying the properties listed in the above theorem, is called a *vector space* over F (see Section 1.3 for the definition of a field). Hence, $\mathbb{R}^d$ is a vector space over the field $\mathbb{R}$.

Using only the vector space axioms listed in Theorem 7.1.1, one can easily derive all of the properties of general vector spaces.

Example 7.1.2. Using only the properties listed in Theorem 7.1.1, prove that if x is an element of a vector space, then $(-1)x$ is an additive inverse for x. That is, prove that $x + (-1)x = 0$.

Solution: By Theorem 7.1.1 (d) and (f) we have

$$x + (-1)x = (1 + (-1))x = 0x = 0.$$

In view of this example, $(-1)x$ is an additive inverse for x and so it makes sense to denote it simply by $-x$.

Other properties of vector spaces will be derived in the exercises.

Inner Product

Definition 7.1.3. The Euclidean inner product of two vectors $u = (u_1, \cdots, u_d)$ and $v = (v_1, \cdots, v_d)$ in $\mathbb{R}^d$ is the real number

$$u \cdot v = u_1 v_1 + u_2 v_2 + \cdots + u_d v_d. \tag{7.1.1}$$

This has the following simple properties. The proof is left to the exercises.

Theorem 7.1.4. *If $u, v, w \in \mathbb{R}^d$ and $a \in \mathbb{R}$, then*

(a) $u \cdot v = v \cdot u$;

(b) $(u + v) \cdot w = u \cdot w + v \cdot w$;

(c) $(au) \cdot v = a(u \cdot v)$;

(d) $u \cdot u > 0$ unless $u = 0$ in which case $u \cdot u = 0$.

More generally, a function from pairs of vectors to scalars which satisfies (a) through (d) above is called an *inner product* on the vector space. A vector space together with an inner product on that vector space is called an *inner product space*. Thus, $\mathbb{R}^d$ is an inner product space with the inner product described in Definition 7.1.3.

There are other inner products on $\mathbb{R}^d$. For example, if each term $u_j v_j$ in (7.1.1) is replaced by $a_j u_j v_j$, where $a_1, \cdots, a_d$ are positive scalars, then the resulting sum defines a new inner product which is different from the original unless all the a_j's are 1. In this text, the only inner product on $\mathbb{R}^d$ that we will use is the Euclidean inner product as define in (7.1.1).

Using (a) and (c) of Theorem 7.1.4, we easily show that $u \cdot (av) = a(u \cdot v)$. Thus, for a scalar a and vectors u and v, there is no ambiguity if we simply write $au \cdot v$ in place of any one of the three equal products

$$a(u \cdot v), \quad (au) \cdot v, \quad u \cdot (av).$$

Example 7.1.5. If X is an inner product space, $x, y \in X$ and $a, b \in \mathbb{R}$, then calculate the inner product of $ax + by$ with itself.

Solution: By (b) and (c) of the previous theorem, we have

$$(ax + by) \cdot (ax + by) = ax \cdot (ax + by) + by \cdot (ax + by).$$

By (a), (b), and (c) we have

$$ax \cdot (ax + by) = a^2 x \cdot x + abx \cdot y,$$
$$by \cdot (ax + by) = abx \cdot y + b^2 y \cdot y.$$

Combining these yields

$$(ax + by) \cdot (ax + by) = a^2 x \cdot x + 2abx \cdot y + b^2 y \cdot y.$$

Components of a Vector

We will typically denote by e_j the vector consisting of the d-tuple with all entries 0 except for the jth entry which is 1. Thus, $e_j = (0, 0, \cdots, 0, 1, 0, \cdots, 0)$ with the 1 occurring in the jth position. Note that

$$e_j \cdot e_k = \delta_{jk},$$

where δ_{jk} is 1 if $j = k$ and is 0 otherwise. This means that $\{e_j\}_{j=1}^n$ is an *orthonormal set* in $\mathbb{R}^d$.

Note that if $x = (x_1, x_2, \cdots, x_d) \in \mathbb{R}^d$, then the jth component x_j of x is given by $x_j = x \cdot e_j$ for $j = 1, \cdots, d$.

Example 7.1.6. Show that each vector in $\mathbb{R}^d$ is a unique linear combination of the vectors e_j for $j = 1, \cdots, d$.

Solution: If $x = (x_1, x_2, \cdots, x_d)$, then

$$x = \sum_{j=1}^d x_j \, e_j = \sum_{j=1}^d (x \cdot e_j) e_j.$$

This is one way of expressing x as a linear combination of the e_j's. On the other hand, if

$$x = \sum_{j=1}^d a_j e_j$$

is any such linear combination, then for $k = 1, \cdots, d$,

$$x_k = x \cdot e_k = \sum_{j=1}^d a_j e_j \cdot e_k = a_k,$$

since $e_j \cdot e_k = 1$ if $j = k$ and is 0 other wise. Thus the coefficients a_j must be the numbers x_j.

Norm and Distance

Definition 7.1.7. In an inner product space, we define the norm $||x||$ of a vector x to be the number
$$||x|| = \sqrt{x \cdot x}.$$
The distance between two vectors x and y is defined to be $||x - y||$.

Note that, by Theorem 7.1.4 (d), the norm of a vector is always non-negative and is zero only if the vector is the zero vector . Thus, the distance between two vectors is always non-negative and is zero if and only if the vectors are equal.

In calculus, it is often shown that for two vectors u and v in $\mathbb{R}^2$ or $\mathbb{R}^3$ the inner product satisfies
$$u \cdot v = ||u||\,||v|| \cos \theta,$$
where θ is the angle between u and v. Since $|\cos \theta| \le 1$, this implies that
$$|u \cdot v| \le ||u||\,||v||.$$
As we show below, this inequality is true in $\mathbb{R}^d$ and, in fact, in any inner product space. In this generality it is known as the Cauchy-Schwarz inequality.

Theorem 7.1.8. (Cauchy-Schwarz Inequality) *If X is an inner product space, then*
$$|u \cdot v| \le ||u||\,||v||$$
for all $u, v \in X$.

Proof. If we take the inner product of a vector with itself, the result is non-negative by (d) of Theorem 7.1.4. Thus, if u and v are vectors in X and $t \in \mathbb{R}$ is a scalar, then
$$0 \le (tu + v) \cdot (tu + v) = t^2 u \cdot u + 2tu \cdot v + v \cdot v = at^2 + 2bt + c,$$
where $a = u \cdot u = ||u||^2$, $b = u \cdot v$, and $c = v \cdot v = ||v||^2$. The expression on the right is a quadratic function of t which is never negative. This means that the quadratic equation
$$at^2 + 2bt + c = 0$$
has at most one real root (since the graph of $at^2 + 2bt + c$ cannot cross the t-axis). By the quadratic formula, the roots of this equation are
$$-b \pm \sqrt{b^2 - ac}.$$
Since there cannot be two real roots, it must be the case that $b^2 \le ac$. On taking the square root of both sides of this inequality, we obtain the inequality of the theorem. □

Let u and v be vectors in an inner product space. In view of the above theorem, the number $\dfrac{u \cdot v}{||u||\,||v||}$ is always between -1 and 1 and, hence, is the cosine of some angle θ with $0 \le \theta \le \pi$. This leads to the following extension to arbitrary inner product spaces of the notion of the angle between two vectors.

Definition 7.1.9. With u, v and θ as above, we will call θ the angle between u and v. This angle is $\pi/2$ if and only if $u \cdot v = 0$. In this case we will say that u and v are mutually orthogonal and write $u \perp v$.

The Triangle Inequality

The triangle inequality is just the vector space version of the statement that the length of one side of a triangle is always less than or equal to the sum of the lengths of the other two sides. It is stated more precisely in part (a) of the following theorem.

Theorem 7.1.10. *If X is an inner product space, $x, y \in X$, and $a \in \mathbb{R}$, then*

(a) $||x + y|| \leq ||x|| + ||y||$;

(b) $||ax|| = |a|\, ||x||$;

(c) $||x|| = 0$ *implies* $x = 0$.

Proof. Using Example 7.1.5 and the Cauchy-Schwarz inequaltiy, we have

$$||x + y||^2 = (x + y) \cdot (x + y) = ||x||^2 + 2x \cdot y + ||y||^2$$
$$\leq ||x||^2 + 2||x||\,||y|| + ||y||^2 = (||x|| + ||y||)^2.$$

Part (a) of the theorem follows from this on taking square roots. Parts (b) and (c) follow from (c) and (d) of Theorem 7.1.4. □

Suppose u, v, and w are points in a vector space X. Then $||u - v||, ||v - w||$, and $||u - w||$ are the lengths of the sides of the triangle with vertices at u, v, and w. If we apply part (a) of the previous theorem to the vectors $x = u - v$ and $y = v - w$, the result is the inequality

$$||u - w|| \leq ||u - v|| + ||v - w||, \tag{7.1.2}$$

which says that a side of a triangle always has length less than or equal to the sum of the lengths of the other two sides.

Norms in General

The norm induced by an inner product is just one type of norm on a vector space. In general, a *norm* on a vector space X is a non-negative function $||\cdot||$ which satisfies (a), (b), and (c) of the previous theorem. A *normed vector space* is a vector space X together with a norm on X. There are norms on $\mathbb{R}^d$ which are different from the Euclidean norm (the norm induced by the Euclidean inner product).

Definition 7.1.11. If $x = (x_1, x_2, \cdots, x_d) \in \mathbb{R}^d$, we set

1. $||x||_1 = |x_1| + |x_2| + \cdots + |x_d|$;

2. $||x||_\infty = \max\{|x_1|, |x_2|, \cdots, |x_d|\}.$

Example 7.1.12. Show that $|| \cdot ||_1$ is a norm on $\mathbb{R}^d$.
    Solution: If $x = (x_1, x_2, \cdots, x_d)$ and $y = (y_1, y_2, \cdots, y_d)$, then

$$||x + y||_1 = \sum_{j=1}^{n} |x_j + y_j| \le \sum_{j=1}^{n} (|x_j| + |y_j|),$$

by the triangle inequality for $\mathbb{R}$. The sum on the right is equal to

$$\sum_{j=1}^{d} |x_j| + \sum_{j=1}^{d} |y_j| = ||x||_1 + ||y||_1.$$

Thus, $|| \cdot ||_1$ satisfies the triangle inequality ((a) of Theorem 7.1.10).
    If $a \in \mathbb{R}$, then

$$||ax||_1 = \sum_{j=1}^{d} |ax_j| = \sum_{j=1}^{d} |a| \, |x_j| = |a| \, ||x||_1.$$

Thus, $|| \cdot ||_1$ also satisfies (b). That (c) holds as well is obvious, since $||x||_1 = 0$ implies that $x_j = 0$ for each j and, hence, that $x = 0$.

We leave to the exercises, the problem of showing that $|| \cdot ||_\infty$ is also a norm on $\mathbb{R}^d$.

Theorem 7.1.13. *The three norms we have defined on $\mathbb{R}^d$ are related as follows:*

$$d^{-1}||x||_1 \le ||x||_\infty \le ||x|| \le ||x||_1$$

for each $x \in \mathbb{R}^d$.

The proof of this is also left to the exercises.

The Normed Vector Space $C(I)$

In mathematics we deal with a great many normed vector spaces. One that does not look at all like $\mathbb{R}^d$ is the space $C(I)$, where I is a closed bounded interval on the real line, and $C(I)$ is the vector space of all continuous real valued functions on I. Addition is pointwise addition of functions and scalar multiplication is multiplication of a function by a constant. It is easy to see that $C(I)$ is a vector space under these two operations (Exercise 7.1.10). There are many norms that can be put on this vector space, but perhaps the most useful is the sup norm, $|| \cdot ||_\infty$, defined by

$$||f||_\infty = \sup_I |f(x)|, \tag{7.1.3}$$

for $f \in C(I)$. The problem of showing that this is a norm is left to the exercises.

Exercise set 7.1

1. For the vectors $x = (1, 0, 2)$ and $y = (-1, 3, 1)$ in $\mathbb{R}^3$ find

    (a) $2x + y$;

    (b) $x \cdot y$;

    (c) $||x||$ and $||y||$;

    (d) the cosine of the angle between x and y;

    (e) the distance from x to y.

2. Using only the properties listed in Theorem 7.1.1, prove that if u, v, w are vectors in a vector space and $u + w = v + w$, then $u = v$.

3. Using only the properties listed in Theorem 7.1.1, prove that if u is a vector in a vector space, a is a scalar, and $au = 0$, then either $a = 0$ or $u = 0$.

4. Prove Theorem 7.1.4.

5. Prove the second form of the triangle inequality. That is, prove that

$$|\,||x|| - ||y||\,| \leq ||x - y||$$

    holds for any pair of vectors x, y in a normed vector space. Hint: use the first form (Theorem 7.1.10(a)) to prove the second form.

6. Prove that equality holds in the Cauchy-Schwarz inequality (Theorem 7.1.8) if and only if one of the vectors u, v is a scalar multiple of the other.

7. For a norm on a vector space X, defined by an inner product as in Definition 7.1.7, prove that the parallelogram law:

$$||x + y||^2 + ||x - y||^2 = 2||x||^2 + 2||y||^2,$$

    holds for all $x, y \in X$.

8. Prove that $|| \cdot ||_\infty$, as defined in Definition 7.1.11, is a norm on $\mathbb{R}^d$.

9. Prove Theorem 7.1.13

10. Prove that the space $C(I)$, defined in the previous subsection, is a vector space.

11. Prove that the sup norm as defined in 7.1.3 is really a norm on $C(I)$.

12. Prove that if $\{x_k\}$ and $\{y_k\}$ are sequences of real numbers such that

$$\sum_{k=1}^{\infty} x_k^2 < \infty \quad \text{and} \quad \sum_{k=1}^{\infty} y_k^2 < \infty, \quad \text{then} \quad \sum_{k=1}^{\infty} |x_k y_k| < \infty.$$

    Hint: what can you say about the corresponding finite sums?

13. Find a non-zero vector in $\mathbb{R}^3$ which is orthogonal to both $(1, 0, 2)$ and $(3, -1, 1)$.

14. Prove that if u and v are vectors in an inner product space and $u \perp v$, then $||u + v||^2 = ||u||^2 + ||v||^2$.

7.2 Convergent Sequences of Vectors

In this section we study convergence of sequences of vectors in $\mathbb{R}^d$. The definitions and theorems in this topic are very similar to those of Chapter 2 on sequences of numbers.

Metric Spaces

As long as we are working in a space with a reasonable notion of distance between points, we can define and study convergent sequences and continuous functions. Such a space is called a *metric space*. The precise conditions for a space to be a metric space are defined below.

Definition 7.2.1. Let X be a set and δ a function which assigns to each pair (x, y) of elements of X a non-negative real number $\delta(x, y)$. Then δ is called a *metric* on X if, for all $x, y, z \in X$, the following conditions hold:

(a) $\delta(x, y) = \delta(y, x)$;

(b) $\delta(x, y) = 0$ if and only if $x = y$; and

(c) $\delta(x, z) \leq \delta(x, y) + \delta(y, z)$.

A set X, together with a metric δ on X is called a *metric space.*

Conditions (a) and (b) above are called the symmetry and identity conditions, while condition (c) is the triangle inequality for metric spaces.

We will show that $\mathbb{R}^d$ is a metric space, as is any normed vector space.

Theorem 7.2.2. *If X is a normed vector space, then X is a metric space if its metric δ is defined by*

$$\delta(x, y) = ||x - y||.$$

In particular, $\mathbb{R}^d$ is a metric space in the Euclidean norm, as is $C(I)$ in the sup norm.

Proof. Parts (a), (b), and (c) of Theorem 7.1.10 are satisfied by the norm in any normed vector space. Part (b) with $a = -1$ implies that $||x - y|| = ||y - x||$ and so δ is symmetric. Part (c) implies that $||x - y|| = 0$, if and only if $x = y$, and so δ satisfies the identity condition. Part (a) implies (7.1.2), which shows that δ satisfies the triangle inequality. Thus, δ is a metric on X. □

Remark 7.2.3. If X is a metric space with metric δ and Y is any subset of X, then Y is also a metric space with the same metric δ. Thus, any subset of $\mathbb{R}^d$ is also a metric space if it is given the usual Euclidean metric.

There are a great many metric spaces other than subsets of $\mathbb{R}^d$ that are important in mathematics. We will explore some of these in the exercises.

Remark 7.2.4. The following statements summarize the relationship between the types of spaces we have introduced so far:

1. $\mathbb{R}^d$ is an inner product space;

2. every inner product space is a normed vector space, with norm defined by $||x|| = \sqrt{x \cdot x}$;

3. every normed vector space is a metric space, with metric defined by $\delta(x, y) = ||x - y||$.

Sequences

The definition of convergence for a sequence $\{x_n\}$ in R^d should look familiar:

Definition 7.2.5. If $\{x_n\}$ is a sequence of vectors in $\mathbb{R}^d$ and $x \in R^d$, then we say $\{x_n\}$ *converges* to x if for every $\epsilon > 0$ there is an $N \in \mathbb{R}$ such that

$$||x - x_n|| < \epsilon \quad \text{whenever} \quad n \geq N.$$

In this case, we write $\lim_{n \to \infty} x_n = x$ or $\lim x_n = x$ or simply $x_n \to x$.

Note that we do not require the N that appears in this definition to be an integer.

Note also that the only thing we use about $\mathbb{R}^d$ in making this definition is the notion of distance between points in $\mathbb{R}^d$. Quite clearly, the same definition can be made for any metric space X if we just replace $||x - x_n||$ by $\delta(x, x_n)$, where δ is the metric on X. Thus, the definition of convergence for a sequence in a general metric space is the following:

Definition 7.2.6. Let X be a metric space with metric δ. If $\{x_n\}$ is a sequence in X and $x \in X$, then we say $\{x_n\}$ *converges* to x if for every $\epsilon > 0$ there is an $N \in \mathbb{R}$ such that

$$\delta(x, x_n) < \epsilon \quad \text{whenever} \quad n \geq N.$$

In this case, we write $\lim_{n \to \infty} x_k = x$ or $\lim x_n = x$ or simply $x_n \to x$.

We will not try to prove everything in this section in the context of general metric spaces; after all, the object of study here is $\mathbb{R}^d$. However, we will point out some theorems we prove for $\mathbb{R}^d$ that can be proved in general metric spaces or normed vector spaces or inner product spaces, and some of the exercises will be devoted to verifying these claims.

Example 7.2.7. Let $x_n = (1/n^2, 1 + 1/n) \in \mathbb{R}^2$. Use Definition 7.2.5 to prove that the sequence $\{x_n\}$ converges to $x = (0, 1)$.

Solution: We have $x - x_n = (-1/n^2, -1/n)$ and so

$$||x - x_n|| = \sqrt{1/n^4 + 1/n^2} \leq \sqrt{2/n^2} = \sqrt{2}/n.$$

Thus, given $\epsilon > 0$, if we choose $N = \sqrt{2}/\epsilon$, then

$$||x - x_n|| < \sqrt{2}/n \leq \sqrt{2}/N = \epsilon \quad \text{whenever} \quad n \geq N.$$

This completes the proof that $\lim x_n = x$.

Many limit proofs for sequences in $\mathbb{R}^d$ follow the same pattern as in the above example. We showed that $||x - x_n|| < \sqrt{2}/n$ and then used the fact that $\sqrt{2}/n$ can be made less than ϵ by making n large enough – that is, we used the fact that $\lim \sqrt{2}/n = 0$. We can save some effort in future proofs by formalizing in a theorem the method that was used here. The theorem is a vector version of Theorem 2.3.1. In fact, it follows immediately from Theorem 2.3.1 and the fact (obvious from the definition of limit) that $\lim x_n = x$ if and only if $\lim ||x_n - x|| = 0$.

Theorem 7.2.8. *Let $\{x_n\}$ be a sequence in $\mathbb{R}^d$ and let x be a vector in $\mathbb{R}^d$. If there is a sequence $\{a_n\}$ of non-negative real numbers such that*

$$||x - x_n|| \leq a_n \quad \text{for all} \quad n$$

and if $\lim a_n = 0$, then $\lim x_n = x$.

Note that, since the proof of this theorem uses nothing about $\mathbb{R}^d$ but the existence of a metric and the definition of limit, it holds in any metric space (if $||x - x_n||$ is replaced by $\delta(x, x_n)$).

Example 7.2.9. If $x_n = (\mathrm{e}^{-n} \sin n, \mathrm{e}^{-n}) \in \mathbb{R}^2$, prove that $\lim x_n = 0$. **Solution:** We have

$$||x_n - 0|| = ||x_n|| = \sqrt{\mathrm{e}^{-2n}(\sin^2 n + 1)} \leq 2\,\mathrm{e}^{-n} = 2/\,\mathrm{e}^n.$$

Since, $\lim 2/\,\mathrm{e}^n = 0$, the previous theorem tells us that $\lim x_n = 0$.

Limit Theorems

The following theorem says that the limit of a sequence, if it exists, is unique. Its proof is identical to the proof of Theorem 2.1.6. We won't repeat it here. The analogous theorem for metric spaces is also true and also has the same proof.

Theorem 7.2.10. *If $\{x_n\}$ is a sequence in $\mathbb{R}^d$ and $x, y \in \mathbb{R}^d$ with $x_n \to x$ and $x_n \to y$, then $x = y$.*

Theorem 7.2.11. *If* $\lim x_n = x$ *for a sequence* $\{x_n\}$ *in* R^d, *then* $\lim ||x_n|| = ||x||$.

Proof. The second form of the triangle inequality tells us that

$$|\,||x|| - ||x_n||\,| \le ||x - x_n||.$$

If $\lim x_n = x$, then the sequence of numbers on the right converges to 0. It follows that the one on the left also converges to 0. Thus, $\lim ||x_n|| = ||x||$. □

The next theorem is the vector version of the Main Limit Theorem (Theorem 2.3.6) for sequences of real numbers.

Theorem 7.2.12. *If* $\{x_n\}$ *and* $\{y_n\}$ *are sequences of vectors in* $\mathbb{R}^d$ *and* a_n *is a sequence of scalars, and if* $x_n \to x \in \mathbb{R}^d$, $y_n \to y \in \mathbb{R}^d$ *and* $a_n \to a$, *then*

(a) $x_n + y_n \to x + y$;

(b) $a_n x_n \to ax$; *and*

(c) $x_n \cdot y_n \to x \cdot y$.

Proof. (a) By the triangle inequality, we have

$$||x + y - (x_n + y_n)|| \le ||x - x_n|| + ||y - y_n||.$$

Since $x_n \to x$ and $y_n \to y$ we have that $||x - x_n|| \to 0$ and $||y - y_n|| \to 0$. Thus, $||x-x_n|| + ||y-y_n|| \to 0$ and it follows from Theorem 7.2.8 that $x_n + y_n \to x + y$.
    (b) We have

$$||ax - a_n x_n|| = ||a(x - x_n) + (a - a_n)x_n|| \le |a|\,||x - x_n|| + |a - a_n|\,||x_n||.$$

Since $||x - x_n|| \to 0$, $|a - a_n| \to 0$ and $||x_n|| \to ||x||$ (by the previous theorem), the expression on the right converges to 0. Hence, by Theorem 7.2.8 again, $\lim a_n x_n = ax$.
    (c) The proof of this is similar to the proof of (b). The details are left to the exercises. □

Note that the proofs of (a) and (b) above use only properties of $\mathbb{R}^d$ that are also true in any normed vector space, and so they hold in this much more general context. The proof of (c) uses only properties of $\mathbb{R}^d$ that hold in any inner product space and so (c) is true in any inner product space.
    The next theorem tells us that a sequence of vectors converges if and only if it converges componentwise.

Theorem 7.2.13. *A sequence* $\{x_n\}$ *in* $\mathbb{R}^d$ *converges to* $x \in \mathbb{R}^d$ *if and only each component of* $\{x_n\}$ *converges to the corresponding component of* x – *that is, if and only if* $\lim x_n \cdot e_j = x \cdot e_j$ *for* $j = 1, \cdots, d$.

Proof. If $\lim_{n\to\infty} x_n = x$, then $\lim_{n\to\infty} x_n \cdot e_j = x \cdot e_j$ for each j by Theorem 7.2.12, part (c).

To prove the converse, we suppose $\lim_{n\to\infty} x_n \cdot e_j = x \cdot e_j$ for each j. We note that this implies that $\lim_{n\to\infty} |(x_n - x) \cdot e_j| = 0$ for each j. We have,

$$||x_n - x|| = \left(\sum_{j=1}^{d} |(x_n - x) \cdot e_j|^2 \right)^{1/2}.$$

Each term in the sum on the right converges to 0 and, hence, the sum and its square root also converge to 0. We conclude that $\lim x_n = x$. □

The Bolzano-Weierstrass Theorem

The conclusion of the Bolzano-Weierstrass Theorem from Chapter 2 (Theorem 2.5.5) also holds for bounded sequences in $\mathbb{R}^d$. A sequence in $\mathbb{R}^d$ is bounded if there is a number M such that $||x_n|| \leq M$ for all n.

Theorem 7.2.14. (Bolzano-Weierstrass Theorem) *Each bounded sequence in $\mathbb{R}^d$ has a convergent subsequence.*

Proof. We will prove this by induction on the dimension d of the Euclidean space. It is, of course, true for $d = 1$ by the single variable version of the Bolzano-Weierstrass Theorem (Theorem 2.5.5).

Suppose $d > 1$ and the theorem is true for Euclidean space of dimension $d - 1$. Let $\{x_n\}$ be a bounded sequence in $\mathbb{R}^d$. Then there is an $M \in \mathbb{R}$ such that $||x_n|| \leq M$ for all n.

We identify $\mathbb{R}^d$ with the Cartesian product $\mathbb{R}^{d-1} \times R$. This is the space of all pairs (y, z), where $y \in R^{d-1}$ and $z \in \mathbb{R}$. That is, if $x = (x_1, \cdots, x_d) \in \mathbb{R}^d$, then we identify x with the pair (y, z), where $y = (x_1, x_2, \cdots, x_{d-1})$ and $z = x_d$. If this is done, notice that

$$||y|| \leq ||x|| \quad \text{and} \quad |z| \leq ||x||.$$

Thus, if we write each element of the sequence $\{x_n\}$ in the form $x_n = (y_n, z_n) \in \mathbb{R}^{d-1} \times \mathbb{R}$, then $||y_n|| \leq ||x_n|| \leq M$ and $|z_n| \leq ||x_n|| \leq M$. This implies that the sequences $\{y_n\}$ and $\{z_n\}$ are both bounded.

By the induction assumption, the sequence $\{y_n\}$ has a convergent subsequence $\{y_{n_i}\}$. The corresponding subsequence $\{z_{n_i}\}$ of the sequence $\{z_n\}$ is still bounded, and so it has a convergent subsequence. By replacing $\{y_{n_i}\}$ by a (still convergent) subsequence of itself, we may assume that $\{z_{n_i}\}$ itself converges.

The component sequences of $\{x_{n_j}\}$ are those of $\{y_{n_j}\}$, which all converge since $\{y_{n_j}\}$ converges, and the sequence $\{z_{n_i}\}$, which converges. Thus, $\{x_{n_i}\}$ converges since all of its component sequences converge.

We conclude that every bounded sequence in $\mathbb{R}^d$ has a convergent subsequence. This completes the induction and finishes the proof of the theorem. □

Cauchy Sequences

Cauchy sequences in $\mathbb{R}^d$ are defined in the same way as Cauchy sequences of numbers were defined in Definition 2.5.7.

Definition 7.2.15. A sequence $\{x_n\}$ in $\mathbb{R}^d$ is said to be a *Cauchy Sequence* if, for every $\epsilon > 0$, there is an N such that

$$||x_n - x_m|| < \epsilon \quad \text{whenever} \quad n, m \geq N.$$

The following theorem is proved using the Bolzano-Weierstrass Theorem in exactly the same way its single variable counterpart (Theorem 2.5.8) was proved. We won't repeat the proof.

Theorem 7.2.16. *A sequence $\{x_n\}$ in $\mathbb{R}^d$ is a Cauchy sequence if and only if it converges.*

To prove directly from the definition that a certain sequence converges, it is necessary to have in hand the element to which it converges. On the other hand, the definition of a Cauchy sequence involves only the elements of the sequence. Hence, the above theorem provides a way to prove that a sequence converges without having already identified the limit.

Clearly, Cauchy sequences can be defined in any metric space – simply replace "$||x_n - x_m||$" in the above definition by "$\delta(x_n, x_m)$", where δ is the metric. However, the analogue of Theorem 7.2.16 is not true in general for metric spaces. A metric space in which it is true is said to be *complete*. Thus, $\mathbb{R}^d$ is a complete metric space. An example of a metric space which is not complete follows.

Example 7.2.17. Let the interval $(0, 1)$ be considered a metric space with the usual distance between points as metric. Show that this is not a complete metric space.

Solution: The sequence $\{1/n\}$ is a Cauchy sequence since it converges in $\mathbb{R}$ to the point 0. However, since $0 \notin (0, 1)$, this sequence does not converge in the metric space $(0, 1)$. Hence, $(0, 1)$ is not a complete metric space.

Exercise Set 7.2

1. Using only the definition of the limit of a sequence in $\mathbb{R}^d$ prove that
   $$\lim \left(\frac{n}{1+n}, \frac{1-n}{n} \right) = (1, -1).$$

   In each of the next four problems, determine whether or not the sequence $\{x_n\}$ converges and find its limit if it does converge. Use limit theorems to justify your answers.

2. $x_n = \left(\dfrac{n^2 + n - 1}{3n^2 + 2}, \dfrac{n-1}{n+1} \right).$

3. $x_n = (1 + (-1)^n, 1/n, 1 + 1/n)$.

4. $x_n = (2^{-n} \sin(n\pi/4), 2^{-n} \cos(n\pi/4))$;

5. $x_n = (\ln(n+1) - \ln n, \sin(1/n))$.

6. Let $\{x_n\}$ and $\{y_n\}$ be sequences in $\mathbb{R}^d$. Prove that if $\lim x_n = 0$ and $\{y_n\}$ is bounded, then $\lim x_n \cdot y_n = 0$.

7. Let $\{x_n\}$ be a bounded sequence in $\mathbb{R}^d$ and a_n a bounded sequence of scalars. Prove that if either sequence has limit 0, then so does the sequence $\{a_n x_n\}$.

8. Prove that every convergent sequence in $\mathbb{R}^d$ is bounded.

9. If $x_n = (\sin n, \cos n, 1 + (-1)^n)$, does the sequence $\{x_n\}$ in $\mathbb{R}^3$ have a convergent subsequence? Justify your answer.

10. Prove part (c) of Theorem 7.2.12.

11. If $x_n = (1/n, \sin(\pi n/2))$, find three convergent subsequences of $\{x_n\}$ which converge to three different limits.

12. If, for $x, y \in \mathbb{R}$, we set $\delta(x, y) = 0$ if $x = y$ and $\delta(x, y) = 1$ if $x \neq y$, prove that the result is a metric on $\mathbb{R}$. Thus, $\mathbb{R}$ with this metric is a metric space – one that is quite different from $\mathbb{R}$ with the usual metric.

13. What are the convergent sequences in the metric space described in the previous exercise.

14. Let a and b be points of $\mathbb{R}^2$ and let X be the set of all smooth parameterized curves joining a to b in $\mathbb{R}^2$, with parameter interval $[0, 1]$. That is, X is the set of all continuously differentiable functions $\gamma : [0, 1] \to \mathbb{R}^2$, with $\gamma(0) = a$ and $\gamma(1) = b$. Show that if

$$\delta(\gamma_1, \gamma_2) = \sup\{\|\gamma_1(t) - \gamma_2(t)\| : t \in [0, 1]\},$$

then δ is a metric on X.

15. Show that the metric space of the previous exercise is not complete.

16. Let S be the surface of a sphere in $\mathbb{R}^3$. For $x, y \in S$ let $\delta(x, y)$ be the length of the shortest path on S joining x to y. Show that this is a metric on S.

17. Imagine a large building with many rooms. Let X be the set of rooms in this building and let $\delta(x, y)$ be the length of the shortest path along the hallways and stairways of the building that leads from room x to room y. Show that δ is a metric on X.

7.3   Open and Closed Sets

The open ball $B_r(x_0)$ and closed ball $\overline{B}_r(x_0)$, centered at $x_0 \in \mathbb{R}^d$, with radius $r > 0$, are defined by

$$B_r(x_0) = \{x \in \mathbb{R}^d : ||x - x_0|| < r\} \quad \text{and} \quad \overline{B}_r(x_0) = \{x \in \mathbb{R}^d : ||x - x_0)|| \leq r\}.$$

Of course, open and closed balls centered at a given point and with a given radius may be defined in any metric space – one simply uses the metric distance $\delta(x, x_0)$ in place of the distance $||x - x_0||$ defined by the norm in $\mathbb{R}^d$.

   Open intervals and closed intervals on the real line play an important part in the calculus of one variable. Open and closed balls are the direct analogues in $\mathbb{R}^d$ of open and closed intervals on the line. However, the geometry of $\mathbb{R}^d$ is much more complicated than that of the line. We will need the concepts of open and closed for sets that are far more complicated than balls. This leads to the following definition.

Definition 7.3.1. If U is a subset of $\mathbb{R}^d$, we will say that U is *open* if, for each point $x \in U$, there is an open ball centered at x which is contained in U. We will say that a subset of $\mathbb{R}^d$ is *closed* if its complement is open. A *neighborhood* of a point $x \in \mathbb{R}^d$ is any open set which contains x.

   It might seem obvious that open balls are open sets and closed balls are closed sets. However, that is only because we have chosen to call them *open* balls and *closed* balls. We actually have to prove that they satisfy the conditions of the preceding definition. We do this in the next theorem.

Theorem 7.3.2. *In $\mathbb{R}^d$,*

   (a) the empty set $\emptyset$ is both open and closed;

   (b) the whole space $\mathbb{R}^d$ is both open and closed;

   (c) each open ball is open;

   (d) each closed ball is closed.

Proof. The empty set $\emptyset$ is open because it has no points, and so the condition that a set be open, stated in Definition 7.3.1, is vacuously satisfied. The set $\mathbb{R}^d$ is open because it contains any open ball centered at any of its points. Thus, $\emptyset$ and $\mathbb{R}^d$ are both open. Since they are complements of one another, they are also both closed.

   To prove (c), we suppose $B_r(x_0)$ is an open ball and y is one of its points. Then $||y - x_0|| < r$ and so, if we set $s = r - ||y - x_0||$, then $s > 0$. Also, if $x \in B_s(y)$, then $||x - y|| < s$ and so

$$||x - x_0|| \leq ||x - y|| + ||y - x_0|| < s + ||y - x_0|| = r,$$

which means $x \in B_r(x_0)$ (see Figure 7.1). Thus, we have shown that, for each $y \in B_r(x_0)$, there is an open ball, $B_s(y)$, centered at y, which is contained in
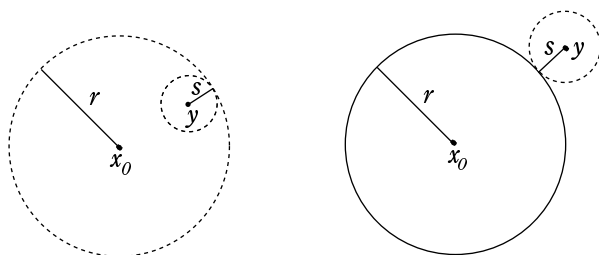
Figure 7.1: Proving Theorem 7.3.2 (c) and (d).

$B_r(x_0)$. By definition, this means that $B_r(x_0)$ is open. This completes the proof of (c).

To prove (d), we consider a closed ball $\overline{B}_r(x_0)$. To prove that it is a closed set, we must show its complement is open. Suppose y is a point in its complement. This means $y \in \mathbb{R}^d$ but $y \notin \overline{B}_r(x_0)$, and so $||y - x_0|| > r$. This time we set $s = ||y - x_0|| - r$ and we claim that the open ball $B_s(y)$ is contained in the complement of $\overline{B}_r(x_0)$. In fact, if $x \in B_s(y)$, then $||x - y|| < s$ and so, by the second form of the triangle inequality (Theorem 2.1.2 (b))

$$||x - x_0|| \geq ||y - x_0|| - ||x - y|| > ||y - x_0|| - s = r,$$

which means x is in the complement of $\overline{B}_r(x_0)$. Thus, we have proved that each point of the complement of $B_r(x_0)$ is the center of an open ball contained in the complement of $\overline{B}_r(x_0)$. This proves that this complement is open, hence, that $\overline{B}_r(x_0)$ is closed. □

The above theorem holds in any metric space and it has the same proof. The same thing is true of the next theorem. It tells us that the collection of all open subsets of $\mathbb{R}^d$ forms what is called a *topology* for $\mathbb{R}^d$. A *topology* for a space X is a collection of sets which are declared to be the open sets of the space. This collection must contain the empty set and the space X and must have the property that it is is closed under arbitrary unions and finite intersections. A space X with a specified topology is called a *topological space*.

Theorem 7.3.3. *In $\mathbb{R}^d$,*

(a) *the union of an arbitrary collection of open sets is open;*

(b) *the intersection of any finite collection of open sets is open;*

(c) *the intersection of an arbitrary collection of closed sets is closed;*

(d) *the union of any finite collection of closed sets is closed.*

Proof. If $\mathcal{V}$ is an arbitrary collection of open sets, and $U = \bigcup \mathcal{V}$ is its union, then x is in U if and only if it is in at least one of the sets in $\mathcal{V}$. Suppose, $x \in V$ with V in $\mathcal{V}$. Then, since V is open, there is a ball $B_r(x)$, centered at x, which

is contained in V. Since $V \subset U$, this ball is also contained in U. This proves that U is open and completes the proof of (a).

Now suppose $\{V_1, V_2, \cdots, V_n\}$ is a finite collection of open sets and

$$x \in U = V_1 \cap V_2 \cap \cdots \cap V_n.$$

Then, since each V_k is open, there exists for each k a radius r_k such that $B_{r_k}(x) \subset V_k$. If $r = \min\{r_1, r_2, \cdots, r_n\}$, then $B_r(x) \subset V_k$ for every k, which implies that $B_r(x) \subset U$. It follows that U is open. This completes the proof of (b).

The proofs of the corresponding statements (c) and (d) for closed sets follow from those for open sets by taking complements. We leave the details to Exercise 7.3.5. ∎

Remark 7.3.4. An easy consequence of the above theorem is that if U is open and K is closed and if $K \subset U$, then the set theoretic difference $U \setminus K$ is open. On the other hand, if $U \subset K$, then $K \setminus U$ is closed (Exercise 7.3.6).

Example 7.3.5. If $0 < r < R$, prove that the annulus

$$A = \{x \in \mathbb{R}^2 : r < ||x|| < R\},$$

is open.

Solution: The ball $B_R(0)$ is open, the ball $\overline{B}_r(0)$ is closed, and A is the set theoretic difference $B_R(0) \setminus \overline{B}_r(0)$. Thus, by the previous remark, A is open.

A similar argument shows that an annulus of the form

$$\{x \in \mathbb{R}^2 : r \leq ||x|| \leq R\}.$$

is closed.

Interior, Closure, and Boundary

If E is a subset of $\mathbb{R}^d$, then the union of all open subsets of E is open, by Theorem 7.3.3. By construction, it is a subset of E which contains all open subsets of E. Thus, every subset of R^d contains a largest open subset – that is, an open subset which contains all other open subsets.

Similarly, the intersection of all closed sets containing E is a closed set containing E and it is contained in every closed set containing E. Thus, it is the smallest closed set containing E.

It is a consequence of this discussion that the following definition makes sense.

Definition 7.3.6. Let E be a subset of $\mathbb{R}^d$. Then:

(a) the largest open subset of E is called the *interior* of E and is denoted $E°$;

(b) the smallest closed set containing E is called the *closure* of E and is denoted $\overline{E}$;
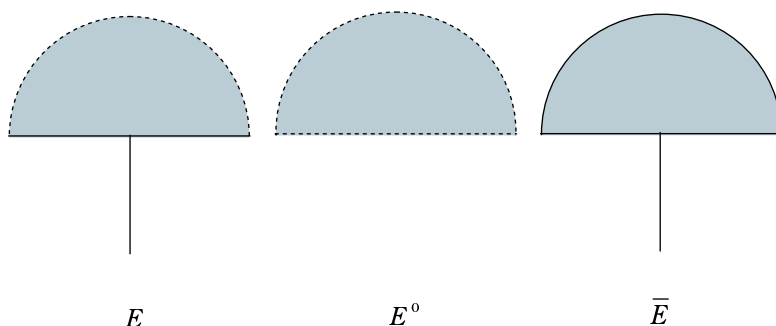
Figure 7.2: The Set E of Example 7.3.8, its Interior E^o, and Closure $\overline{E}$.

(c) the set $\overline{E} \setminus E^\circ$ is called the boundary of E and is denoted ∂E.

Note that these concepts can be defined in exactly the same way in any topological space and, in particular, in any metric space.

Recall that a neighborhood of a point $x \in \mathbb{R}^d$ is any open set containing x. The proof of the following theorem is elementary and is left to the exercises. This theorem also holds in any metric space.

Theorem 7.3.7. *Let E be a subset of $\mathbb{R}^d$ and x an element of $\mathbb{R}^d$. Then:*

(a) *$x \in E^\circ$ if and only if there is a neighborhood of x that is contained in E;*

(b) *$x \in \overline{E}$ if and only if every neighborhood of x contains a point of E;*

(c) *$x \in \partial E$ if and only if every neighborhood of x contains points of E and points of the complement of E.*

Example 7.3.8. Find the interior, closure and boundary for the set

$$E = \{(x, y) \in \mathbb{R}^2 : ||(x, y)|| < 1, \ y \geq 0\} \cup \{(0, -y) : y \in [0, 1]\}.$$

Solution: It is immediate from the previous theorem that

$$E^\circ = \{(x, y) \in \mathbb{R}^2 : ||(x, y)|| < 1, \ y > 0, \}$$
$$\overline{E} = \{(x, y) \in \mathbb{R}^2 : ||(x, y)|| \leq 1, \ y \geq 0\} \cup \{(0, -y) : y \in [0, 1]\},$$
$$\partial E = \{(x, y) \in \mathbb{R}^2 : ||(x, y)|| = 1, \ y \geq 0\} \cup [-1, 1] \cup \{(0, -y) : y \in [0, 1]\}.$$

See Figure 7.2

Sequences

The concepts of open and closed sets are intimately connected to the concept of convergence of a sequence.

Theorem 7.3.9. *A sequence $\{x_n\}$ in $\mathbb{R}^d$ converges to $x \in \mathbb{R}^d$ if and only if, for every neighborhood U of x, there is a number N such that $x_n \in U$ whenever $n \geq N$.*

Proof. If for every neighborhood U of x there is an N such that $x_n \in U$ whenever $n \geq N$, then this is true, in particular, for each neighborhood of the form $B_\epsilon(x)$ with $\epsilon > 0$. This means that for each $\epsilon > 0$ there is an N such that $||x - x_n|| < \epsilon$ whenever $n \geq N$. That is, $\lim x_n = x$.

Conversely, if $\lim x_n = x$ and U is any neighborhood of x, we may choose an $\epsilon > 0$ such that the ball $B_\epsilon(x)$ is contained in U. By the definition of limit, for this ϵ there is an N such that $||x - x_n|| < \epsilon$ whenever $n \geq N$. Then $x_n \in B_\epsilon(x) \subset U$ whenever $n \geq N$. This completes the proof. ◻

Theorem 7.3.10. *If A is a subset of $\mathbb{R}^d$, then $\overline{A}$ is the set of all limits of convergent sequences in A. The set A is closed if and only if every covergent sequence in A converges to a point of A.*

Proof. If $x \in \overline{A}$, then each neighborhood of x contains a point of A by Theorem 7.3.7(b). In particular, each neighborhood of the form $B_{1/n}(x)$, for $n \in \mathbb{N}$, contains a point of A. We choose one and call it x_n. Since $||x - x_n|| < 1/n$, the sequence $\{x_n\}$ converges to x. Thus, each point in the closure of A is the limit of a sequence in A.

Conversely, suppose $x = \lim x_n$ for some sequence $\{x_n\}$ in A. By the previous theorem, each neighborhood of x contains points in this sequence. In particular, each neighborhood of x contains a point of A. Hence, $x \in \overline{A}$ by Theorem 7.3.7(b).

Since a set is closed if and only if it is its own closure, it follows that A is closed if and only if it contains all limits of convergent sequences in A. ◻

Exercise Set 7.3

1. Prove that the set $\{(x,y) \in R^2 : y > 0\}$ is an open subset of R^2.

2. Prove that every finite subset of $\mathbb{R}^d$ is closed.

3. Find the interior, closure, and boundary for the set

$$\{(x,y) \in \mathbb{R}^2 : 0 \leq x < 2,\ 0 \leq y < 1\}.$$

4. Find the interior, closure, and boundary for the set

$$\{(x,y) \in \mathbb{R}^2 : ||(x,y)|| < 1\} \cup \{(x,y) \in \mathbb{R}^2 : y = 0,\ -2 < x < 2\}.$$

5. Prove (c) and (d) of Theorem 7.3.3

6. Let A be an open set and B a closed set. If $B \subset A$, prove that $A \setminus B$ is open. If $A \subset B$, prove that $B \setminus A$ is closed.

7. Prove Theorem 7.3.7.

8. If E is a subset of $\mathbb{R}^d$, is the interior of the closure of E necessarily the same as the interior of E? Justify your answer.

9. If A and B are subsets of $\mathbb{R}^d$ show that $\overline{A \cup B} = \overline{A} \cup \overline{B}$. Is the analogous statement true for $A \cap B$? Justify your answer.

10. If A and B are subsets of $\mathbb{R}^d$, prove that $(A \cap B)^\circ = A^\circ \cap B^\circ$. Is the analogous statement true for $A \cup B$? Justify your answer.

11. Let $\{x_n\}$ be a convergent sequence in $\mathbb{R}^d$ with limit x. Set

$$A = \{x_1, x_2, x_3, \cdots\} \cup \{x\},$$

that is, A is the set consisting of all the points occuring in the sequence together with the limit x. Show that A is a closed set.

12. Let $\{x_n\}$ be any sequence in $\mathbb{R}^d$ and let A be the set consisting of the points that occur in this sequence. Prove that the closure of A consists of A together with all limits of convergent subsequences of A.

13. Show that Theorem 7.3.10 remains true if $\mathbb{R}^d$ is replaced by any metric space.

14. Find the interior and closure of the set Q of rationals in $\mathbb{R}$.

15. If E is a subset of $\mathbb{R}^d$, show that $(\overline{E})^c = (E^c)^\circ$.

7.4 Compact Sets

In this section and the next, we study two topological properties, compactness and connectedness, that a subset of $\mathbb{R}^d$ may or may not have. A topological property of a set E is one that can be described using only knowledge of the open sets of $\mathbb{R}^d$ and their relationship to E. Thus, they are properties that can be defined in any toplological space. Compactness and connectedness are two such properties.

Open Covers

An open cover of a set $E \subset \mathbb{R}^d$ is a collection of open sets whose union contains E. An open cover of a set E may or may not have a finite subcover – that is, there may or may not be finitely many sets in the collection which also form a cover of E.

Example 7.4.1. The collection $\mathcal{U}$ of all open intervals of length $1/2$ and with rational endpoints is clearly an open cover of the interval $[0,1]$. Show that it has a finite subcover.

  Solution: The three intervals $(-1/8, 3/8)$, $(1/4, 3/4)$, and $(5/8, 9/8)$ belong to $\mathcal{U}$ and they cover $[0,1]$.

Example 7.4.2. The collection $\{(1/n, 1) : n = 1, 2, \cdots\}$ is a collection of open sets which covers $(0, 1)$. Does it have a finite subcover?

Solution: No. Since this collection of intervals is nested upward, any finite subcollection has a largest interval $(1/m, 1)$. Then the union of the sets in the subcollection is just $(1/m, 1)$ and this does not contain $(0, 1)$.

Compactness

The above discussion leads to the following definition:

Definition 7.4.3. A subset K of $\mathbb{R}^d$ is called *compact* if every open cover of K has a finite subcover.

Note that Example 7.4.2 shows that the open interval $(0, 1)$ is not compact, since it has an open cover with no finite subcover.

A subset E of $\mathbb{R}^d$ is bounded if there is a number R such that $||x|| \leq R$ for every $x \in E$ – that is, if $E \subset \overline{B}_R(0)$ for some R.

Theorem 7.4.4. *Every compact subset K of $\mathbb{R}^d$ is bounded.*

Proof. We have $K \subset \mathbb{R}^d = \cup_n B_n(0)$. This means that the open balls $B_n(0)$ for $n = 1, 2, \cdots$ form an open cover of K. Since K is compact, finitely many of these balls must also form a cover of K. This implies K is contained in one these balls, say $B_m(0)$, since they form a sequence which is nested upward. Since K is contained in $B_m(0) \subset \overline{B}_m(0)$, it is bounded. $\square$

Theorem 7.4.5. *Every compact subset K of $\mathbb{R}^d$ is closed.*

Proof. We will prove this by showing that $K = \overline{K}$. If $x \in \overline{K}$ and n is a positive integer, we let U_n be the complement in $\mathbb{R}^d$ of $\overline{B}_{1/n}(x)$. The union of the nested sequence of open sets $\{U_n\}$ is $\mathbb{R}^d \setminus \{x\}$.

If some finite subcollection of $\{U_n\}$ covers K then some one of these sets, say U_m, contains K. This means that $B_{1/m}(x) \cap K = \emptyset$, which is impossible, since $x \in \overline{K}$. Because K is compact, this means that $\{U_n\}$ cannot be an open cover of K. Since x is the only point of $\mathbb{R}^d$ not covered by $\{U_n\}$, x must be in K.

We conclude that $K = \overline{K}$ and K is closed. $\square$

The Heine-Borel Theorem

The last two theorems show that a compact subset of $\mathbb{R}^d$ is both closed and bounded. The Heine-Borel Theorem says the the converse is also true – every closed bounded subset of $\mathbb{R}^d$ is compact. Before we prove this, we prove the following analogue of the Nested Interval Theorem (Theorem 2.5.1).

Theorem 7.4.6. *If $A_1 \supset A_2 \supset \cdots \supset A_n \supset A_{n+1} \supset \cdots$ is a nested sequence of non-empty bounded closed subsets of $\mathbb{R}^d$, then $\cap_n A_n \neq \emptyset$.*

Proof. Since each A_n is non-empty, we may choose a point $x_n \in A_n$ for each n. These points are all in A_1, which is bounded. Hence, $\{x_n\}$ is a bounded sequence. By the Bolzano–Weierstrass Theorem (Theorem 7.2.14) this sequence has a convergent subsequence $\{x_{n_k}\}$. Let x be the limit of this subsequence.

Since A_1 is closed and $x_{n_k} \in A_1$ for every k, we have that $x \in A_1$. In fact, for each n, $n_k \geq n$ if $k \geq n$, and so, beginning with the nth term, each term of the sequence $\{x_{n_k}\}$ belongs to A_n. Since A_n is closed, we have $x \in A_n$. We conclude that $x \in \cap_n A_n$. Hence, $\cap_n A_n \neq \emptyset$. $\square$

In the proof of the following theorem, we will make use of the concept of an d-cube in $\mathbb{R}^d$. This is a set of the form $C = I_1 \times I_2 \times \cdots \times I_d$, where each I_j is a closed bounded interval in $\mathbb{R}$ of length L. The intervals I_j are called the *edges* of C and the number L is called the *edge length* of C. Note that a 2-cube is just a square in $\mathbb{R}^2$ with sides parallel to the coordinate axes, while a 3-cube is a cube in $\mathbb{R}^3$ with edges parallel to the axes.

Theorem 7.4.7. (Heine-Borel Theorem) *A subset of $\mathbb{R}^d$ is compact if and only if it is closed and bounded.*

Proof. We already know that every compact subset of $\mathbb{R}^d$ is closed and bounded. Thus, to complete the proof we just need to show that every closed bounded subset of $\mathbb{R}^d$ is compact.

Let K be a closed bounded subset of $\mathbb{R}^d$ and $\mathcal{V}$ an open cover of K. Suppose $\mathcal{V}$ has no finite subcover. We will show that this leads to a contradiction.

Since K is bounded, it lies inside some d-cube C_1. Let L be the edge length of C_1. By partitioning each edge of C_1 at its midpoint, we may partition C_1 into 2^d d-cubes of edge length $L/2$. By intersecting each of these smaller cubes with K, we partition K into finitely many subsets. If each of these is covered by finitely many of the sets in $\mathcal{V}$, then K itself is also. Since it is not, we conclude that the intersection of K with at least one of these smaller d-cubes is not covered by finitely many sets in $\mathcal{V}$. Choose one and call it C_2.

By continuing in this way (actually, by induction), we may construct a nested sequence of d-cubes (see Figure 7.3)

$$C_1 \supset C_2 \supset \cdots \supset C_n \supset C_{n+1} \supset \cdots,$$

where, for each n, C_n is a closed d-cube of edge length $L/2^{n-1}$ and with the property that $C_n \cap K$ cannot be covered by finitely many of the sets in $\mathcal{V}$.

The sets $C_n \cap K$ form a sequence of closed, bounded sets, nested downward, as in the previous theorem. By that theorem $\cap_n(C_n \cap K)$ is not empty, Let x be a point in this intersection. Then $x \in K$ and, since $\mathcal{V}$ is an open cover of K, there is some open set V in the collection $\mathcal{V}$ such that $x \in V$. Since V is open, there is an open ball $B_r(x)$, centered at x, which is contained in V.

The diameter of C_n (maximum distance between two points of C_n) is less than $dL/2^{n-1}$. Hence, for large enough n, the diameter of C_n is less than r. Then C_n must be contained in $B_r(x)$ since it contains x. This implies that $C_n \subset V$. This is a contradiction, since C_n was chosen so that no finite
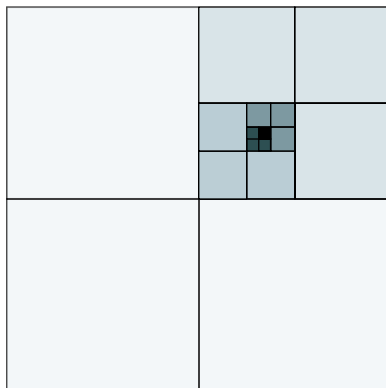
Figure 7.3: Nested Cubes of Theorem 7.4.7

subcollection of the sets in $\mathcal{V}$ covers $C_n \cap K$. Thus, our assumption that K is not covered by any finite subcollection of $\mathcal{V}$ has led to a contradiction.

We conclude that every open cover of K has a finite subcover and, hence, that K is compact. $\qquad\square$

Corollary 7.4.8. *Each closed subset of a compact set in $\mathbb{R}^d$ is also compact.*

Proof. If A is closed and contained in a compact set K, then A is bounded because K is bounded. Since A is closed and bounded, it is compact by the Heine-Borel Theorem. $\qquad\square$

Applications of Compactness

The next chapter will contain a large number of applications of compactness to function theory. The next example illustrates a technique that is often used in such applications.

Example 7.4.9. Let K be a compact subset of $\mathbb{R}^d$ and let ρ be a function defined on K with $\rho(x) > 0$ for each $x \in K$. Prove there exists a finite set of points $\{x_1, x_2, \cdots, x_m\}$ such that K is contained in the union of the open balls $B_{\rho(x_i)}(x_i)$ for $i = 1, 2, \cdots, m$.

Solution: The collection of open sets $\{B_{\rho(x)}(x) : x \in K\}$ is an open cover of K (since, for each $y \in K$, $y \in B_{\rho(y)}(y) \subset \cup\{B_{\rho(x)}(x) : x \in K\}$). Since K is compact, there is a finite subcover $\{B_{\rho(x_i)}(x_i) : i = 1, \cdots m\}$. This means K is contained in the union of the $B_{\rho(x_i)}(x_i)$ for $i = 1, 2, \cdots, m$.

The next theorem is an application of this technique. It is a *separation theorem* which shows that a compact set is *separated* from the complement of any open set that contains it.

Theorem 7.4.10. *Suppose K is a compact subset and U an open subset of $\mathbb{R}^d$ with $K \subset U$. Then there exists an open set V such that $\overline{V}$ is compact and $K \subset V \subset \overline{V} \subset U$.*

Proof. Since U is open and contains K, for each $x \in K$ there is an open ball centered at x which lies in U. Then the ball, centered at x, of half this radius has its closure contained in U. Let $\rho(x)$be the radius of this smaller ball. Then $x \in B_{\rho(x)}(x) \subset \overline{B}_{\rho(x)}(x) \subset U$. By the previous example, there are finitely many points $x_1, \cdots, x_m$ such that K is contained in the union V of the sets $B_{\rho(x_i)}(x_i)$. The closure of V is contained in the compact set which is the union of the sets $\overline{B}_{\rho(x_i)}(x_i)$, and this is contained in U. Thus, $\overline{V}$ is compact, since it a closed subset of a compact set, and $K \subset V \subset \overline{V} \subset U$. $\qquad\square$

Compact Metric Spaces

Since compactness is a topological property, it makes perfectly good sense in any metric space. The definition of a compact subset of a metric space X is exactly the same as Definition 7.4.3 except that $\mathbb{R}^d$ is replaced by X. If the space X itself is compact, then X is called a *compact metric space*.

Any compact subset of $\mathbb{R}^d$ is a compact metric space if it is considered a space by itself and is given the same metric it has as a subset of $\mathbb{R}^d$.

Exercise Set 7.4

1. If K is a compact subset of $\mathbb{R}^d$ and $U_1 \subset U_2 \subset \cdots \subset U_k \subset \cdots$ is a nested upward sequence of open sets with $K \subset \cup_k U_k$, then prove that K is contained in one of the sets U_k.

2. Let K be a compact subset of $\mathbb{R}^d$ and $A_1 \supset A_2 \supset \cdots \supset A_j \supset \cdots$ a nested downward sequence of closed subsets of $\mathbb{R}^d$. Show that if $A_k \cap K \neq \emptyset$ for each k, then $(\cap_k A_k) \cap K \neq \emptyset$.

3. Show that if $K_1 \supset K_2 \supset \cdots \supset K_j \supset \cdots$ is a nested downward sequence of compact sets and U is an open set which contains $\cap_j K_j$, then U contains one of the sets K_j.

4. Prove that if K is a compact subset of $\mathbb{R}^d$, then K contains a point of maximal norm. That is, there is a point $x_1 \in K$ such that

$$||x|| \leq ||x_1|| \quad \text{for all} \quad x \in K.$$

   Hint: Set $m = \sup\{||x|| : x \in K\}$ and consider the open balls $B_{m-1/n}(0)$.

5. Prove that if K is a compact subset of $\mathbb{R}^d$ and y is a point of $\mathbb{R}^d$ which is not in K, then there is a closest point to y in K. That is, there is an $x_0 \in K$ such that

$$||x_0 - y|| \leq ||x - y|| \quad \text{for all} \quad x \in K.$$

6. Prove that the conclusion of the previous exercise also holds if we only assume that K is a closed subset of $\mathbb{R}^d$. Hint: replace K by its intersection with a suitably large closed ball centered at y.

7. Prove that if K_1, K_2 is a disjoint pair of compact sets, then there exists a disjoint pair of open sets V_1, V_2 such that $K_1 \subset V_1$ and $K_2 \subset V_2$. Hint: Use Theorem 7.4.10.

8. Prove that a set $K \subset \mathbb{R}^d$ is compact if and only if every sequence in K has a subsequence which converges to an element of K. Hint: use the Bolzano – Weierstrass and Heine–Borel Theorems.

9. Show that it is true that the union of any finite collection of compact subsets of $\mathbb{R}^d$ is compact, but it is not true that the union of an infinite collection of compact subsets is necessarily compact. Show the latter statement by finding an example of an infinite union of compact sets which is not compact.

10. Prove that if A and B are compact subsets of a metric space, then $A \cup B$ and $A \cap B$ are also compact.

11. Prove that if X is a compact metric space, then every sequence in X has a convergent subsequence.

12. Prove that if X is a compact metric space, then every closed subset of X is also compact.

13. Prove that a compact metric space is complete (that is, every Cauchy sequence converges).

14. We will say a metric space X is *bounded* if, for some $M > 0$ and $x \in X$, the entire space X is contained in $B_M(x) = \{y \in X : \delta(x,y) \leq M\}$. Show that a compact metric space is bounded.

15. Consider the metric space of Exercise 7.2.12. Show that it is complete and bounded, but not compact. Thus, the analogue of the Heine-Borel Theorem does not hold in general metric spaces.

7.5   Connected Sets

Consider the three sets A, B, C described in Figure 7.4. Each of these sets is the union of two closed discs of radius one in $\mathbb{R}^2$. In A the distance between the centers of the two discs is greater than 2; in B it is less than 2 and in C it is exactly 2. The point about these three sets that we wish to discuss is this: set A is *disconnected* – one cannot pass from one of the discs making up this set to the other without leaving the set. On the other hand, B and C are *connected* – one can pass from any point in the set to any other point in the set without leaving the set. As stated so far, these are not very precise ideas. The precise definition of connectedness is as follows.
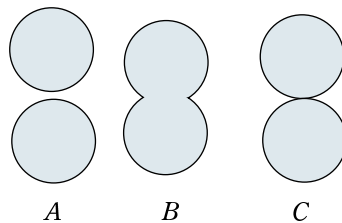
Figure 7.4: Disconnected and Connected Sets

Definition 7.5.1. A subset E of $\mathbb{R}^d$ is said to be *separated* by a pair of open sets U and V in $\mathbb{R}^d$ if

(a) $E \subset U \cup V$;

(b) $(E \cap U) \cap (E \cap V) = \emptyset$;

(c) $E \cap U \neq \emptyset$, and $E \cap V \neq \emptyset$.

If no pair of open subsets of $\mathbb{R}^d$ separates E, then we will say that E is *connected*.

The above definition becomes somewhat simpler to state if we give a special name to subsets of E of the form $E \cap U$ where U is an open set.

Definition 7.5.2. Let E be a subset of $\mathbb{R}^d$. A subset A of E is said to be *relatively open* (in E) if it has the form $A = E \cap U$ for some open subset U of $\mathbb{R}^d$. Similarly, a subset B is said to be *relatively closed* (in E) if it has the form $E \cap C$ for some closed subset C of $\mathbb{R}^d$.

Using these concepts, the definition of connecteness can be rephrased as follows.

Remark 7.5.3. A subset E of $\mathbb{R}^d$ is connected if and only if it is not the disjoint union of two non-empty relatively open subsets.

Connected Subsets of $\mathbb{R}$

The connected subsets of $\mathbb{R}$ are easily characterized.

Theorem 7.5.4. *A non-empty subset of $\mathbb{R}$ is connected if and only if it is an interval.*

Proof. Suppose E is a non-empty subset of $\mathbb{R}$. Let

$$a = \inf E \quad \text{and} \quad b = \sup E.$$

Now a and b may not be finite, but E is certainly contained in the interval consisting of (a, b) together with $\{a\}$ if a is finite and $\{b\}$ if b is finite. The set E will be an interval if and only if it contains (a, b).

Suppose E is not an interval. Then there is an $x \in (a, b)$ such that $x \notin E$. Then E is contained in the set $(-\infty, x) \cup (x, \infty)$. Furthermore, since $a = \inf E$ and $a < x$, there must be points of E which are less than x – that is, $E \cap (-\infty, x) \neq \emptyset$. Similarly, since $b = \sup E$ and $x < b$, $E \cap (x, \infty) \neq \emptyset$. Thus, by Definition 7.5.1, the set E is separated by the pair of open sets $(-\infty, x)$ and (x, ∞) and, hence, is not connected. Thus, if E is connected, it must be an interval.

Conversely, suppose E is an interval. Then E is (a, b) possibly together with one or more of its endpoints. Suppose U and V are open subsets of $\mathbb{R}$ with $(U \cap E) \cap (V \cap E) = \emptyset$ and $E \subset U \cup V$. We define a function f on E by $f(x) = 0$ if $x \in E \cap U$ and $f(x) = 1$ if $x \in E \cap V$.

We claim f is a continuous function on the interval E. If $x \in E$ and $\epsilon > 0$, then x is in one of the sets U or V. Since they are both open, there is an interval $(x - \delta, x + \delta)$ which is also contained in whichever of these sets contains x. Then f has the same value at any $y \in E \cap (x - \delta, x + \delta)$ that it has at x. Thus,

$$|f(x) - f(y)| = 0 < \epsilon \quad \text{whenever} \quad y \in E \quad \text{and} \quad |x - y| < \delta.$$

This proves that f is continuous on E. However, its only possible values are 0 and 1. By the Intermediate Value Theorem (Theorem 3.2.3) it cannot take on both these values, since it would then have to take on every value in between. This means one of the sets $E \cap U$, $E \cap V$ is empty. Hence, E is not separated by U and V. We conclude that no pair of open sets separates E and, hence, E is connected. $\square$

If L is a straight line in $\mathbb{R}^d$, then the intersection of an open ball in $\mathbb{R}^d$ with L is an open interval in L (or is empty). It follows that the relatively open subsets of L are exactly the open subsets of L considered as a copy of $\mathbb{R}$. It follows from the above theorem that intervals in L are connected subsets of $\mathbb{R}^d$. Thus, the line segment joining two points in $\mathbb{R}^d$ is a connected set.

Connected Components

Theorem 7.5.5. *If A and B are connected subsets of $\mathbb{R}^d$ and $A \cap B \neq \emptyset$, then $A \cup B$ is also connected.*

Proof. Suppose U and V are disjoint relatively open subsets of $A \cup B$ such that $A \cup B = U \cup V$. Then $U \cap A$ and $V \cap A$ are disjoint relatively open subsets of A. Since A is connected, U and V cannot both have non-empty intersection with A. Since A is contained in their union and can't meet both of them, A must be contained in either U or V. Similarly, B must be contained in either U or V. Since U and V are disjoint and A and B are not, A and B must be contained in the same one of the sets U, V and must both be disjoint from the other. Since $U \cup V = A \cup B$ one of the sets U, V is empty. This shows that U and V do not separate $A \cup B$. Hence, $A \cup B$ is connected. $\square$

Basically the same argument shows that the union of any collection of connected sets with at least one point is common is also connected (Exercise 7.5.6).
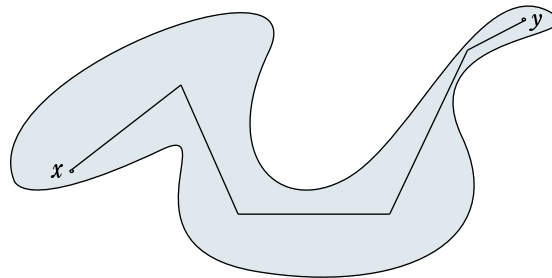
Figure 7.5: A piecewise linear path in E

In particular, if $x \in E$ where E is some subset of $\mathbb{R}^d$, then the union of all connected subsets of E containing x is itself connected. Thus, for each point $x \in E$ there is a connected subset of E which contains all connected subsets containing x – that is, a *maximal* connected subset containing x.

Definition 7.5.6. If E is a subset of $\mathbb{R}^d$ and $x \in E$, then the union of all connected subsets of E containing x is called the *connected component* of E containing x.

Clearly, the connected components of E are the maximal connected subsets of E. Any two distinct components are disjoint since, otherwise, their union would be a connected set larger than at least one of them. Two points x and y of E are in the same component of E if and only if there is some connected subset of E that contains both x and y. In particular, if the line segment joining two points x and y of E also lies in E, then x and y are in the same connected component of E.

Since every point in an open or closed ball is joined by a line segment to the center of the ball, we have:

Theorem 7.5.7. *Every open or closed ball in $\mathbb{R}^d$ is a connected set.*

More generally, a piecewise linear path joining x and y in E is a finite set of line segments $\{[x_{i-1}, x_i]\}_{i=1}^m$, each contained in E, with each line segment beginning where the preceding one ends, and with $x_0 = x$ and $x_m = y$. One easily proves by induction that the union of the line segments in such a path is a connected set (see Figure 7.5). It follows that:

Theorem 7.5.8. *If E is a subset of $\mathbb{R}^d$ and x and y are points of E that may be joined by a piecewise linear path in E, then x and y are in the same component of E. If every pair of points in E can be joined by a piecewise linear path in E, then E is connected.*

Example 7.5.9. Find a subset of $\mathbb{R}^2$ with infinitely many components.
  Solution: This is easy. The set of integers on the x-axis is such a set. Since the only connected subsets of this set are the single point subsets, each
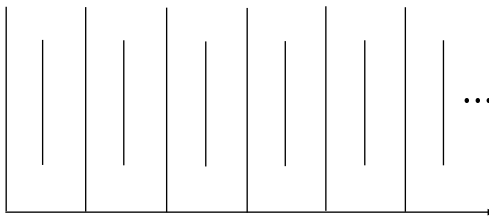
Figure 7.6: A set with infinitely many components

point is a component. A more complicated example is illustrated in Figure 7.6. The vertical lines that touch the bottom horizontal line together with this horizontal line form one component, while each of the shorter vertical lines is itself a component.

Components of an Open Set

Theorem 7.5.10. *If U is an open subset of $\mathbb{R}^d$, then each of its connected components is also open.*

Proof. Let V be a connected component of the open set U and let x be a point of V. Since U is open, there is an open ball $B_r(x)$, centered at x, such that $B_r(x) \subset U$. Since V is the union of all connected subsets of U containing x and $B_r(x)$ is connected, it must be true that $B_r(x) \subset V$. Since every point of V is the center of an open ball contained in V, the set V is open. $\square$

The components of an open set U form a pairwise disjoint family of open connected subsets of U with union U, Conversely:

Theorem 7.5.11. *If an open set U can be written as the union of a pairwise disjoint family $\mathcal{V}$ of open connected subsets, then these subsets must be the components of U.*

Proof. If V is one of the open sets in $\mathcal{V}$, then V must have non-empty intersection with at least one component of U, call it C. Then $V \subset C$ since V is a connected set containing a point of the component C.

We must also have $C \subset V$, since, otherwise, V and the union of all the sets in $\mathcal{V}$ other than V would be two open sets which separate C. Thus, $V = C$.

We now have that every set in $\mathcal{V}$ is a component of U. Since the union of the sets in $\mathcal{V}$ is U, every component of U must occur in $\mathcal{V}$. This completes the proof. $\square$

Example 7.5.12. What are the components of the complement of the set $D \cup E$ where

$$D = \{(x,y) \in R^2 : ||(x+1, y)|| = 1\} \quad \text{and} \quad E = \{x \in R^2 : ||(x-1, y)|| = 1\}.$$

Solution: The complement of $D \cup E$ is the union of the open sets

$$A = \{(x, y) \in \mathbb{R}^2 : ||(x + 1, y)|| < 1\},$$
$$B = \{(x, y) \in \mathbb{R}^2 : ||(x - 1, y)|| < 1\}, \quad \text{and} \qquad (7.5.1)$$
$$C = \{(x, y) \in \mathbb{R}^2 : ||(x + 1, y)|| > 1 \text{ and } ||(x - 1, y)|| > 1\}.$$

These three sets are pairwise disjoint and each of them is connected. Hence, they must be the components of the complement of $D \cup E$, by the previous theorem.

Exercise Set 7.5

In the first four exercises below, tell whether or not the set A is connected. If A is not connected, describe its connected components. Justify your answers.

1. $A = \{(x, y) \in \mathbb{R}^2 : ||(x, y)|| < 1\} \cup \{(x, y) \in \mathbb{R}^2 : 1 \leq x \leq 2, \ y = 0\}$.

2. $A = \{(x, y) \in \mathbb{R}^2 : ||(x, y)|| < 1\} \cup \{(x, y) \in \mathbb{R}^2 : 1 < x \leq 2, \ y = 0\}$.

3. $A = \{(x, y) \in \mathbb{R}^2 : 1 < ||(x, y)|| < 2\}$.

4. $A = \{(x, y) \in \mathbb{R}^2 : 1 < ||(x, y)|| < 2\} \cup \{(x, y) \in \mathbb{R}^2 : ||(x, y)|| < 1\}$.

5. What are the connected components of the complement of the set of integers in $\mathbb{R}$?

6. Prove that the union of a collection of connected subsets of $\mathbb{R}^d$ with a point in common is also connected.

7. Which subsets of $\mathbb{R}$ are both compact and connected? Justify your answer.

8. Give an example of two connected subsets of $\mathbb{R}^2$ whose intersection is not connected.

9. Prove that if E is an open connected subset of $\mathbb{R}^d$, then each pair of points in E can be connected by a piecewise linear path in E. Hint: fix a point $x_0 \in E$ and consider two sets: (1) the set U of all points in E that can be connected to x_0 by a piecewise linear path in E, and (2) the set V of points in E that cannot be connected to x_0 by a piecewise connected path in E.

10. Prove that the closure of a connected set is connected.

11. Is the interior of a connected set necessarily connected? Justify your answer.

12. Are the components of a closed set necessarily closed? Justify your answer.

13. Connected sets in a metric space (or any topological space) are defined in the same way as they are in $\mathbb{R}^d$. Is it true in general for metric spaces that open balls are connected?

14. A subset of a metric space is said to be *totally disconnected* if its components are all single points. Find a compact, totally disconnected subset of $\mathbb{R}$ which is not a finite set.

15. Find a compact, totally disconnected subset of $\mathbb{R}$ (see the previous exercise) which has no isolated points (a point $x \in E$ is an isolated point of E if $\{x\}$ is relatively open in E – that is, if there is an open set U such that $U \cap E = \{x\}$).

Chapter 8

Functions on Euclidean Space

In this chapter we begin the study of functions defined on a subset of the Euclidean space $\mathbb{R}^p$ with values in the Euclidean space $\mathbb{R}^q$. Our first objective is to define and study continuity for such functions.

8.1   Continuous Functions of Several Variables

For two natural numbers p and q, we shall study functions F, defined on a subset D of $\mathbb{R}^p$ and with values in $\mathbb{R}^q$. Such a function is sometimes called a *transformation* from D to $\mathbb{R}^q$. We will denote this situation by $F : D \to \mathbb{R}^q$. The definition of continuity in this context follows the familiar pattern.

Definition 8.1.1. Let D be a subset of $\mathbb{R}^p$ and $F : D \to \mathbb{R}^q$ a function. We say that F is continuous at $a \in D$ if for each $\epsilon > 0$ there is a $\delta > 0$ such that

$$||F(x) - F(a)|| \leq \epsilon \quad \text{whenever} \quad x \in D \quad \text{and} \quad ||x - a|| < \delta.$$

If F is continuous at each point of D, then F is said to be continuous on D.

Note that this definition depends very much on the domain D of the function due to the fact that the condition on $||F(x) - F(a)||$ is only required to hold for $x \in D$. If the domain of the function is changed, then what it means for a function to be continuous at a may change even if a is in both domains.

Example 8.1.2. The function $f : \mathbb{R}^p \to \mathbb{R}$ which is 1 on $\overline{B}_1(0)$ and 0 everywhere else is clearly not continuous at boundary points of $\overline{B}_1(0)$. Show that, if the domain of f is changed to $\overline{B}_1(0)$, then the new function is continuous on all of $\overline{B}_1(0)$.

  Solution: The new function is just the identically 1 function on its domain and, hence, is continuous at each point of its domain – including points of the boundary.

Example 8.1.3. Consider the function $f : \mathbb{R}^2 \to \mathbb{R}$ defined by

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

Show that f is not continuous at $(0,0)$.

    Solution: This function has the value 0 at $(0,0)$, but every disc centered at $(0,0)$ contains points of the form (x,x) with $x \neq 0$ and, at such a point, f has the value $1/2$. So the condition for continuity at $(0,0)$ will not be satisfied when ϵ is $1/2$ or less.

Example 8.1.4. Show that the function with domain $\mathbb{R}^2$ defined by

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

is continuous at $(0,0)$.

    Solution: Since $(x+y)^2 \geq 0$ and $(x-y)^2 \geq 0$, it follows that $-2xy \leq x^2 + y^2$ and $2xy \leq x^2 + y^2$. Taken together, these two inequalities imply that

$$2|xy| \leq x^2 + y^2$$

On dividing by $2\sqrt{x^2 + y^2}$ this becomes

$$|f(x,y) - f(0,0)| = \left| \frac{xy}{\sqrt{x^2 + y^2}} \right| \leq \frac{1}{2}\sqrt{x^2 + y^2} = \frac{1}{2}||(x,y) - (0,0)||.$$

Thus, given $\epsilon > 0$, if $\delta = 2\epsilon$, then

$$|f(x,y) - f(0,0)| < \epsilon \quad \text{whenever} \quad ||(x,y) - (0,0)|| < \delta.$$

We conclude that f is continuous at $(0,0)$.

Vector Valued Functions

The previous two examples involved real valued functions, We will also be concerned with functions with values in $\mathbb{R}^q$ for some natural number $q > 1$. Given such a function F with domain $D \subset \mathbb{R}^p$, for each $x \in D$ let $f_j(x) = e_j \cdot F(x)$ be the jth component of the vector $F(x) \in \mathbb{R}^q$. Then each f_j is a real valued function on D. We will sometimes denote the function F by

$$F(x) = (f_1(x), f_2(x), \cdots, f_q(x)).$$

The real valued function f_j is called the jth component function of F.

Theorem 8.1.5. *A function $F : D \to \mathbb{R}^q$ is continuous at a point $a \in D$ if and only if each of its component functions is continuous at a.*

Proof. It follows from Theorem 7.1.13 that, for each k and each $x \in D$,

$$|f_k(x) - f_k(a)| \leq ||F(x) - F(a)|| \leq \sum_{j=1}^{q} |f_j(x) - f_j(a)|.$$

Given $\epsilon > 0$, It follows from the first inequality that if $||F(x) - F(a)|| < \epsilon$, then also $|f_k(x) - f_k(a)| < \epsilon$ for each k. Hence, if F is continuous at x_0, then so is each f_k. It follows from the second inequality that if $|f_j(x) - f_j(a)| < \epsilon/q$ for each j, then $||F(x) - F(a)|| < \epsilon$. This implies that if each f_j is continuous at a, then so is F. □

Sequences and Continuity

Recall that Theorem 3.1.5 says that a function f of one variable is continuous at a point a of its domain D if and only if it takes sequences in D which converge to a to sequences which converge to $f(a)$. The same theorem is true of functions of several variables, in fact, it is true of any function from one metric space to another. The proof is also the same and we won't repeat it.

Theorem 8.1.6. *Let D be a subset of $\mathbb{R}^p$, $a \in D$, and $F : D \to \mathbb{R}^q$ a transformation. Then F is continuous at a if and only if, whenever $\{x_n\}$ is a sequence in D which converges to a, then the sequence $\{F(x_n)\}$ converges to $F(a)$.*

If F and G are two functions with domain $D \subset \mathbb{R}^p$ and with values in $\mathbb{R}^q$ and if h is a real valued function with domain D, then we can define new functions, hF, $F + G$, and $F \cdot G$ by

$$\begin{aligned}
(hF)(x) &= h(x)F(x), \\
(F + G)(x) &= F(x) + G(x), \\
(F \cdot G)(x) &= F(x) \cdot G(x).
\end{aligned} \tag{8.1.1}$$

Theorems 7.2.12 and 8.1.6 combine to prove the following theorem. The details are left to the exercises.

Theorem 8.1.7. *With F, G, h, and D as above, if F, G, and h are continuous at $a \in D$, then so are hF, $F + G$, and $F \cdot G$.*

Composition of Functions

If $G : D \to \mathbb{R}^p$ is a function with domain $D \subset \mathbb{R}^d$ and $F : E \to \mathbb{R}^q$ is a function with domain $E \subset \mathbb{R}^p$, then $F(G(x))$ is defined as long as $x \in D$ and $G(x) \in E$. Thus,

$$(F \circ G)(x) = F(G(x))$$

defines a function with domain $D \cap G^{-1}(E)$ and with values in $\mathbb{R}^q$. This is the *composition* of the function G with the function F.

The following theorem follows immediately from two applications of Theorem 8.1.6. The details are left to the exercises.

Theorem 8.1.8. *With F and G as above, if $a \in D \cap G^{-1}(E)$, G is continuous at a, and F is continuous at $G(a)$, then $F \circ G$ is continuous at a.*

Limits

Whether or not a function F is defined at a point $a \in \mathbb{R}^p$, it may have a limit as x approaches a. In order for this concept to make sense, it must be the case that there are points of the domain of F which are arbitrarily close but not equal to a.

If D is a subset of $\mathbb{R}^p$ and $a \in \mathbb{R}^p$, then we will say that a is a *limit point* of D if every neighborhood of a contains points of D different from a (note that a may or may not be in D).

Definition 8.1.9. If $D \subset \mathbb{R}^p$, a is a limit point of D, and $F : D \to \mathbb{R}^q$ is a function with domain D, then we will say that the limit of F as x approaches a is b if, for each $\epsilon > 0$, there is a $\delta > 0$ such that

$$||F(x) - b|| < \epsilon \quad \text{whenever} \quad x \in D \quad \text{and} \quad 0 < ||x - a|| < \delta.$$

In this case, we write $\lim_{x \to a} F(x) = b$.

If we compare this definition with the definition of continuity at a (Definition 8.1.1), we see that a function $F : D \to \mathbb{R}^q$ is continuous at a point $a \in D$ which is a limit point of D if and only if $\lim_{x \to a} F(x) = F(a)$.

On the other hand, if $a \in D$ but a is not a limit point of D, then a function F, with domain D is automatically continuous at a (since, for small enough δ, there are no points $x \in D$ with $||x - a|| < \delta$ other than $x = a$), but the limit of F as x approaches a is not defined. A point of D which is not a limit point of D is called an *isolated point* of D. For example, the set $D = B_1((0,0)) \cup \{(1,1)\}$ is a subset of $\mathbb{R}^2$ with $(1,1)$ as an isolated point.

Note that Examples 8.1.3 and 8.1.4 show that

$$\lim_{(x,y) \to (0,0)} \frac{xy}{\sqrt{x^2 + y^2}} = 0,$$

while

$$\lim_{(x,y) \to (0,0)} \frac{xy}{x^2 + y^2}$$

does not exist. In fact, this function has limit

$$\frac{a}{1 + a^2}$$

as (x,y) approaches $(0,0)$ along the line $y = ax$. Since the function approaches different numbers as (x,y) approaches $(0,0)$ from different directions, the limit does not exist.

Curves and Surfaces

A continuous function $\gamma : I \to \mathbb{R}^q$, where I is an interval in $\mathbb{R}$, is called a *parameterized curve* with parameter interval I. The variable t in $\gamma(t)$ is called the *parameter* for the curve. Intuitively, as t ranges through the parameter interval, $\gamma(t)$ traces out something like a curved line in $\mathbb{R}^q$.

If the parameter interval I is a closed bounded interval $[a, b]$ with $\gamma(a) = x$ and $\gamma(b) = y$, then γ is called a curve in $\mathbb{R}^q$ joining x to y. The points x and y are called the *endpoints* of the curve. If $x = y$, then γ is called a closed curve.

Example 8.1.10. Give examples of a closed curve, a curve with endpoints which is not closed, and a curve with no endpoints.

Solution: The curve $\gamma(t) = (\cos t, \sin t)$, $t \in [0, 2\pi]$, is a closed curve in $\mathbb{R}^2$. It is closed because $\gamma(0) = (1, 0) = \gamma(2\pi)$.

The curve $\gamma(t) = (t^2, t^3)$, $t \in [0, 1]$, is a curve joining $x = (0, 0)$ and $y = (1, 1)$. It has these points as endpoints. It is not closed, since the endpoints are not the same.

The curve $\gamma(t) = (t \cos t, t \sin t, t)$, $t \in (-\infty, \infty)$ is a spiral curve in $\mathbb{R}^3$ with no endpoints.

Generally, a curve is a one dimensional object, but there are exceptions. A curve may be *degenerate* – that is, $\gamma(t)$ may be a constant vector in $\mathbb{R}^q$. Then the image of γ is a single point, which is a zero dimensional object.

A *parameterized surface* in $\mathbb{R}^q$ ($q \geq 2$) is a continuous function $F : A \to \mathbb{R}^q$, where A is an open subset of $\mathbb{R}^2$ or an open subset of $\mathbb{R}^2$ together with all or part of the boundary of this open subset.

Example 8.1.11. Give three examples of parameterized surfaces.

Solution: The image of the surface

$$F(\theta, \phi) = (\cos \theta \cos \phi, \sin \theta \cos \phi, \sin \phi) \quad \text{with} \quad \theta \in [0, 2\pi), \; \phi \in [0, \pi]$$

is the sphere of radius 1 centered at the origin. The parameter set A in this case is the rectangle $[0, 2\pi) \times [0, \pi]$. The parameterization is the one given by expressing the sphere in spherical coordinates. Note that this sphere is just $\overline{B}_1(0) \setminus B_1(0)$ and, hence, is a closed set (Exercise 7.3.6) even though its parameter set is not closed.

The closed upper half of the above sphere may be parameterized as above but with parameter set $[0, 2\pi) \times [0, \pi/2]$ or it may be parameterized by

$$G(x, y) = (x, y, \sqrt{1 - x^2 - y^2}) \quad \text{with} \quad x^2 + y^2 \leq 1.$$

Here, the set A is the closed disc of radius 1 centered at the origin in $\mathbb{R}^2$.

If we change the parameter set for G in the above example to the open disc of radius 1 centered at 0, then we obtain a surface which is not a closed set – the upper half of the unit sphere not including the circle $\{(x, y, z) : x^2 + y^2 = 1, \; z = 0\}$.

Generally , the image of a parameterized surface is a two dimensional object, but there are exceptions. A surface may be *degenerate.* The parameter function F could have image contained in a set of dimension less than 2 – it could be a point, or a curve. For example, the image of

$$F(u, v) = (\cos(u + v), \sin(u + v), u + v) \quad \text{with} \quad (u, v) \in \mathbb{R}^2$$

is actually the spiral curve $(\cos t, \sin t, t)$, as we can see by making the substitution $t = u + v$.

Conditions that guarantee that a curve or surface is not degenerate will be obtained in the next chapter.

Exercise Set 8.1

1. Consider the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

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

Is this function continuous at $(0, 0)$? Justify your answer.

2. Give a simple reason why the function $\gamma : \mathbb{R} \rightarrow \mathbb{R}^4$ defined by

$$\gamma(t) = (t, \sin t, \mathrm{e}^t, t^2)$$

is continuous on $\mathbb{R}$.

3. Does the function $f : \mathbb{R}^2 \setminus \{(0, 0)\} \rightarrow \mathbb{R}$, defined by

$$f(x, y) = \frac{x}{\sqrt{x^2 + y^2}},$$

have a limit as (x, y) approaches $(0, 0)$. Justify your answer.

4. Consider the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

$$f(x, y) = \begin{cases} xy & \text{if } xy > 0 \\ 0 & \text{if } xy \leq 0. \end{cases}$$

At which points of $\mathbb{R}^2$ is this function continuous?

5. For the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

$$f(x, y) = \frac{x^2 y}{x^4 + y^2}$$

Show that f has limit 0 as $(x, y) \rightarrow (0, 0)$ along any straight line through the origin, but it does not have a limit as $(x, y) \rightarrow (0, 0)$ in $\mathbb{R}^2$.

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

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

   At which points of $\mathbb{R}^2$ is this function continuous?

7. Prove Theorem 8.1.7.

8. Prove Theorem 8.1.8.

9. Prove that a is a limit point of a set $D \subset \mathbb{R}^p$ if and only if there is a sequence of points in D but not equal to a which converges to a.

10. Let D be a subset of $\mathbb{R}^p$ and $F : D \to \mathbb{R}^q$ a function. If a is a limit point of D, prove that $\lim_{x \to a} F(x) = b$ if and only if $\lim_{n \to \infty} F(x_n) = b$ whenever $\{x_n\}$ is a sequence in D which converges to a.

11. Let $F : D \to \mathbb{R}^q$ be a transformation with domain $D \subset \mathbb{R}^p$ and let a be a limit point of D. Prove that if $\{F(x_n)\}$ converges whenever $\{x_n\}$ is a sequence in D which converges to a, then $\lim_{x \to a} F(x)$ exists.

12. Let $B_1(0)$ be the open unit ball in $\mathbb{R}^2$. Does every continuous function $f : B_1(0) \to \mathbb{R}$ take Cauchy sequences to Cauchy sequences?

13. Let $\overline{B}_1(0)$ be the closed unit ball in $\mathbb{R}^2$. Does every continuous function $f : \overline{B}_1(0) \to \mathbb{R}$ take Cauchy sequences to Cauchy sequences?

14. Find a parameterized curve $\gamma(t)$ in $\mathbb{R}^2$, with parameter interval $[0, \infty)$, that begins at $(1, 0)$, spirals inward in the counterclockwise direction, and approaches $(0, 0)$ as $t \to \infty$.

15. Find a parameterization of the cylindrical surface in $\mathbb{R}^3$ defined by the equation $x^2 + y^2 = 1$ (z is unrestricted). That is, find a continuous function $F : A \to \mathbb{R}^3$ with $A \subset \mathbb{R}^2$, such that F has the cylinder as image.

8.2 Properties of Continuous Functions

The theme of this section is that continuous functions are the functions that behave well with respect to topological properties of sets.

Continuity and Open and Closed Sets

Recall that if D is a subset of $\mathbb{R}^p$, then a *relatively open* subset of D is a set of the form $U \cap D$, where U is open in $\mathbb{R}^p$. The relatively open subsets of D are the open subsets of D considered as a metric space by itself (rather than a subset of $\mathbb{R}^p$). Relatively closed sets are defined analogously.

Theorem 8.2.1. *If $D \subset \mathbb{R}^p$ and $F : D \to \mathbb{R}^q$ is a function, then F is continuous on D if and only $F^{-1}(U)$ is a relatively open subset of D whenever U is an open subset of $\mathbb{R}^q$. Equivalently, F is continuous if and only if $F^{-1}(A)$ is a relatively closed subset of D whenever A is a closed subset of $\mathbb{R}^q$.*

Proof. Suppose F is continuous and U is an open subset of $\mathbb{R}^q$. If $a \in F^{-1}(U)$, then $b = F(a) \in U$. Since U is open, there is an $\epsilon > 0$ such that $B_\epsilon(b) \subset U$. Since F is continuous on D, there is a $\delta > 0$ such that

$$||F(x) - F(a)|| < \epsilon \quad \text{whenever} \quad x \in D \quad \text{and} \quad ||x - a|| < \delta.$$

This implies that $F(B_\delta(a) \cap D) \subset B_\epsilon(b) \subset U$, and , hence, that

$$B_\delta(a) \cap D \subset F^{-1}(U).$$

Since we can do this at each $a \in F^{-1}(U)$, we conclude that $F^{-1}(U)$ is the intersection of D with the union of the resulting collection of open balls $B_\delta(a)$. Hence, it is relatively open in D.

On the other hand, suppose $F^{-1}(U)$ is relatively open in D for each open set U in $\mathbb{R}^q$, In particular, this implies that if $a \in D$, $b = F(a)$, and $\epsilon > 0$, then the set $F^{-1}(B_\epsilon(b))$ is relatively open in D. Thus,

$$F^{-1}(B_\epsilon(b)) = D \cap V$$

for some open set $V \subset \mathbb{R}^p$. Since $a \in V$ and V is open, there is a $\delta > 0$ such that $B_\delta(a) \subset V$. Then $x \in D$ and $||x - a|| < \delta$ implies $x \in V \cap D = F^{-1}(B_\epsilon(b))$. This means that

$$||F(x) - F(a)|| < \epsilon \quad \text{whenever} \quad x \in D \quad \text{and} \quad ||x - a|| < \delta.$$

Hence, F is continuous at a. Since this is true for all points $a \in D$, we conclude that F is continuous on D.

The analogous result for closed sets follows from the above by taking complements and using the fact that a subset of D is relatively closed if and only if it is the complement in D of a set which is relatively open. The details are left to the exercises. $\square$

If D is open, then the relatively open subsets of D are just the open subsets of D. Hence, we have the following corollary of the above theorem.

Corollary 8.2.2. *If $D \subset \mathbb{R}^p$ is open and $F : D \to \mathbb{R}^q$ is a function, then F is continuous on D if and only if $F^{-1}(U)$ is open for every open set $U \subset \mathbb{R}^q$.*

Continuity and Compactness

The proof of the following theorem is very simple, but it has a lot of very useful consequences.

Theorem 8.2.3. *If K is a compact subset of $\mathbb{R}^p$ and $F : K \to \mathbb{R}^q$ is a continuous function, then $F(K)$ is a compact subset of $\mathbb{R}^q$.*

Proof. Let $\mathcal{U}$ be an open cover of $F(K)$ and let $\mathcal{V}$ be the collection of all open subsets $V \subset \mathbb{R}^p$ such that $V \cap K = F^{-1}(U)$ for some $U \in \mathcal{U}$. There is at least one such V for each $U \in \mathcal{U}$ since $F^{-1}(U)$ is relatively open in K by the previous theorem.

Since $\mathcal{U}$ is a cover of $F(K)$, $\mathcal{V}$ is an open cover of K. Since K is compact, there is a finite subcollection $\{V_j\}_{j=1}^n$ of $\mathcal{V}$ which also covers K. For each V_j we may choose a $U_j \in \mathcal{U}$ such that $V_j \cap K = F^{-1}(U_j)$.

If $y \in F(K)$, then $y = F(x)$ for some $x \in K$. This x belongs to $V_j \cap K$ for some j because $\{V_j\}_{j=1}^n$ is a cover of K. Then $y \in U_j$. This proves that the collection $\{U_j\}_{j=1}^n$ is a cover of $F(K)$. It is, in fact, a finite subcover of $\mathcal{U}$. Since we can do this for every open cover of $F(K)$, we have proved that $F(K)$ is compact. $\qquad\blacksquare$

A function $F : D \to \mathbb{R}^q$ is said to be *bounded* on D if there is a number M such that

$$||F(x)|| \leq M \quad \text{for all} \quad x \in D.$$

That is, F is bounded on D if the set of non-negative numbers $\{||F(x)|| : x \in D\}$ is bounded above. The least upper bound of this set is denoted $\sup_D ||F(x)||$. It may or may not be a member of the set – that is, there may or may not be a point $x_0 \in D$ such that $||F(x_0)|| = \sup_D ||F(x)||$. If there is such a point x_0, then we say that $||F(x)||$ assumes a maximum value on D.

A compact set contains points of maximal norm and points of minimal norm (Exercise 7.4.4). Combining this with the previous theorem yields the following:

Theorem 8.2.4. *If $K \subset \mathbb{R}^p$ is compact and $F : K \to \mathbb{R}^q$ is continuous, then F is bounded on K and $||F(x)||$ assumes a maximum value on K.*

Proof. By the previous theorem, $F(K)$ is compact and, hence, bounded. Furthermore, it contains a point of maximum norm by Exercise 7.4.4. This point is in $F(K)$ and so it the form $F(x_0)$ for some $x_0 \in K$. $\qquad\blacksquare$

Corollary 8.2.5. *If $K \subset \mathbb{R}^p$ is compact and $f : K \to \mathbb{R}$ is a continuous real valued function on K, then f assumes a maximal value and a minimal value on K.*

Proof. It follows from the previous theorem that $\{|f(x)| : x \in K\}$ is bounded above by some number M. Then the function $g(x) = f(x) + M$ is a non-negative function and so $|g(x)| = g(x)$. By the previous theorem, there is a point $x_0 \in K$ with

$$g(x) \leq g(x_0) \quad \text{for all} \quad x \in K.$$

Since $f(x) = g(x) - M$, it follows that x_0 is a point at which f achieves its maximal value.

Since the above argument applies equally well to $-f(x)$, and, since a maximum for $-f(x)$ on K will be the negative of a minimum for $f(x)$ on K, it follows that $f(x)$ has a minimum value on K as well. $\qquad\blacksquare$

Example 8.2.6. Let K be a compact subset of $\mathbb{R}^p$. Show that $f : K \to \mathbb{R}$ is a real valued continuous function on K which is strictly positive at each point of K, then there is a number $\delta > 0$ such that $f(x) \geq \delta$ for all $x \in K$.

Solution: By Corollary 8.2.5, the function f has a minimum value δ on K. This minimum value cannot be 0, since f is positive at all points of K. Thus, $\delta > 0$ and $f(x) \geq \delta$ for all $x \in K$.

Continuity and Connectedness

Continuous functions also take connected sets to connected sets.

Theorem 8.2.7. *If $D \subset \mathbb{R}^p$ is connected and $F : D \to \mathbb{R}^q$ is continuous, then $F(D)$ is also connected.*

Proof. Suppose U and V are open subsets of $\mathbb{R}^q$ such that $F(D) \subset U \cup V$ and $(U \cap F(D)) \cap (V \cap F(D)) = \emptyset$. Then $F^{-1}(U)$ and $F^{-1}(V)$ are relatively open subsets of D, $F^{-1}(U) \cap F^{-1}(V) = \emptyset$, and $D \subset F^{-1}(U) \cup F^{-1}(V)$. Thus, one of the sets $F^{-1}(U) \cap D$ and $F^{-1}(V) \cap D$ must be empty since, otherwise, they would separate D. However, if $F^{-1}(U) \cap D = \emptyset$, then $U \cap F(D) = \emptyset$ and a similar statement holds for V. Thus, either U or V has empty intersection with $F(D)$ which implies that the two sets do not separate $F(D)$. Hence, $F(D)$ is connected. $\square$

The following is the several variable version of the Intermediate Value Theorem, since it says that if a continuous real valued function on a connected set takes on two values, it also takes on every value in between the two.

Corollary 8.2.8. *If $D \subset \mathbb{R}^p$ is connected and $f : D \to \mathbb{R}$ is a continuous function, then $f(D)$ is an interval.*

Proof. By the previous theorem, $f(D)$ is a connected subset of the line $\mathbb{R}$. By Theorem 7.5.4 the only such sets are intervals. $\square$

Now suppose E is a subset of $\mathbb{R}^d$ and $\gamma : I \to E$ is a parameterized curve with parameter interval $I = [a, b]$. Since I is connected by Theorem 7.5.4, its image $\gamma(I)$ is a connected subset of E. Thus, if $x = \gamma(a)$ and $y = \gamma(b)$, then x and y must be in the same component of E. Thus, we have proved the following.

Theorem 8.2.9. *If E is a subset of $\mathbb{R}^d$ and x and y are points of E that may be joined by a curve in E, then x and y are in the same connected component of E. If each pair of points of E may be joined by a curve in E, then E is connected.*

Example 8.2.10. Show that the unit circle T (the set of points $(x, y) \in \mathbb{R}^2$ with $x^2 + y^2 = 1$) is connected.

Solution: Each point on the circle T is of the form $(\cos t, \sin t)$. Each pair of such points $(\cos a, \sin a)$ and $(\cos b, \sin b)$ with $a < b$, are joined by the curve

$$\gamma(t) = (\cos t, \sin t) \quad t \in [a, b]$$

which lies in the circle. Hence, the circle T is connected.

Uniform Continuity

Definition 8.2.11. Let D be a subset of $\mathbb{R}^p$ and $F : D \to \mathbb{R}^q$ a function. Then F is said to be *uniformly continuous* on D if for each $\epsilon > 0$ there is a $\delta > 0$ such that

$$||F(x) - F(y)|| < \epsilon \quad \text{whenever} \quad x, y \in D \quad \text{and} \quad ||x - y|| < \delta.$$

As with uniform continuity for functions of one variable, discussed in Section 3.3, the point here is that the choice of δ does not depend on x or y.

Uniform continuity is an important concept and it will play a key role in our proof of the existence of the Riemann integral of a function of several variables.

We proved in Theorem 3.3.4 that a continuous function on closed, bounded interval is uniformly continuous. The analogous theorem holds for functions of several variables, but compact sets replace closed, bounded intervals.

Theorem 8.2.12. *If K is a compact subset of $\mathbb{R}^p$ and $F : K \to \mathbb{R}^q$ is continuous on K, then F is uniformly continuous on K.*

Proof. Since F is continuous on K, given $\epsilon > 0$ we may choose for each $x \in K$ a number $\delta(x) > 0$ such that

$$||F(y) - F(x)|| < \epsilon/2 \quad \text{whenever} \quad y \in K \quad \text{and} \quad ||y - x|| < \delta(x). \quad (8.2.1)$$

We set $\rho(x) = \delta(x)/2$. Then $\rho(x)$ is a positive valued function defined on K, just as in Example 7.4.9. In that example, we showed that a consequence of the compactness of K is that there is a finite set of points $\{x_1, x_2, \cdots, x_n\}$ such that K is contained in the union of the balls $B_{\rho(x_j)}(x_j)$ for $j = 1, \cdots n$.

We set $\rho = \min\{\rho(x_j) : j = 1, \cdots, n\}$. Then given any two points $x, y \in K$ with $||x - y|| < \rho$, x must be in $B_{\rho(x_j)}(x_j)$ for some j. This implies that $||x - x_j|| < \rho(x_j) < \delta(x_j)$ and

$$||y - x_j|| \le ||y - x|| + ||x - x_j|| < \rho + \rho(x_j) \le 2\rho(x_j) = \delta(x_j).$$

Since both x and y are within $\delta(x_j)$ of x_j, it follows from (8.2.1) that

$$||F(x) - F(y)|| \le ||F(x) - F(x_j)|| + ||F(x_j) - F(y)|| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Hence, F is uniformly continuous on K. $\qquad\qquad\square$

In Theorem 3.3.6 we showed that a function is uniformly continuous on a bounded interval if and only if it has a continuous extension to the closure of the interval. The analogous theorem holds for functions from $\mathbb{R}^p$ to $\mathbb{R}^q$.

Theorem 8.2.13. *If $D \subset \mathbb{R}^p$ is a bounded set and $F : D \to \mathbb{R}^q$ is a function, then F is uniformly continuous on D if and only if F can be extended to a continuous function $\hat{F} : \overline{D} \to \mathbb{R}^q$.*

Proof. Note that, since D is bounded, $\overline{D}$ is compact. Thus, if F has an extension to a continuous function $\hat{F} : \overline{D} \to \mathbb{R}^q$, then $\hat{F}$ is uniformly continuous on $\overline{D}$, by the previous theorem. Then $\hat{F}$ is also uniformly continuous on the smaller set D. But $\hat{F} = F$ on D, and so F is uniformly continuous on D.

Conversely, suppose F is uniformly continuous on D. Then $\{F(x_n)\}$ is a Cauchy sequence in $\mathbb{R}^q$ whenever $\{x_n\}$ is a Cauchy sequence in D (Exercise 8.2.11). If $x \in \overline{D}$, then there is a sequence $\{x_n\}$ in D that converges to x (Theorem 7.3.10). Such a sequence is necessarily Cauchy and so $\{F(x_n)\}$ is also Cauchy. But Cauchy sequences in $\mathbb{R}^q$ converge by Theorem 7.2.16.

If $\{y_n\}$ is another sequence in D which converges to x, then we may construct a third sequence $\{z_n\}$ converging to x by intertwining the sequences $\{x_n\}$ and $\{y_n\}$ – that is, let $z_{2n} = y_n$ and $z_{2n-1} = x_n$. Then, $\{z_n\}$ not only converges to x, it has both $\{x_n\}$ and $\{y_n\}$ as subsequences. By the above argument, the sequence $\{F(z_n)\}$ must converge to a point $u \in \mathbb{R}^q$. Both subsequences $\{F(x_n)\}$ and $\{F(y_n)\}$ must then converge to the same point u. Thus, we have proved that no matter what sequence $\{x_n\}$ converging to x we choose, the limit of the sequence $\{F(x_n)\}$ is the same. Therefore, it makes sense to define an extension $\hat{F}$ of F to $\overline{D}$ by setting

$$\hat{F}(x) = \lim F(x_n)$$

for any sequence $\{x_n\}$ in D converging to x. The resulting function is obviously equal to F on D, since we may just choose $x_n = x$ for all n if $x \in D$.

We now have an extension $\hat{F}$ of F to $\overline{D}$. It remains to prove that it is continuous on $\overline{D}$. We will do this by applying Theorem 8.1.6. If $\{x_n\}$ is a sequence in $\overline{D}$ which converges to $x \in \overline{D}$, we may choose for each n a point $y_n \in D$ such that $\|x_n - y_n\| < 1/n$ and $\|F(y_n) - \hat{F}(x_n)\| < 1/n$. Then

$$\|x - y_n\| \leq \|x - x_n\| + \|x_n - y_n\| < \|x - x_n\| + 1/n.$$

Since $\|x - x_n\| \to 0$ and $1/n \to 0$, it follows that $y_n \to x$ and, hence, $F(y_n) \to \hat{F}(x)$ by our definition of $\hat{F}$. However, it also follows that $\hat{F}(x_n) \to \hat{F}(x)$ since,

$$\|\hat{F}(x) - \hat{F}(x_n)\| \leq \|\hat{F}(x) - F(y_n)\| + \| F(y_n) - \hat{F}(x_n)\|,$$

and both $\|F(y_n) - \hat{F}(x_n)\|$ and $\|\hat{F}(x) - F(y_n)\|$ converge to 0.

Since $\hat{F}(x_n) \to \hat{F}(x)$ whenever $\{x_n\}$ is a sequence in $\overline{D}$ converging to $x \in \overline{D}$, the function $\hat{F}$ is continuous on $\overline{D}$ by Theorem 8.1.6. $\square$

Exercise Set 8.2

1. If $A = \{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, 0 \leq y \leq 1\}$. Which of the following sets cannot be the image of the set A under a continuous function $F : A \to \mathbb{R}^2$? Justify your answers.

   (a) $\overline{B}_2(0, 0)$;

   (b) $B_1(0)$;

   (c) $\{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 1, \ 0 \leq y\}$;

    (d) $\overline{B}_1(0,0) \cup \overline{B}_1(3,0)$.

    (e) $\{(t,t) \in \mathbb{R}^2 : t \in \mathbb{R}; \ 0 \le t \le 1\}$.

2. Finish the proof of Theorem 8.2.1, by proving that a function is continuous if and only if the inverse image of each closed set is relatively closed. Hint: you may use the first part of the theorem (that a function is continuous if and only if the inverse image of each open set is relatively open).

3. If K is a compact, connected subset of $\mathbb{R}^p$ and $f : K \to \mathbb{R}$ is a continuous function, what can you say about $f(K)$?

4. If $F : \mathbb{R}^p \to \mathbb{R}^q$ is continuous and A is a bounded subset of $\mathbb{R}^p$, prove that $\overline{F(A)} = F(\overline{A})$. Is this necessarily true if A is not bounded?

5. The image of a compact set under a continuous function is compact, hence closed, by Theorem 8.2.3. Is the image of a closed set under a continuous function necessarily closed? Prove that it is or give an example where it is not.

6. Is the image of an open set under a continuous function necessarily an open set? Prove that it is or give an example where it is not.

7. Is the sphere $\{(x,y,z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$ connected? How do you know?

8. Prove that if $f : T \to \mathbb{R}$ is a continuous real valued function on the unit circle $T = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$, then there is a pair of diametrically opposed points (x,y) and $(-x,-y)$ on T at which f has the same value.

9. Find an example of a closed set $A \subset \mathbb{R}^2$, which is connected, but which contains two points that cannot be joined by a curve in A.

10. Is the function $f : \mathbb{R}^2 \setminus \{(2,0)\} \to \mathbb{R}$ defined by

$$f(x,y) = \frac{1}{(x-2)^2 + y^2}$$

uniformly continuous on $B_1(0,0)$? Is it uniformly continuous on $B_2(0,0)$? Justify your answers.

11. If $D \subset \mathbb{R}^p$, prove that if a function $F : D \to \mathbb{R}^q$ is uniformly continuous on D then $\{F(x_n)\}$ is a Cauchy sequence in $\mathbb{R}^q$ whenever $\{x_n\}$ is a Cauchy sequence in D.

12. Show that the converse of the statement in the previous exercise is not true in general, but it is true if the set D is bounded. That is, show that there exist a D and a continuous function $F : D \to \mathbb{R}^q$ which is not uniformly continuous but which does take each Cauchy sequence in D to a Cauchy sequence in $\mathbb{R}^d$. However, show there are no such functions if D is bounded.

13. Does uniform continuity make sense for a function from one metric space to another? If so, how would you define it?

8.3 Sequences of Functions

Uniform convergence of sequences of functions will play the same role in functions of several variables that it did in earlier chapters on functions of a single variable. It preserves continuity and allows the limit to be taken inside an integral.

The results of Section 3.4 on uniform convergence hold in the several variable context and have almost the same proofs.

Uniform convergence

Definition 8.3.1. Let $\{F_n\}$ be a sequence of functions from D to $\mathbb{R}^q$, where $D \subset \mathbb{R}^p$. We say this sequence converges pointwise to $F : D \to \mathbb{R}^q$ on D if the sequence $\{F_n(x)\}$ converges to $F(x)$ for each $x \in D$.

We say $\{F_n\}$ converges *uniformly* to $F : D \to \mathbb{R}^q$ on D if, for each $\epsilon > 0$, there is an N such that

$$||F(x) - F_n(x)|| < \epsilon \quad \text{whenever} \quad x \in D \quad \text{and} \quad n \geq N.$$

The difference between pointwise and uniform convergence is that, in the latter, the choice of N must be independent of x.

The following test for uniform convergence is the several variable analogue of Theorem 3.4.6. The proof is simple and is left to the exercises.

Theorem 8.3.2. *Let F be a function and $\{F_n\}$ a sequence of functions defined on a set $D \subset \mathbb{R}^p$ and having values in $\mathbb{R}^q$. If there is a sequence of non-negative numbers $\{b_n\}$, such that $b_n \to 0$, and*

$$||F(x) - F_n(x)|| \leq b_n \quad \text{for all} \quad x \in D,$$

then $\{F_n\}$ converges uniformly to F on D.

Example 8.3.3. Examine the convergence of the sequence $\{(x^2 + y^2)^n\}$ on the closed disc $\overline{B}_r(0,0)$ in $\mathbb{R}^2$ for each $r \leq 1$.
    Solution: Note that $x^2 + y^2 \leq r^2$ on $\overline{B}_r(0,0)$. Thus,

$$|(x^2 + y^2)^n| \leq r^{2n} \quad \text{on} \quad \overline{B}_r(0,0).$$

If $r < 1$, then $r^{2n} \to 0$ and, hence, $\{(x^2 + y^2)^n\}$ converges uniformly to 0 on $\overline{B}_r(0,0)$ by the previous theorem.

On $\overline{B}_1(0,0)$, the sequence $\{(x^2 + y^2)^n\}$ converges to 0 if (x,y) is in the interior of the disc and to 1 if (x,y) is on the boundary of the disc. The limit function is not continuous on $\overline{B}_1(0,0)$ and, by the next theorem, this means the convergence is not uniform. Without using the next theorem, we can still

easily see that the convergence is not uniform – in fact, not uniform even on the smaller set $B_1(0,0)$. Given an ϵ with $0 < \epsilon < 1$, if $(x,y) \in B_1(0,0)$ and we set $r = ||(x,y)|| < 1$, then $|(x^2 + y^2)^n| = r^{2n}$ and so

$$|(x^2 + y^2)^n| < \epsilon \qquad (8.3.1)$$

if and only $r^{2n} < \epsilon$, which holds if and only if

$$n > N_r = \frac{\ln \epsilon}{2 \ln r}.$$

Thus, an N with the property that (8.3.1) holds for all $r < 1$ must be larger than N_r for all $r < 1$. There is no such N, since $\lim_{r \to 1} N_r = \infty$.

Uniform Convergence and Continuity

One of the main reasons uniform convergence is important is the following theorem. Its proof is the same as the proof of the analogous theorem for real valued functions of a real variable (Theorem 3.4.4), and we will not repeat it.

Theorem 8.3.4. *If $\{F_n\}$ is a sequence of continuous functions from a subset D of $\mathbb{R}^p$ to $\mathbb{R}^q$, which converges uniformly on D to a function F, then F is also continuous on D.*

As we saw in example 8.3.3, a sequence of continuous functions which converges only pointwise may not converge to a continuous function.

Example 8.3.5. Define a sequence $\{F_n\}$ of functions from the unit ball $B_1(0,0)$ in $\mathbb{R}^2$ to $\mathbb{R}^2$ by

$$F_n(x,y) = \left(\frac{x^2 - ny^2}{1 + ny^2}, \frac{nx}{1 + nx^2} \right).$$

Show that this sequence converges pointwise, but not uniformly on $B_1(0,0)$.

Solution: Each of the functions F_n is continuous on $B_1(0,0)$. The sequence clearly converges pointwise to the function F defined on $B_1(0,0)$ by

$$F(x) = \begin{cases} (-1, 1/x) & \text{if} \quad x \neq 0, \ y \neq 0 \\ (-1, 0) & \text{if} \quad x = 0, \ y \neq 0 \\ (x^2, 1/x) & \text{if} \quad x \neq 0, \ y = 0 \\ (0, 0) & \text{if} \quad x = 0, \ y = 0 \end{cases}$$

This function is not continuous on $B_1(0,0)$ – in fact, it is discontinuous at all points on the x and y axes – and so, by the previous theorem, the convergence of $\{F_n\}$ to F cannot be uniform on $B_1(0,0)$.

Uniformly Cauchy Sequences

Definition 8.3.6. If $D \subset \mathbb{R}^p$ and $\{F_n\}$ is a sequence of functions from D to $\mathbb{R}^q$, then $\{F^n\}$ is said to be *uniformly Cauchy* if, for each $\epsilon > 0$, there is an N such that

$$||F_n(x) - F_m(x)|| < \epsilon \quad \text{whenever} \quad x \in D \text{ and } n, m \geq N.$$

Another several variable analogue of a singe variable theorem (Theorem 3.4.10) is the following. Since the proof of the single variable version was left to the exercises, we will actually prove this version.

Theorem 8.3.7. *If $D \subset \mathbb{R}^p$, a sequence of functions $F_n : D \to \mathbb{R}^q$ is uniformly Cauchy if and only if it converges uniformly to some function $F : D \to \mathbb{R}^q$.*

Proof. If $\{F_n\}$ converges uniformly on D to a function F and $\epsilon > 0$, then there is an N such that

$$||F(x) - F_n(x)|| < \epsilon/2 \quad \text{whenever} \quad x \in D, \ n \geq N.$$

Then

$$||F_n(x) - F_m(x)|| \leq ||F_n(x) - F(x)|| + ||F(x) - F_m(x)|| < \epsilon/2 + \epsilon/2 = \epsilon$$

whenever $x \in D$ and $n, m \geq N$. Thus, $\{F_n\}$ is uniformly Cauchy.

On the other hand, if $\{F_n\}$ is uniformly Cauchy, then for each $x \in D$, $\{F_n(x)\}$ is a Cauchy sequence of vectors in $\mathbb{R}^q$ and, hence, converges to some vector $F(x) \in \mathbb{R}^q$ by Theorem 7.2.16. That is, $\{F_n\}$ converges pointwise to a function $F : D \to \mathbb{R}^q$. It remains to prove that the convergence is uniform.

Since the sequence is uniformly Cauchy, for each $\epsilon > 0$ there is an N such that

$$||F_n(x) - F_m(x)|| < \epsilon/2 \quad \text{whenever} \quad x \in D \text{ and } n, m \geq N.$$

If $m > n \geq N$ we have

$$||F(x) - F_n(x)|| \leq ||F(x) - F_m(x)|| + ||F_m(x) - F_n(x)|| < ||F_m(x) - F(x)|| + \epsilon/2.$$

The left side of this inequality does not depend on m and the right side holds for all $m > n$. For each $x \in D$, $\lim ||F(x) - F_m(x)|| = 0$. Hence, on taking the limit of the above inequality as $m \to \infty$, we conclude that

$$||F(x) - F_n(x)|| \leq \epsilon/2 < \epsilon \quad \text{for all} \quad x \in D \text{ and } n \geq N.$$

This proves that $\{F_n\}$ converges uniformly to F on D. $\square$

The Sup Norm

If D is a compact subset of $\mathbb{R}^p$, each continuous function F from D to $\mathbb{R}^q$ is bounded, by Theorem 8.2.4. That is, $\sup_D ||F(x)||$ is finite and, in fact, $||F(x)||$ actually assumes this value at some point of D. We set,

$$||F||_D = \sup_D ||F(x)||.$$

This is a norm on the vector space of all continuous functions from D to $\mathbb{R}^q$.

Example 8.3.8. Find $||\gamma||_I$ if I is the interval $[0, \pi]$ and $\gamma : I \to \mathbb{R}^2$ is the curve defined by

$$\gamma(t) = (\cos t, 1 + \sin t).$$

We have

$$||\gamma(t)|| = \sqrt{\cos^2 t + (1 + \sin t)^2} = \sqrt{2 + 2\sin t}.$$

This attains its maximum value on $[0, \pi]$ at $t = \pi/2$, where it has the value 2, Thus, $||\gamma||_I = 2$.

Theorem 8.3.9. *If D is a compact subset of $\mathbb{R}^p$ and $\{F_n\}$ is a sequence of continuous functions from D to $\mathbb{R}^q$, then $\{F_n\}$ converges uniformly to a function $F : D \to \mathbb{R}^q$ if and only if $\lim_{n\to\infty} ||F - F_n||_D = 0$.*

Proof. Given any $\epsilon > 0$ and any n, the inequality $||F(x) - F_n(x)|| < \epsilon$ holds for all $x \in D$ if and only if $||F - F_n||_D < \epsilon$. Thus, $\{F_n\}$ converges uniformly to F if and only if $\lim_{n\to\infty} ||F - F_n||_D = 0$. $\qquad\qquad\qquad\qquad\qquad\square$

The space, $\mathcal{C}(K; \mathbb{R}^q)$, of all continuous functions on a compact set $K \subset \mathbb{R}^p$, with values in $\mathbb{R}^q$ is a vector space under the operations of pointwise addition and scalar multiplication of functions. If we define the norm of an element F of this space to be the Sup norm $||F||_K$, then it is easy to see that $\mathcal{C}(K; \mathbb{R}^q)$ is a normed vector space (Exercise 8.3.11). In particular, it is a metric space in which the distance between two elements F and G is defined to be $||F - G||_K$. It turns out that this is a complete metric space (meaning that all Cauchy sequences converge).

Theorem 8.3.10. *The normed vector space $\mathcal{C}(K; \mathbb{R}^q)$ is complete.*

Proof. A Cauchy sequence in $\mathcal{C}(K; \mathbb{R}^q)$, is by definition a sequence of continuous functions which is Cauchy in the metric defined by the norm $|| \cdot ||_K$. Such a sequence is uniformly Cauchy on K. By Theorem 8.3.7 such a sequence converges uniformly on K. The limit function is continuous, by 8.3.4. By the previous theorem, the sequence converges in the metric defined by $|| \cdot ||_K$ to this limit. Thus, each Cauchy sequence in the metric space $\mathcal{C}(K; \mathbb{R}^q)$ converges to an element of $\mathcal{C}(K; \mathbb{R}^q)$ and, hence, this space is complete. $\qquad\square$

Series of Functions

Given a series

$$\sum_{k=1}^{\infty} F_k(x) \qquad\qquad\qquad (8.3.2)$$

whose terms F_k are functions from a domain $D \subset \mathbb{R}^p$ into $\mathbb{R}^q$, we define its associated sequence of partial sums $\{S_n\}$ in the usual way:

$$S_n(x) = \sum_{k=1}^{n} F_k(x).$$

The series converges pointwise if its sequence of partial sums converges pointwise, It converges uniformly on D if its sequence of partial sums converges uniformly on D.

As in the single variable case, there is a simple condition (the Weierstrass M-test) which ensures that a series converges uniformly. The proof is the same as the proof of Theorem 6.4.4 and so we will not repeat it.

Theorem 8.3.11. (Weierstrass M-test) *If there is a convergent series of non-negative numbers*

$$\sum_{k=1}^{\infty} M_k,$$

such that $||F_k(x)|| \leq M_k$ for all k and all $x \in D$, then the series (8.3.2) converges uniformly on D.

Example 8.3.12. Show that the series

$$\sum_{k=1}^{\infty} \frac{1}{k^2} \sin kx \cos ky \qquad\qquad (8.3.3)$$

converges uniformly on $\mathbb{R}^2$.

Solution: Since

$$\left| \frac{1}{k^2} \sin kx \cos ky \right| \leq \frac{1}{k^2} \quad \text{for all} \quad k, x, y,$$

and the series $\sum_{k=1}^{\infty} 1/k^2$ converges (it's a p-series with $p = 2$), the Weierstrass M-test tells us that the series (8.3.3) converges uniformly on $\mathbb{R}^2$.

Exercise Set 8.3

1. Show that the sequence $\{\gamma_n(t)\}$, where

$$\gamma_n(t) = \left(\frac{1}{1+nt}, \frac{t}{n} \right)$$

   does not converge uniformly on $[0, 1]$.

2. Show that the sequence $\{\lambda_n(t)\}$, where

$$\lambda_n(t) = \left(\frac{t}{1+nt}, \frac{t}{n} \right)$$

   does converge uniformly on $[0, 1]$.

3. Does the sequence $\{(k^{-1} \sin kx, k^{-1} \cos ky)\}$ converge pointwise on $\mathbb{R}^2$? Does it converge uniformly on $\mathbb{R}^2$? Justify your answers.

4. Does the sequence $\{\sin(x/k), \cos(y/k)\}$ converge pointwise on $\mathbb{R}^2$? Does it converge uniformly on $\mathbb{R}^2$? Justify your answer.

5. Find $||F||_D$ if $D = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ and $F : \mathbb{R}^2 \to \mathbb{R}^2$ is defined by

$$F(x, y) = (x + 1, y + 1).$$

6. Find $||\gamma||_I$ if $I = [0, \pi]$ and $\gamma : I \to \mathbb{R}^2$ is defined by

$$\gamma(t) = (2 \cos t, 3 \sin t).$$

7. Prove that if $\{F_n\}$ is a sequence of bounded functions from a set $D \subset \mathbb{R}^p$ into $\mathbb{R}^q$ and if $\{F_n\}$ converges uniformly to F on D, then F is also bounded.

8. Does the series $\sum_{k=0}^{\infty} x^k y^k$ converge uniformly on the square

$$\{(x, y) \in \mathbb{R}^2 : -1 < x < 1, -1 < y < 1\}?$$

Justify your answer.

9. Does the series $\sum_{k=0}^{\infty} x^k y^k$ converge uniformly on the disc

$$\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}?$$

Justify your answer.

10. Does the series $\sum_{k=0}^{\infty} (x^n, (1 - x)^n)$ converge pointwise on $[0, 1]$? Does it converge pointwise on $(0, 1)$? On which subsets of $(0, 1)$ does it converge uniformly? Justify your answers.

11. If K is a compact subset of $\mathbb{R}^p$, show that $|| \cdot ||_K$ is a norm on the vector space $\mathcal{C}(K; \mathbb{R}^q)$ of continuous functions on K with values in $\mathbb{R}^q$.

12. Prove that if D is a subset of $\mathbb{R}^p$ and $\{F_n\}$ is a sequence of functions from D to $\mathbb{R}^q$, then $\{F_n\}$ fails to converge uniformly to 0 if and only if there is a sequence $\{x_n\}$ in D such that the sequence of numbers $\{F_n(x_n)\}$ does not converge to 0.

13. If $K \subset \mathbb{R}^q$ is compact, show that a series $\sum_{k=1}^{\infty} F_k(x)$ of functions from K to $\mathbb{R}^q$ converges uniformly on K if the series of numbers $\sum_{k=1}^{\infty} ||F_k(x)||_K$ converges.

8.4   Linear Functions, Matrices

Other than constants, linear functions are the simplest functions from $\mathbb{R}^p$ to $\mathbb{R}^q$. For example, the linear functions from $\mathbb{R}$ to $\mathbb{R}$ are the functions of the form

$$L(x) = mx,$$

where m is a constant – that is, they are functions whose graphs are straight lines through the origin. In this section we introduce and study linear functions between Euclidean spaces. In the next chapter we will show how to use linear functions to approximate more complicated functions.

Linear Functions

Definition 8.4.1. A function $L : \mathbb{R}^p \to \mathbb{R}^q$ is said to be *linear* if, whenever $x, y \in \mathbb{R}^p$ and $a \in \mathbb{R}$,

(a) $L(x + y) = L(x) + L(y)$; and

(b) $L(ax) = aL(x)$.

Linear functions are often called *linear transformations* or *linear operators*.

Combining (a) and (b) of this definition we see that a linear function preserves linear combinations of vectors. That is,

$$L(ax + by) = aL(x) + bL(y) \tag{8.4.1}$$

for all pairs of vectors $x, y \in \mathbb{R}^p$ and all pairs of scalars a, b. An induction argument shows that the analogous result holds for linear combinations of more than two vectors.

Note that, since the definition uses only addition and scalar multiplication, linear functions between any two vector spaces may be defined in the same way as linear functions between $\mathbb{R}^p$ and $\mathbb{R}^q$.

Example 8.4.2. Determine whether the functions F, G from $\mathbb{R}^2$ to $\mathbb{R}^2$ and H from R^2 to $\mathbb{R}$ are linear, where

$$F(x, y) = (2x + y, x - y),$$
$$G(x, y) = (x^2, x + y),$$
$$H(x, y) = \begin{cases} \dfrac{x^3 + y^3}{x^2 + y^2} & \text{if} \quad (x, y) \neq (0, 0) \\ 0 & \text{if} \quad (x, y) = (0, 0) \end{cases}$$

Solution: The function F is linear since, given two vectors $u = (x_1, y_1)$ and $v = (x_2, y_2)$ in $\mathbb{R}^2$ and a scalar a, we have:

$$\begin{aligned} F(u + v) &= F(x_1 + x_2, y_1 + y_2) \\ &= (2(x_1 + x_2) + (y_1 + y_2), (x_1 + x_2) - (y_1 + y_2)) \\ &= ((2x_1 + y_1) + (2x_2 + y_2), (x_1 - y_1) + (x_2 - y_2)) = F(u) + F(v) \end{aligned}$$

and

$$\begin{aligned} F(au) = F(ax_1, ay_1) &= (2(ax_1) + ay_1, ax_1 - ay_1) \\ &= (a(2x_1 + y_1), a(x_1 - y_1)) = aF(u). \end{aligned}$$

The function G is not linear since, if $u = (1, 0)$, then

$$G(2u) = ((2)^2, 2) = (4, 2),$$

while

$$2G(u) = 2(1^2, 1) = (2, 2).$$

These are not equal and so (b) of the above definition does not hold for G.

The function H is also not linear. If $u = (1, 0)$ and $v = (0, 1)$, then

$$H(u) = H(v) = H(u + v) = 1.$$

Thus, $H(u + v) \neq H(u) + H(v)$ and (a) of the definition does not hold (note that (b) does hold for this function).

Linear Functions and Matrices

Recall that each vector $x \in \mathbb{R}^p$ may be written as a linear combination of the vectors e_j, where

$$e_j = (0, \cdots, 0, 1, 0, \cdots, 0)$$

with the 1 in the jth place. Specifically,

$$x = \sum_{j=1}^{p} x_j \, e_j \tag{8.4.2}$$

where x_j is the jth component of the vector x.

If we apply a linear function $L : \mathbb{R}^p \to \mathbb{R}^q$ to the vector x and use the fact that linear functions preserve linear combinations, we conclude that

$$L(x) = \sum_{k=1}^{p} x_j L(e_j).$$

The vector $L(e_j) \in \mathbb{R}^q$ has ith component $e_i \cdot L(e_j)$. If we set

$$a_{ij} = e_i \cdot L(e_j), \tag{8.4.3}$$

then the ith component y_i of the vector $y = L(x)$ is

$$y_i = \sum_{j=1}^{p} a_{ij} x_j. \tag{8.4.4}$$

The numbers (a_{ij}), appearing in (8.4.4), form a $q \times p$ matrix – that is a rectangular array

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ a_{q1} & a_{q2} & \cdots & a_{qp} \end{pmatrix}$$

with q rows and p columns. The equation $y = L(x)$ can be expressed in vector–matrix notation as

$$
\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_q \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ a_{q1} & a_{q2} & \cdots & a_{qp} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ x_p. \end{pmatrix}.
\tag{8.4.5}
$$

In this notation, the vectors x and y are written as column vectors. The expression on the right is the vector–matrix product of the matrix $A = (a_{ij})$ and the vector $x = (x_j)$. It is defined to be the vector whose ith component is the inner product of the ith row of A with the vector x.

At this point, we have shown that, to each linear function $L : \mathbb{R}^p \to \mathbb{R}^q$, there corresponds a $q \times p$ matrix A such that

$$
L(x) = Ax,
$$

where Ax is the vector–matrix product of A with x, as in (8.4.5). On the other hand, each $q \times p$ matrix A determines a linear function in this way, since vector matrix multiplication satisfies

$$
A(x + y) = Ax + Ay \quad \text{and} \quad A(cx) = c(Ax),
$$

for every pair of vectors $x, y \in \mathbb{R}^p$ and every scalar $c \in \mathbb{R}$ (Exercise 8.4.11).

Note that, in the correspondence between a linear function L and its matrix A, the jth column of A is the vector $L(e_j)$. The following theorem summarizes the above discussion.

Theorem 8.4.3. *A function $L : \mathbb{R}^p \to \mathbb{R}^q$ is linear if and only if there is a $q \times p$ matrix A such that*

$$
L(x) = Ax \quad \text{for all} \quad x \in \mathbb{R}^p.
$$

Example 8.4.4. If a function L from $\mathbb{R}^3$ to $\mathbb{R}^3$ is defined by

$$
L(x, y, z) = (x + 2y - z, y + z, 3x - y + z),
$$

then is L linear? If so, what matrix represents it?

Solution: If we write $L(x, y, z)$ as a column vector, then it clearly is given by

$$
L(x, y, z) = \begin{pmatrix} x + 2y - z \\ y + z \\ 3x - y + z \end{pmatrix} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \\ 3 & -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.
$$

Since L is given by a matrix through vector–matrix multiplication, it is linear by Theorem 8.4.3.

Matrix Operations

The sum of two linear functions $L : \mathbb{R}^p \to \mathbb{R}^q$ and $M : \mathbb{R}^p \to \mathbb{R}^q$ is defined pointwise, as is the sum of any two functions with a common domain. That is, $(L + M)(x) = L(x) + M(x)$. The function $L + M$ is also a linear function since

$$(L + M)(x + y) = L(x + y) + M(x + y)$$
$$= L(x) + L(y) + M(x) + M(y) = (L + M)(x) + (L + M)(y),$$

for all $x, y \in \mathbb{R}^p$, and

$$(L + M)(ax) = L(ax) + M(ax) = aL(x) + aM(x) = a(L + M)(x),$$

for all $x \in \mathbb{R}^p$ and $a \in \mathbb{R}$.

Similarly, the product of a scalar c with a linear function L is defined by $(cL)(x) = cL(x)$. This is also, clearly, a linear function.

If $M : \mathbb{R}^p \to \mathbb{R}^q$ and $L : \mathbb{R}^q \to \mathbb{R}^s$ are linear functions, then the composition $L \circ M : \mathbb{R}^p \to \mathbb{R}^s$ is defined, where

$$L \circ M(x) = L(M(x)).$$

This is also a linear function, since

$$(L \circ M)(x + y) = L(M(x + y)) = L(M(x) + M(y))$$
$$= L(M(x)) + L(M(y)) = L \circ M(x) + L \circ M(y).$$

for all $x, y \in \mathbb{R}^q$, and

$$L \circ M(ax) = L(M(ax)) = L(aM(x)) = aL(M(x)) = aL \circ M(x),$$

for all $x \in \mathbb{R}^q$ and all $a \in \mathbb{R}$.

In view of the above, it is natural to ask, for linear functions L and M represented by matrices A and B, what are the matrices representing $L + M$, cL, and $M \circ L$? The answer is given in the next two theorems. They have simple proofs based on the fact that, if the matrix A represents the linear function L, then the jth row of A is $L(e_j)$ (this is just equation (8.4.3)). The details are left to the exercises.

Theorem 8.4.5. *If $L : \mathbb{R}^p \to \mathbb{R}^q$ and $M : \mathbb{R}^p \to \mathbb{R}^q$ are linear functions represented by matrices $A = (a_{ij})$ and $B = (b_{ij})$, respectively, and $c \in \mathbb{R}$, then $L + M$ and cL are represented by the matrices*

$$A + B = (a_{ij} + b_{ij}) \quad and \quad cA = (ca_{ij}).$$

These are the usual operations of addition and scalar multiplication of matrices. The entry in the ith row and jth column of $A + B$ is $a_{ij} + b_{ij}$, while that of cA is ca_{ij}.

Theorem 8.4.6. *If $L : \mathbb{R}^q \to \mathbb{R}^s$ and $M : \mathbb{R}^p \to \mathbb{R}^q$ are linear functions represented by matrices $A = (a_{ij})$ and $B = (b_{jk})$, then $L \circ M : \mathbb{R}^p \to \mathbb{R}^s$ is represented by the matrix $AB = (c_{ik})$, where*

$$c_{ik} = \sum_{j=1}^{q} a_{ij} b_{jk}.$$

This is the usual operation of matrix multiplication. The entry in the ith row and kth column of AB is the inner product of the ith row of A with the kth column of B.

Example 8.4.7. If $A = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 & 3 \\ 1 & 0 & 1 \end{pmatrix}$, then find $2A - B$.

Solution: We have

$$2A - B = \begin{pmatrix} 2 - 0 & 4 - 1 & -2 - 3 \\ 0 - 1 & 2 - 0 & 2 - 1 \end{pmatrix} = \begin{pmatrix} 2 & 3 & -5 \\ -1 & 2 & 1 \end{pmatrix}$$

The transpose A^t of a matrix A is the matrix obtained by interchanging the rows and columns of A. That is. If $A = (a_{ij})$, then $A^t = (b_{ij})$, where $b_{ij} = a_{ji}$.

Example 8.4.8. If A is the matrix of the previous example, then find A^t, AA^t and $A^t A$.

Solution: By definition, we have

$$A^t = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ -1 & 1 \end{pmatrix},$$

while

$$AA^t = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 6 & 1 \\ 1 & 2 \end{pmatrix}$$

and

$$A^t A = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 5 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

Norm of a Linear Transformation

Definition 8.4.9. A linear transformation L from a normed vector space X to a normed vector space Y is said to be *bounded* if the set

$$\left\{ \frac{||L(x)||}{||x||} : x \in X, x \neq 0 \right\} \tag{8.4.6}$$

is bounded above. In this case, the least upper bound of this set is called the *operator norm* of L and is denoted $||L||$.

Equivalently, a linear transformation L is bounded if there is a number B such that

$$||L(x)|| \leq B||x|| \quad \text{for all} \quad x \in X.$$

The operator norm $||L||$ of L is the least such number B.

Theorem 8.4.10. *If X and Y are normed vector spaces, then every bounded linear transformation $L : X \to Y$ is uniformly continuous on X.*

Proof. If $x_1, x_2 \in X$, then

$$||L(x_1) - L(x_2)|| = ||L(x_1 - x_2)|| \leq ||L||\,||x_1 - x_2||.$$

Hence, given $\epsilon > 0$, if we choose $\delta = \epsilon/||L||$, then

$$||L(x_1) - L(x_2)|| \leq ||L||\,||x_1 - x_2|| < \epsilon \quad \text{whenever} \quad ||x_1 - x_2|| < \delta.$$

This shows that L is uniformly continuous on X. $\qquad\square$

Theorem 8.4.11. *Every linear transformation from $L : \mathbb{R}^p \to \mathbb{R}^q$ is bounded and, hence, uniformly continuous. Furthermore,*

$$||L|| \leq \left(\sum_{ij} |a_{ij}|^2 \right)^{1/2},$$

where $A = (a_{ij})$ is the matrix which determines L.

Proof. Let A be the matrix which determines L and let r_i be the ith row of A. Then the ith component of $y = L(x) = Ax$ is the inner product $y_i = r_i \cdot x$. By the Cauchy-Schwarz Inequality (Theorem 7.1.8)

$$|y_i| \leq ||r_i||\,||x||.$$

Thus,

$$||L(x)|| = \left(y_1^2 + \cdots + y_q^2 \right)^{1/2} \leq \left(||r_1||^2 + \cdots + ||r_q||^2 \right)^{1/2} ||x||$$

$$= \left(\sum_{ij} |a_{ij}|^2 \right)^{1/2} ||x||.$$

This implies that L is bounded and $||L|| \leq \left(\sum_{ij} |a_{ij}|^2 \right)^{1/2}$. $\qquad\square$

Inverse of a Matrix

Of particular interest in matrix theory are *square matrices* – that is, $p \times p$ matrices for some p. The product of two $p \times p$ matrices is another one and so the set of $p \times p$ matrices is closed under multiplication.

There is a multiplicative identity I in the set of $p \times p$ matrices. This is the matrix $I = (\delta_{ij})$ where $\delta_{ij} = 1$ if $i = j$ and $\delta_{ij} = 0$ otherwise. It has the property that

$$AI = IA = A,$$

for any $p \times p$ matrix A.

If A is a $p \times p$ matrix, then an *inverse* for A is a $p \times p$ matrix A^{-1} such that

$$AA^{-1} = A^{-1}A = I.$$

By Cramer's rule, a square matrix has an inverse if and only if its determinant $\det A$ is non-zero and, in this case,

$$A^{-1} = \frac{1}{\det A}(A^c)^t,$$

where A^c is the matrix of cofactors of A – that is, $A^c = ((-1)^{i+j} \det A_{ij})$, where A_{ij} is the $(p-1) \times (p-1)$ matrix obtained by deleting the ith row and jth column from A.

A matrix is said to be *non-singular* if it has an inverse, that is, if its determinant is non-zero. A square matrix is *singular* if it fails to have an inverse.

Note that if $L : \mathbb{R}^d \to \mathbb{R}^d$ is a linear transformation with matrix A, then A has an inverse matrix A^{-1} if and only L has an inverse transformation L^{-1} and, in this case, the linear transformation L^{-1} has A^{-1} as its associated matrix.

Example 8.4.12. Let

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & -1 \\ -2 & 1 \end{pmatrix}.$$

For each of A and B, determine if the matrix has an inverse and, if it does, find it.

Solution: The matrices A and B have determinants

$$\det A = 2 + 1 = 3 \quad \text{and} \quad \det B = 2 - 2 = 0.$$

Thus, A has an inverse and B does not. By Cramer's rule, the inverse of A is

$$\frac{1}{3}\begin{pmatrix} 1 & -1 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 1/3 & -1/3 \\ 1/3 & 2/3 \end{pmatrix}.$$

Remark 8.4.13. In what follows, we will often ignore the difference between a linear function L and the matrix which represents it. They are not exactly the same. The matrix of a linear transformation depends on a choice of coordinate systems in $\mathbb{R}^p$ and $\mathbb{R}^q$, while the linear transformation is independent of the

choice of coordinates. To ignore the distinction will not cause problems as long as we stick with one coordinate system. There will, however, be occasions where we change coordinate systems in $\mathbb{R}^p$ or R^q or both while dealing with a given linear transformation. It should be understood that the matrix corresponding to the linear transformation will, as a result, also change.

Exercise Set 8.4

The first five exercises involve the matrices

$$A = \begin{pmatrix} 3 & -1 \\ 2 & 1 \end{pmatrix}, B = \begin{pmatrix} 2 & 5 \\ -2 & 2 \end{pmatrix}, C = \begin{pmatrix} 1 & -1 \\ 4 & -6 \\ -1 & 2 \end{pmatrix}, D = \begin{pmatrix} 2 & 0 & 1 \\ -1 & 1 & 3 \end{pmatrix}.$$

1. Find $2A + B$, $A - B$, AB and BA.

2. Find $\det A$ and $\det B$ and A^{-1} and B^{-1}.

3. Find CD and DC.

4. Based on the result of the previous exercise, can you tell what $(CD)^2$ is without doing any further calculation?

5. Find $\det CD$.

6. Is the function $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $F(x, y) = (x + y, xy)$ a linear transformation? If so, what is its matrix?

7. Is the the function $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $F(x, y) = (x + y, x - y)$ a linear transformation? If so, what is its matrix?

8. Is the transformation of $\mathbb{R}^2$ to itself which rotates every vector through an angle θ (counterclockwise rotations have positive angle and clockwise rotations have negative angle) a linear transformation? If so, what is its matrix?

9. What is the matrix for the linear transformation of $\mathbb{R}^2$ which reflects each point through the diagonal line $y = x$ (this transformation interchanges the x and y coordinates of each point).

10. Find a linear transformation $L : \mathbb{R}^3 \to \mathbb{R}^3$ such that $L(1, 2, 1) = (1, 2, 1)$ and $L(u) = 0$ for every vector $u \in \mathbb{R}^3$ which is orthogonal to $(1, 2, 1)$.

11. Prove that if A is a $q \times p$ matrix, then

$$A(x + y) = Ax + Ay \quad \text{and} \quad A(cx) = c(Ax),$$

for every pair of vectors $x, y \in \mathbb{R}^p$ and every scalar $c \in \mathbb{R}$.

12. Prove Theorem 8.4.5.

13. Prove Theorem 8.4.6.

14. Prove that if K and L are linear transformations from $\mathbb{R}^p \to \mathbb{R}^q$, then

$$||K + L|| \leq ||K|| + ||L||.$$

15. Prove that if $K : \mathbb{R}^p \to \mathbb{R}^q$ and $L : \mathbb{R}^q \to R^r$ are linear transformations, then
$$||L \circ K|| \leq ||L|| \, ||K||.$$

16. Prove that the operator norm of a $p \times p$ diagonal matrix has norm equal to the largest absolute value of the elements on the diagonal.

8.5   Dimension, Rank, Lines, and Planes

A vector space X has finite dimension if it contains a finite set $\{x_1, x_2, \cdots, x_k\}$ of vectors which span X – that is, every vector in X is a linear combination of the vectors x_j. If this set is also *linearly independent*, meaning the only linear combination of the vectors x_j that equals 0 is the one in which all coefficients are zero, then the set $\{x_1, x_2, \cdots, x_k\}$ is called a *basis* for X. In this case, each element of X is a unique linear combination of the vectors x_j. Every finite dimensional vector space X has a basis. In fact X has many bases, but each of them has the same number of elements. This number is called the *dimension* of X and written $\dim(X)$.

A subset M of a vector space X is called a *linear subspace* if it is closed under addition and scalar multiplication – that is, $x + y \in M$ and $ax \in M$ whenever $x, y \in M$ and $a \in \mathbb{R}$. It follows that a linear subspace M of a vector space is itself a vector space, with addition and scalar multiplication in M defined in the same way they are defined in X. If X is finite dimensional, then so is the subspace M and any basis $\{x_1, x_2, \cdots, x_m\}$ for M can be expanded to a basis $\{x_1, x_2, \cdots, x_m, x_{m+1} \cdots, x_n\}$ for X. Thus

$$\dim(M) \leq \dim(X).$$

The set $\{e_1, \cdots e_p\}$ is a basis for $\mathbb{R}^p$, where recall that e_j is the p-tuple which has 1 for its jth component and 0 for all the others. However, this is not the only basis for $\mathbb{R}^p$.

Example 8.5.1. Show that the vectors $u = (1, 0, 1)$, $v = (1, 1, 0)$, and $w = (0, 1, 1)$ form a basis for $\mathbb{R}^3$.
Solution: Consider the vector equation

$$au + bv + cw = y. \tag{8.5.1}$$

To show that $\{u, v, w\}$ spans $\mathbb{R}^3$, we must show that this equation has a solution for every y. To show that $\{u, v, w\}$ is a linearly independent set, we must show that if $y = 0$, then this equation has only the zero solution for (a, b, c). Taken

together, these two statements mean that equation (8.5.1) should have a unique solution for every $y \in \mathbb{R}^3$. The vector equation (8.5.1) is equivalent to the system of linear equations

$$
\begin{aligned}
a + b + 0 &= y_1 \\
0 + b + c &= y_2, \\
a + 0 + c &= y_3
\end{aligned}
$$

which, in turn, may be written as the vector matrix equation

$$
\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}.
$$

The matrix in this equation has determinant 2 and so the matrix has an inverse. This implies that the equation has a unique solution (a, b, c) for each $y = (y_1, y_2, y_3)$ and, hence, that $\{u, v, w\}$ is a basis for $\mathbb{R}^3$.

Definition 8.5.2. If $L : X \to Y$ is a linear transformation between vector spaces, then the *image* of L, denoted $\operatorname{im}(L)$ is the set

$$
L(X) = \{L(x) : x \in X\},
$$

while the *kernel* of L, denoted $\ker(L)$, is the set

$$
\{x \in X : L(x) = 0\}.
$$

Since L is linear, it follows easily that its kernel and image are linear subspaces of X and Y, respectively.

Theorem 8.5.3. *If $L : X \to Y$ is a linear transformation between finite dimensional vector spaces, then*

$$
\dim(\ker(L)) + \dim(\operatorname{im}(L)) = \dim(X).
$$

Proof. Let $\dim(\ker(L)) = m$ and let $\{x_1, x_2, \cdots, x_m\}$ be a basis for $\ker(L)$. We may expand this to a basis $\{x_1, x_2, \cdots, x_m, x_{m+1} \cdots, x_n\}$ for X.

Set $y_j = L(x_{m+j})$ for $j = 1, \cdots, n - m$. Since every vector in X is a linear combination of the vectors $x_1, \cdots, x_n$ and $L(x_k) = 0$ for $k = 1, \cdots, m$, we conclude that every vector in $\operatorname{im}(L)$ is a linear combination of the vectors $y_1, \cdots, y_{n-m}$. This set of vectors is linearly independent, since if

$$
a_1 y_1 + a_2 y_2 + \cdots + a_{n-m} y_{n-m} = 0,
$$

then $a_1 x_{m+1} + a_2 x_{m+2} + \cdots + a_{n-m} x_n \in \ker(L)$. This implies that there are numbers $b_1, \cdots, b_m$ such that

$$
a_1 x_{m+1} + a_2 x_{m+2} + \cdots + a_{n-m} x_n = b_1 x_1 + b_2 x_2 + \cdots + b_m x_m.
$$

However, since $\{x_1, \cdots, x_n\}$ is a linearly independent set, the a_js and b_ks must all be 0. The fact that the a_js must all be 0 shows that the set $\{y_1, \cdots, y_{n-m}\}$ is linearly independent and, hence, forms a basis for $\operatorname{im}(L)$.

We now have $\dim(X) = n$, $\dim(\ker(L)) = m$ and $\dim(\operatorname{im}(L)) = n - m$. Thus, $\dim(\ker(L)) + \dim(\operatorname{im}(L)) = \dim(X)$, as claimed. $\square$

Definition 8.5.4. Let A be a $q \times p$ matrix and let $L : \mathbb{R}^p \to \mathbb{R}^q$ be the linear transformation it determines. Then $\mathrm{Rank}(A)$ is defined to be $\dim(\mathrm{im}(L))$. Equivalently, by the previous theorem, it is also equal to $\dim(X) - \dim(\ker(L))$. If L is a linear transformation whose matrix has rank r, then we will also say that L has rank r.

A *submatrix* of a matrix A is a matrix obtained from A by deleting some of its rows and columns.

The following is proved in most linear algebra texts. We won't repeat the proof here.

Theorem 8.5.5. *The rank of a $q \times p$ matrix A is r, where $r \times r$ is the dimension of the largest square submatrix of A with non-zero determinant.*

Example 8.5.6. What is the rank of the matrix

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 4 \\ 1 & -1 \end{pmatrix}?$$

Solution: This matrix has

$$\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}$$

as a 2×2 submatrix with determinant -3. It has no square submatrices of larger dimension. Therefore, the matrix A has rank 2.

Example 8.5.7. What is the rank of the matrix

$$B = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & -1 & -2 \end{pmatrix}?$$

Solution: This matrix also has

$$\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}$$

as a 2×2 submatrix with determinant -3. The only square submatrix of larger dimension is the matrix B and this has determinant 0 Therefore, the matrix B also has rank 2.

Affine Functions

Definition 8.5.8. An *affine function* $F : \mathbb{R}^p \to \mathbb{R}^q$ is a function of the form

$$F(x) = b + L(x),$$

where $b \in \mathbb{R}^q$ and $L : R^p \to \mathbb{R}^q$ is a linear function. The *rank* of an affine transformation F is the rank of its linear part L.

An *affine subspace* M of $\mathbb{R}^p$ is a translate $b + N$ of a linear subspace N of $\mathbb{R}^p$. In this case, the dimension of M is defined to be the dimension of N.

The image of an affine function $F(x) = b + L(x)$ is an affine subspace $b + \text{im}(L)$ – that is, it is the translate $b + \text{im}(L)$ of the linear subspace $\text{im}(L)$. The dimension of this subspace is the rank of L.

Similarly, if $F(x) = b + L(x)$ is an affine function, then the set of solutions to the vector equation $F(x) = 0$ is also an affine subspace. In fact, if a is one such solution (so that $F(a) = b + L(a) = 0$), then x is also a solution if and only if

$$L(x - a) = -b + b = 0.$$

Hence, x is a solution if and only if $x \in a + \ker(L)$. Thus, the set of solutions of the vector equation $F(x) = 0$ is the translate $a + \ker(L)$ of the linear subspace $\ker(L)$ of $\mathbb{R}^p$ and, hence, is an affine subspace. The dimension of this subspace is $p - \text{Rank}(L)$.

In general, if $M = b + N$ an affine subspace, with N the corresponding linear subspace, then we will say that N is the set of vectors *parallel* to the affine subspace M.

Lines in $\mathbb{R}^3$

Lines in $\mathbb{R}^p$ are one dimensional affine subspaces of $\mathbb{R}^p$. The above discusssion suggests expressing them as either images of rank 1 affine transformations or as kernels of rank $p - 1$ affine transformations with domain $\mathbb{R}^p$.

A rank 1 affine transformation $\gamma : \mathbb{R} \to \mathbb{R}^q$ has the form

$$\gamma(t) = a + tu. \tag{8.5.2}$$

The image of this transformation is a line which contains the point $a = F(0)$ and is parallel to the vector $u = \gamma(1) - \gamma(0)$.

On the other hand, given a line in $\mathbb{R}^q$, if we choose distinct points a and b on the line, and we set $u = b - a$, then the image of the affine transformation (8.5.2) is a line which contains both $a = \gamma(0)$ and $b = \gamma(1)$ and, hence, is the line we started with.

Thus, the lines in $\mathbb{R}^q$ are exactly the images of affine transformations of the form (8.5.2). This situation is often expressed as a vector equation

$$x = a + tu,$$

which describes the points x on the line as the values assumed by the right side of the equation as t ranges over $\mathbb{R}$ This is a *parametric vector equation* for the line.

In $\mathbb{R}^3$, a parametric vector equation for a line takes the form $(x, y, z) = (a_1, a_2, a_3) + t(u_1, u_2, u_3)$, which is equivalent to the system of parametric equations

$$
\begin{aligned}
x &= a_1 + tu_1 \\
y &= a_2 + tu_2 \\
z &= a_3 + tu_3
\end{aligned}
$$

Example 8.5.9. Find parametric equations for the line in $\mathbb{R}^3$ which contains the point $(1, 0, 0)$ and is parallel to the vector $u = (-3, 4, 5)$.

  Solution: A parametric vector equation for this line is

$$(x, y, z) = (1, 0, 0) + t(-3, 4, 5).$$

The corresponding system of parametric equations is

$$
\begin{aligned}
x &= 1 - 3t \\
y &= 4t \\
z &= 5t
\end{aligned}
$$

Example 8.5.10. Find parametric equations for the line in $\mathbb{R}^3$ containing the points $(2, 1, 1)$ and $(5, -1, 3)$.

  Solution: If we set $u = (5, -1, 3) - (2, 1, 1) = (3, -2, 2)$, then the parametric equation for our line in vector form is

$$(x, y, z) = (2, 1, 1) + t(3, -2, 2) = (2 + 3t, 1 - 2t, 1 + 2t).$$

This can also be expressed as the system of parametric equatiions

$$
\begin{aligned}
x &= 2 + 3t \\
y &= 1 - 2t \\
z &= 1 + 2t
\end{aligned}
$$

To express a line in $\mathbb{R}^q$ as the kernel of an affine transformation, we choose a point a on the line and a vector u parallel to the line (we may choose $u = b - a$ where b is a point on the line distinct from a). If A is a matrix whose rows form a basis for the linear subspace

$$\{y \in \mathbb{R}^p : y \cdot u = 0)\},$$

then A is a $p-1 \times p$ matrix of rank $p-1$ and $Au = 0$. This means that the kernel of the linear transformation determined by A has dimension 1 and contains u. Hence, this kernel is $\{tu : t \in \mathbb{R}\}$. The line $\{a + tu : t \in \mathbb{R}\}$ contains a and is parallel to u. Thus, it must be our original line. By the construction of A, it also has the form

$$\{x \in \mathbb{R}^p : A(x - a) = 0\} = \{x \in \mathbb{R}^p : Ax - c = 0\} \quad \text{where} \quad c = Aa.$$

Thus, our line is the kernel of the affine transformation F defined by $F(x) = Ax - c$.

If we apply the above discussion to $\mathbb{R}^3$, we conclude that the typical line in $\mathbb{R}^3$ is the set of solutions (x, y, z) to an equation of the form

$$
\begin{pmatrix} v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}
\begin{pmatrix} x \\ y \\ z \end{pmatrix}
=
\begin{pmatrix} c_1 \\ c_2 \end{pmatrix},
$$

Where (v_1, v_2, v_3) and (w_1, w_2, w_3) are linearly independent vectors. In other words, it is the set of all simultaneous solutions of the pair of linear equations

$$
\begin{aligned}
v_1 x + v_2 y + v_3 z &= c_1 \\
w_1 x + w_2 y + w_3 z &= c_2.
\end{aligned}
$$

Example 8.5.11. Express the line in Example 8.5.10 as the set of solutions of a pair of linear equations.

  Solution: We need to find two linearly independent vectors which are orthogonal to $u = (3, -2, 2)$. Such a pair is $(2, 3, 0)$ and $(2, 1, -2)$. If we apply the matrix with these two vectors as rows to the vector $a = (2, 1, 1)$, the result is

$$
\begin{pmatrix} 2 & 3 & 0 \\ 2 & 1 & -2 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 7 \\ 3 \end{pmatrix},
$$

Thus, in vector matrix form, the equation of our line is

$$
\begin{pmatrix} 2 & 3 & 0 \\ 3 & 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 7 \\ 3 \end{pmatrix},
$$

This is equivalent to the pair of simultaneous equations

$$
\begin{aligned}
2x + 3y &= 7 \\
3x + y - 2z &= 3
\end{aligned}
$$

Planes in $\mathbb{R}^3$

A plane in $\mathbb{R}^p$ is a two dimensional affine subspace of $\mathbb{R}^p$ – that is, a translate of a two dimensional linear subspace of $\mathbb{R}^p$. Such an object can be described as the image of an affine transformation of rank 2 or the kernel of an affine transformation of rank $p - 2$ with domain $\mathbb{R}^p$.

  If u and v are linearly independent vectors in $\mathbb{R}^p$, then they form a basis for a 2-dimensional linear subspace of $\mathbb{R}^p$. If we translate this subspace by adding a to each of its points, we obtain a plane which contains a and is parallel to u and v. It consists of all points of the form

$$
x = a + su + tv;
$$

that is, it is the image of the affine transformation $F : \mathbb{R}^2 \to \mathbb{R}^p$ defined by

$$
F(s, t) = a + su + tv.
$$

This is the vector parametric form for the equation of a plane.

  In the case where $p = 3$, a vector parametric equation of a plane has the form

$$
\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} + \begin{pmatrix} u_1 & v_1 \\ u_2 & v_2 \\ u_3 & v_3 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix},
$$

or, when written as a system of equations,

$$x = a_1 + su_1 + tv_1$$
$$y = a_2 + su_2 + tv_2$$
$$z = a_3 + su_3 + tv_3$$

Given three points a, b, c in $\mathbb{R}^p$ which do not lie on the same line, the vectors $u = b - a$ and $v = c - a$ are linearly independent (Exercise 8.5.15). Hence, a, u, and v determine an affine function F with image a plane, as above. This plane contains the points $a = F(0,0)$, $b = F(1,0)$, and $c = F(0,1)$.

Example 8.5.12. Find parametric equations for the plane that contains the three points $(1,0,1)$, $(1,1,2)$, $(-1,2,0)$.

Solution: We choose $a = (1,0,1)$, $u = (1,1,2) - (1,0,1) = (0,1,1)$, and $v = (-1,2,0) - (1,0,1) = (-2,2,-1)$. Then, according to the above discussion, the plane we seek has parametric equations

$$
\begin{aligned}
x &= 1 & - 2t \\
y &= & s + 2t \\
z &= & 1 + s - t.
\end{aligned}
$$

We can also express a plane in $\mathbb{R}^3$ as the kernel of a rank 1 affine transformation from $\mathbb{R}^3$ to $\mathbb{R}$. If $a = (a_1, a_2, a_3)$ is a fixed point in the plane, $u = (x, y, z)$ the general point of the plane, and $v = (v_1, v_2, v_3)$ a vector perpendicular to the plane, then $v \cdot (u - a) = 0$. Thus, the plane is the kernel of the affine transformation $f : \mathbb{R}^3 \to \mathbb{R}$ defined by $f(u) = v \cdot u - b$, where $b = v \cdot a$. The equation of the plane is then

$$v_1 x + v_2 y + v_3 z = b.$$

Example 8.5.13. Find an equation for the plane of Example 8.5.12.

Solution: We choose $a = (1,0,1)$ as a point in the plane. Now we need a vector perpendicular to the plane. The vectors $(0,1,1)$ and $(-2,2,-1)$ are parallel to the plane and so we need to find a vector orthogonal to each of these. In fact, $(3,2,-2)$ is orthogonal to each of these vectors. Also,

$$(3, 2, -2) \cdot (1, 0, 1) = 1.$$

Hence, an equation for our plane is

$$3x + 2y - 2z = 1.$$

Exercise Set 8.5

1. Do the vectors $(1,2,1), (2,0,1)$, and $(1,-1,1)$ form a basis for $\mathbb{R}^3$. Justify your answer.

2. Do the vectors $(1,2,1), (2,0,1)$, and $(0,4,1)$ form a basis for $\mathbb{R}^3$. Justify your answer.

3. What is the rank of the matrix $\begin{pmatrix} 1 & 2 & 1 \\ 2 & 3 & -1 \\ 1 & 1 & -2 \end{pmatrix}$?

4. What is the rank of the matrix $\begin{pmatrix} 1 & -2 & 3 \\ -2 & 4 & -6 \end{pmatrix}$?

5. What is the rank of the matrix $\begin{pmatrix} 1 & -2 & 3 \\ -2 & 3 & -6 \end{pmatrix}$?

6. Find parametric equations for the line in $\mathbb{R}^3$ which contains the point $(1, 2, 3)$ and is parallel to the vector $(1, 1, 1)$.

7. Find parametric equations for the line in $\mathbb{R}^3$ containing both $(1, 1, 1)$ and $(3, -1, 3)$.

8. Express the line of the previous exercise as the set of simultaneous solutions of a pair of linear equations.

9. Find parametric equations for the plane that contains the three points $(1, 0, -1)$, $(2, 1, 2)$, $(-1, 2, 3)$.

10. Express the plane of the previous exercise as the set of solutions of a linear equation.

11. Find parametric equations for a line which passes through the origin and is perpendicular to the plane $x - y + 3z = 5$. Use this line to determine the distance from the plane to the origin.

12. Find the distance from the line with parametric vector equation $(x, y, z) = (1 + 2t, 2 - t, 4 + t)$ to the origin.

13. Find a formula for the point on the one dimensional subspace of $\mathbb{R}^p$ generated by a non-zero vector u which is closest to the point $a \in \mathbb{R}^p$.

14. Prove that, in $\mathbb{R}^3$, a plane and a line not parallel to it must meet in exactly one point.

15. Prove that if a, b, and c are three points in $\mathbb{R}^p$ which do not lie on the same line, then the vectors $u = b - a$ and $v = c - a$ are linearly independent.

16. Prove that if M is a linear subspace of $\mathbb{R}^p$ and we set

$$M^{\perp} = \{y \in \mathbb{R}^p : y \perp x \quad \text{for all} \quad x \in M\},$$

then $M^{\perp}$ is a also a linear subspace of $\mathbb{R}^p$ and every vector in $u \in \mathbb{R}^p$ may be written in a unique way as $u = x + y$ with $x \in M$ and $y \in M^{\perp}$ (see Definition 7.1.9).

Chapter 9

Differentiation in Several Variables

The most powerful method available for studying a function in several variables is to approximate it locally, near a given point, by an affine function. When this can be done, it provides a wealth of information about the original function. Affine approximation leads to the definition of the *differential* of a function of several variables. The differential of a function F, when it exists, is a matrix of partial derivatives of coordinate functions of F. For this reason, we precede the discussion of the differential with a brief review of partial derivatives.

9.1 Partial Derivatives

In this section, f will be a real valued function defined on an open set in $\mathbb{R}^p$.

Definition 9.1.1. The partial derivative of f with respect to its jth variable at $x = (x_1, \cdots, x_j, \cdots, x_p)$ is denoted $\dfrac{\partial f}{\partial x_j}(x)$ and is defined by

$$\frac{\partial f}{\partial x_j}(x) = \frac{d}{dt} f(x_1, \cdots, x_{j-1}, t, x_{j+1}, \cdots, x_p)|_{t=x_j},$$

provided this derivative exists.

Thus, the partial derivative of a function f, with respect to its jth variable, at a point x in its domain is obtained by fixing all of the variables of f, except the jth one, at the appropriate values $x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_p$, then differentiating with respect to the remaining variable and evaluating at x_j.

Remark 9.1.2. When it is not necessary to explicitly exhibit the point x at which the partial derivative is being computed (because it is understood from the context or because x is a generic point of the domain of f) we will simply write $\dfrac{\partial f}{\partial x_j}$ for the partial derivative of x with respect to its jth variable.

Two other notations that are often used for the partial derivative of f with respect to x_j are f_{x_j} and f_j. We won't use these in this text.

Example 9.1.3. Find the partial derivatives of the function

$$f(x_1, x_2, x_3, x_4) = x_1^2 + x_1 x_3 - 4x_2^2 x_4^3.$$

Solution: To find $\dfrac{\partial f}{\partial x_1}$, we consider x_2, x_3, x_4 to be fixed constants and we differentiate with respect to the remaining variable and evaluate at x_1. The result is

$$\frac{\partial f}{\partial x_1} = 2x_1 + x_3.$$

Similarly, we have

$$\frac{\partial f}{\partial x_2} = -8x_2 x_4^3,$$
$$\frac{\partial f}{\partial x_3} = x_1,$$
$$\frac{\partial f}{\partial x_4} = -12x_2^2 x_4^2.$$

Example 9.1.4. Find the partial derivatives of the function

$$f(x, y, z) = z^2 \cos xy.$$

Solution: We have

$$\frac{\partial f}{\partial x} = -yz^2 \sin xy,$$
$$\frac{\partial f}{\partial y} = -xz^2 \sin xy,$$
$$\frac{\partial f}{\partial z} = 2z \cos xy.$$

The Partial Derivatives as Limits

If we use the definition of the devivative of a function of one variable as the limit of a difference quotient, the result is

$$\frac{\partial f}{\partial x_j}(x_1, \cdots, x_p) = \lim_{h \to 0} \frac{f(x_1, \cdots, x_j + h, \cdots, x_p) - f(x_1, \cdots, x_j, \cdots, x_p)}{h}.$$

The notation involved in this statement becomes much simpler if we note that the point $(x_1, \cdots, x_j + h, \cdots, x_p)$ may be written as $x + h\,e_j$, where e_j is the basis vector with 1 in the jth entry and 0 elsewhere. Then,

$$\frac{\partial f}{\partial x_j}(x) = \lim_{h \to 0} \frac{f(x + h\,e_j) - f(x)}{h}. \tag{9.1.1}$$

Higher Order Partial Derivatives

The partial derivatives defined so far are *first order* partial derivatives. We define second order partial derivatives of f in the following fashion: for $i, j = 1, \cdots, p$ we set

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial x_j} \right). \tag{9.1.2}$$

The meaning of this is as follows: If the partial derivative $\dfrac{\partial f}{\partial x_j}$ exists in a neighborhood of a point $x \in \mathbb{R}^p$, then we may attempt to take the partial derivative with respect to x_i of the resulting function at the point x. The result, if it exists, is the right side of the above equation. The expression on the left is the notation that is commonly used for this second order partial derivative. In the case where $i = j$, we modify this notation slightly and write

$$\frac{\partial^2 f}{\partial x_j^2} = \frac{\partial}{\partial x_j} \left(\frac{\partial f}{\partial x_j} \right).$$

A useful way to think of this process is as follows: the expression $\dfrac{\partial}{\partial x_j}$ is an *operator* – that is, a transformation which takes a function f on an open set U to another function $\dfrac{\partial f}{\partial x_j}$ on U (provided this derivative exists on U). In fact, this operator is a linear operator (preserves sums and scalar products) because the derivative of a sum is the sum of the derivatives and the derivative of a constant times a function is the constant times the derivative of the function. Such operators may be *composed* – that is, we may first apply one such operator, $\dfrac{\partial}{\partial x_j}$, to a function and then apply another, $\dfrac{\partial}{\partial x_i}$, to the result. In fact, we may continue to compose such operators, applying one after another, as long as the resulting function has the appropriate partial derivatives on the given open set. From this point of view, the second order partial derivative of (9.1.2) is just the result of applying to f the second order differential operator

$$\frac{\partial^2}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i} \circ \frac{\partial}{\partial x_j}.$$

We may, of course, define higher order partial differential operators in an analogous fashion. Given integers $j_1, j_2, \cdots, j_m$ between 1 and p, we set

$$\frac{\partial^m}{\partial x_{j_1} \cdots \partial x_{j_m}} = \frac{\partial}{\partial x_{j_1}} \circ \frac{\partial}{\partial x_{j_2}} \circ \cdots \circ \frac{\partial}{\partial x_{j_m}}.$$

The resulting operator is a partial differential operator of *total degree m*.

Example 9.1.5. Find $\dfrac{\partial^5 f}{\partial x \partial y \partial z \partial y \partial x}$ if $f(x, y, z) = x^2 y^3 z^4 + x^2 + y^4 + xyz$.

Solution: We proceed one derivative at a time:

$$\text{apply} \quad \frac{\partial}{\partial x} : \qquad \frac{\partial f}{\partial x} = 2xy^3z^4 + 2x + yz,$$

$$\text{apply} \quad \frac{\partial}{\partial y} : \qquad \frac{\partial^2 f}{\partial y \partial x} = 6xy^2z^4 + z,$$

$$\text{apply} \quad \frac{\partial}{\partial z} : \qquad \frac{\partial^3 f}{\partial z \partial y \partial x} = 24xy^2z^3 + 1,$$

$$\text{apply} \quad \frac{\partial}{\partial y} : \qquad \frac{\partial^4 f}{\partial y \partial z \partial y \partial x} = 48xyz^3,$$

$$\text{apply} \quad \frac{\partial}{\partial x} : \qquad \frac{\partial^5 f}{\partial x \partial y \partial z \partial y \partial x} = 48yz^3.$$

Equality of Mixed Partials

It is natural to ask whether or not, in a mixed higher order partial derivative, the order in which the derivatives are taken makes a difference. Some additional calculation using the previous example (Exercise 9.1.5) shows that, at least for the function f of that example, the order in which the five partial derivative operators are applied makes no difference. This is not always the case, but it is the case under rather mild continuity assumptions. When it is the case, we may change the order in which the partial derivatives are taken so as to collect partial derivatives with respect to the same variable together. For example, the 5th order mixed partial derivative of the previous example can be re-written as

$$\frac{\partial^5 f}{\partial x \partial x \partial y \partial y \partial z} = \frac{\partial^5 f}{\partial x^2 \partial y^2 \partial z}.$$

The next theorem tells us when interchanging the order of a mixed partial derivative is legitimate.

Theorem 9.1.6. *Suppose f is a function defined on an open disc $B_r(a,b) \subset \mathbb{R}^2$. Also suppose that both first order partial derivatives exist in $B_r(a,b)$ and that $\frac{\partial^2 f}{\partial y \partial x}$ exists in $B_r(a,b)$ and is continuous at (a,b). Then $\frac{\partial^2 f}{\partial x \partial y}$ exists at (a,b) and is equal to $\frac{\partial^2 f}{\partial y \partial x}(a,b)$.*

Proof. We introduce a function $\lambda(h,k)$, defined for (h,k) in the disc $B = B_r(0,0)$, by

$$\lambda(h,k) = f(a+h, b+k) - f(a+h, b) - f(a, b+k) + f(a,b).$$

It follows from the hypotheses of the theorem that the partial derivative of $\lambda(h,k)$ with respect to h exists for all (h,k) in the disc B. If $(h,k) \in B$, the rectangle with vertices $(0,0)$, $(0,k)$, $(h,0)$ and (h,k) is also contained in this disc and so the partial derivative of λ with respect to its first variable exists on an open set containing this rectangle.

Now for fixed k,

$$\lambda(h,k) = g(h) - g(0) \quad \text{where} \quad g(u) = f(a+u, b+k) - f(a+u, b).$$

The function g is differentiable on an open interval containing $[0,h]$, and so we may apply the Mean Value Theorem to g to conclude there is a number $s \in (0,h)$ such that $g(h) - g(0) = hg'(s)$. This means

$$\lambda(h,k) = h\left(\frac{\partial f}{\partial x}(a+s, b+k) - \frac{\partial f}{\partial x}(a+s, b)\right). \qquad (9.1.3)$$

Of course, the number s depends on h and k.

Since $\dfrac{\partial^2 f}{\partial y \partial x}$ exists on B, $\dfrac{\partial f}{\partial x}$ is a differentiable function of its second variable on B. Hence, we may apply the Mean Value Theorem to this function as well. We conclude that there is a point $t \in (0,k)$ such that

$$\frac{\partial f}{\partial x}(a+s, b+k) - \frac{\partial f}{\partial x}(a+s, b) = k\frac{\partial^2 f}{\partial y \partial x}(a+s, b+t). \qquad (9.1.4)$$

Combining (9.1.3) and (9.1.4) yields

$$\frac{1}{hk}\lambda(h,k) = \frac{\partial^2 f}{\partial y \partial x}(a+s, b+t).$$

By hypothesis, the second order partial derivative on the right is continuous at (a,b). This implies that

$$\lim_{(h,k)\to(0,0)} \frac{\lambda(h,k)}{hk} = \frac{\partial^2 f}{\partial y \partial x}(a,b).$$

This conclusion uses the fact that the point $(a+s, b+t)$, wherever it is, is at least closer to (a,b) than the point $(a+h, b+k)$.

We complete the proof by noting that the above limit exists independently of how (h,k) approaches $(0,0)$. In particular, the result will be the same if we first let k approach 0 and then h. However,

$$\lim_{h\to 0}\lim_{k\to 0}\frac{1}{hk}\lambda(h,k)$$

$$= \lim_{h\to 0}\lim_{k\to 0}\frac{1}{h}\left(\frac{f(a+h, b+k) - f(a+h, b)}{k} - \frac{f(a, b+k) - f(a,b)}{k}\right)$$

$$= \lim_{h\to 0}\frac{1}{h}\left(\lim_{k\to 0}\frac{f(a+h, b+k) - f(a+h, b)}{k} - \lim_{k\to 0}\frac{f(a, b+k) - f(a,b)}{k}\right)$$

$$= \lim_{h\to 0}\frac{1}{h}\left(\frac{\partial f}{\partial y}(a+h, b) - \frac{\partial f}{\partial y}(a,b)\right)$$

$$= \frac{\partial^2 f}{\partial x \partial y}(a,b).$$

Hence, this second order partial derivative also exists and it equals $\dfrac{\partial^2 f}{\partial y \partial x}(a,b)$.
Note that distributing the limit with respect to k across the difference in the second step above requires that we know the two limits involved exist. This follows from the assumption that $\dfrac{\partial f}{\partial y}$ exists in $B_r(a,b)$. $\square$

Obviously, the same result holds, with the same proof, if x and y are reversed in the statement of the above theorem. That is, if we assume either one of the second order mixed partials exists in a neighborhood of (a,b) and is continuous at (a,b), then the other one also exists at (a,b) and the two are equal at (a,b).

The following example shows that the continuity of the mixed partial that is assumed to exist is a necessary assumption in the above theorem.

Example 9.1.7. For the function

$$f(x,y) = \begin{cases} \dfrac{x^3 y - xy^3}{x^2 + y^2} & \text{if} \quad (x,y) \neq (0,0) \\ 0 & \text{if} \quad (x,y) = (0,0) \end{cases},$$

show that the first order partial derivatives exist and are continuous everywhere. Then show that the mixed second order partial derivatives $\dfrac{\partial^2 f}{\partial x \partial y}$ and $\dfrac{\partial^2 f}{\partial y \partial x}$ exist everywhere, but they are not equal at $(0,0)$. Why doesn't this contradict the above theorem?

Solution: Except at the point $(0,0)$ where the denominator vanishes, we may use the standard rules of differentiation to show that

$$\begin{aligned}
\frac{\partial f}{\partial x} &= \frac{(3x^2 y - y^3)(x^2 + y^2) - 2x(x^3 y - xy^3)}{(x^2 + y^2)^2}, \\
\frac{\partial f}{\partial y} &= \frac{(x^3 - 3xy^2)(x^2 + y^2) - 2y(x^3 y - xy^3)}{(x^2 + y^2)^2}.
\end{aligned} \tag{9.1.5}$$

These expressions may be differentiated again to show that each of the second order partial derivatives also exists, except possibly at $(0,0)$.

In order to calculate $\dfrac{\partial f}{\partial x}(0,0)$ we set $y = 0$ in the expression for f. The resulting function of x is identically 0 and, hence, has derivative 0 with respect to x. Similar reasoning leads to the same conclusion for $\dfrac{\partial f}{\partial y}(0,0)$. Since both the expressions in (9.1.5) have limit 0 as $(x,y) \to (0,0)$, the first order partial derivatives are continuous everywhere, including at $(0,0)$, where they both have the value 0.

To calculate $\dfrac{\partial^2 f}{\partial x \partial y}$, we first note that $\dfrac{\partial f}{\partial y}(x,0) = x$, for all x. Hence, $\dfrac{\partial^2 f}{\partial x \partial y}(0,0) = 1$.

On the other hand, $\dfrac{\partial f}{\partial x}(0,y) = -y$, and so $\dfrac{\partial^2 f}{\partial y \partial x}(0,0) = -1$.

The two mixed partials are not equal at $(0,0)$ even though they both exist everywhere. Why doesn't this contradict the previous theorem? It must be the case that niether of these mixed partial derivatives is continuous at $(0,0)$ – a fact that will be verified in the exercises.

An important hypothesis in many theorems is that a function f belongs to the class $\mathcal{C}^k(U)$ defined below.

Definition 9.1.8. If U is an open subset of $\mathbb{R}^p$ then a function $F : U \to \mathbb{R}^q$ is said to be $\mathcal{C}^k$ on U if, for each coordinate function f_j of F, all partial derivatives of f_j of total order less than or equal to k exist and are continuous on U.

Functions which are $\mathcal{C}^1$ on U will be called *smooth* functions on U.

By using Theorem 9.1.6 to interchange pairs of adjacent first order partial differential operators, the following theorem may be proved:

Theorem 9.1.9. *If a real valued function f is $\mathcal{C}^k$ on $U \subset \mathbb{R}^p$ and $m \leq k$, then the mth order partial derivative $\dfrac{\partial^m f}{\partial x_{j_1} \cdots \partial x_{j_m}}$ is independent of the order in which the first order partial derivatives $\dfrac{\partial}{\partial j_i}$ are applied.*

Exercise Set 9.1

1. If $f(x,y) = \sqrt{x^2 + y^2}$, find $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$. Are there any points in the plane where they don't exist?

2. If $f(x,y) = xy^2 + xy + y^3$, find all first and second order partial derivatives of f.

3. If $f(x,y) = x \cos y$, find $\dfrac{\partial f}{\partial x}$, $\dfrac{\partial f}{\partial y}$, $\dfrac{\partial^2 f}{\partial x \partial y}$, and $\dfrac{\partial^2 f}{\partial y \partial x}$.

4. If $f(x,y) = e^{xy} \sin y$, find $\dfrac{\partial f}{\partial x}$, $\dfrac{\partial f}{\partial y}$, $\dfrac{\partial^2 f}{\partial x \partial y}$, and $\dfrac{\partial^2 f}{\partial y \partial x}$.

5. If f is the function of Example 9.1.5 directly calculate

$$\frac{\partial^5 f}{\partial x^2 \partial y^2 \partial z}.$$

Verify that it is the same as the mixed partial derivative of f calculated in the example.

6. Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable on $\mathbb{R}$ and define a function $g : \mathbb{R}^2 \to \mathbb{R}$ by $g(x,y) = f(x+y)$. Use (9.1.1) to show that $\dfrac{\partial g}{\partial x} = \dfrac{\partial g}{\partial y}$ on $\mathbb{R}^2$.

7. Theorem 9.1.6 is a statement about a function of two variables. Show how it can be applied several times in a step by step procedure to prove that if $U \subset \mathbb{R}^3$ and f is $\mathcal{C}^3$ on U, then

$$\frac{\partial^3 f}{\partial x \partial y \partial z} = \frac{\partial^3 f}{\partial z \partial y \partial x}.$$

8. If $p > 0$, let f be the function

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

For which values of p is $\dfrac{\partial f}{\partial x}$ continuous at $(0, 0)$?

9. If f is the function of Example 9.1.7, show by direct calculation that $\dfrac{\partial^2 f}{\partial x \partial y}$ is not continuous at $(0, 0)$. A similar calculation shows that $\dfrac{\partial^2 f}{\partial y \partial x}$ is not continuous at $(0, 0)$ (you need not do both calculations).

10. If f is defined on $\mathbb{R}^2$ by

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

show that both $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$ exist everywhere, but they are not continuous at $(0, 0)$. In fact, f itself is not continuous at $(0, 0)$ (see Example 8.1.3).

9.2   The Differential

Let f be a real valued function defined on an interval on the line. Recall that the equation of the tangent line to the curve $y = f(x)$, at a point a where f is differentiable, is:

$$y = f(a) + f'(a)(x - a)$$

This is the equation of the line which best approximates the curve when x is near a. The right side, is an affine function,

$$T(x) = f(a) + f'(a)(x - a),$$

of x. What is special about T that makes its graph the line which best approximates the curve $y = f(x)$ near a? For convenience of notation let $h = x - a$, so that $x = a + h$. Then

$$f(a + h) - T(a + h) = f(a + h) - f(a) - f'(a)h$$

and so

$$\lim_{h \to 0} \frac{f(a+h) - T(a+h)}{h} = \lim_{h \to a} \frac{f(a+h) - f(a)}{h} - f'(a) = 0.$$

In other words, not only do f and T have the same value at a, but as h approaches 0, the difference between $f(a+h)$ and $T(a+h)$ approaches zero faster than h does. No affine function other than T has this property (Exercise 9.2.7).

Example 9.2.1. What is the best affine approximation to $f(x) = x^3 - 2x + 1$ at the point $(2, 5)$?

   Solution: Here, $a = 2$, $f(a) = 5$, and $f'(a) = f'(2) = 22$, so the best affine approximation to $f(x)$ at $x = 2$ is $T(x) = 5 + 22(x - 2) = 22x - 39$.

Affine Approximation in Several Variables

By analogy with the single variable case, if $F : D \to \mathbb{R}^q$ is a function defined on a subset D of $\mathbb{R}^p$, then the best affine approximation to F at $a \in D$ would be an affine function $T : \mathbb{R}^p \to \mathbb{R}^q$ such that $F(a+h) - T(a+h)$ goes to 0 faster than h as the vector h approaches 0. In order for this to make sense at all, a must be a limit point of D and, in fact, we will require that a be an interior point of D. This ensures that there is an open ball, centered at a, which is contained in D.

   It must also be the case that F and its affine approximation T have the same value at a. However, if T is affine, then $T(a+h) = b + L(a+h)$ where $L : \mathbb{R}^p \to \mathbb{R}^q$ is linear and $b \in \mathbb{R}^q$ is a constant. If $T(a) = F(a)$, then $b = T(a) - L(a)$ and so T has the form $T(a+h) = F(a) - L(a) + L(a+h)$. Then, since L is linear,

$$T(a+h) = F(a) + L(h).$$

   A function which has a best affine approximation at a is said to be *differentiable* at a. The precise definition of this concept is as follows:

Definition 9.2.2. Let $F : D \to \mathbb{R}^q$ be a function with domain $D \subset \mathbb{R}^p$, and let a be an interior point of D. We say that F is differentiable at a if there is a linear function $L : \mathbb{R}^p \to \mathbb{R}^q$ such that

$$\lim_{h \to 0} \frac{F(a+h) - F(a) - L(h)}{||h||} = 0. \qquad (9.2.1)$$

   In this case, we call the linear function L the *differential* of F at a and denote it by $dF(a)$.

   Just as in the single variable case, if F is differentiable, then the function

$$T(x) = F(a) + dF(a)(x - a)$$

is the best affine approximation to $F(x)$ for x near a.

   Also, as in the single variable case, differentiability implies continuity. We state this in the following theorem, the proof of which is left to the exercises.

Theorem 9.2.3. *If $F : D \to \mathbb{R}^q$ is differentiable at $a \in D$, then F is continuous at a.*

Example 9.2.4. Let F be the function from $\mathbb{R}^2$ to $\mathbb{R}^2$ defined by

$$F(x, y) = (x^2 + y^2, xy).$$

Show that F is differentiable at $(1, 2)$ and its differential is the linear function with matrix

$$A = \begin{pmatrix} 2 & 4 \\ 2 & 1 \end{pmatrix}.$$

Find the affine function which best approximates F near $(1, 2)$.

   Solution: With $a = (1, 2)$ and $h = (x - 1, y - 2) = (s, t)$, we have $F(a) = (5, 2)$ and

$$
\begin{aligned}
&F(a + h) - F(a) - Ah \\
&= ((1 + s)^2 + (2 + t)^2 - 5 - 2s - 4t, (1 + s)(2 + t) - 2 - 2s - t) \\
&= (s^2 + t^2, st)
\end{aligned}
$$

Thus, the error $F(a+h) - F(a) - Ah$ if $F(a+h)$ is approximated by $F(a) + Ah$ is

$$(s^2 + t^2, st).$$

Then,

$$||F(a + h) - F(a) - Ah||^2 = (s^2 + t^2)^2 + (st)^2 \le 2||h||^4.$$

This implies,

$$\frac{||F(a + h) - F(a) - Ah||}{||h||} \le \sqrt{2}||h||,$$

which has limit 0 as $h \to 0$. This shows that F is differentiable at $(1, 2)$ and that $dF(1, 2) = A$.

   The best affine approximation to $F(x, y)$ near $(1, 2)$ is

$$
\begin{aligned}
T(x, y) &= (5, 2) + \begin{pmatrix} 2 & 4 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x - 1 \\ y - 2 \end{pmatrix} \\
&= (5 + 2(x - 1) + 4(y - 2), 2 + 2(x - 1) + (y - 2)) \\
&= (-5 + 2x + 4y, -2 + 2x + y).
\end{aligned}
$$

The Differential Matrix

Let $F : D \to \mathbb{R}^q$ be a function with $D \subset \mathbb{R}^p$ and a an interior point of D. If F is differentiable at a, then it is easy to compute the matrix (c_{ij}) of its differential $dF(a)$. This is called the *differential matrix* of F at a. As usual, we will tend to ignore the technical difference between the linear function $dF(a)$ and its corresponding matrix (see Remark 8.4.13).

   We suppose that $F(x) = (f_1(x), f_2(x), \cdots, f_q(x))$, so that f_i is the ith coordinate function of F. For $j = 1, \cdots, p$, we apply (9.2.1) in the special case in

which h approaches 0 along the line $h = te_j$ – that is, along the jth coordinate axis. Since the vector expression in 9.2.1 converges to 0, the same thing is true of each of its coordinate functions. This means,

$$\lim_{t \to 0} \frac{f_i(a + te_j) - f_i(a) - c_{ij}t}{t} = 0,$$

which implies

$$c_{ij} = \lim_{t \to 0} \frac{f_i(a + te_j) - f_i(a)}{t}.$$

The limit that appears in this equation is just the partial derivative

$$\frac{\partial f_i}{\partial x_j}(a),$$

of f_i with respect to its jth variable at the point a. This is true for each i and each j. Thus, we have proved the following theorem.

Theorem 9.2.5. *If $F : D \to \mathbb{R}^q$ is differentiable at an interior point a of $D \subset \mathbb{R}^p$, then its differential at a is the linear function $dF(a) : \mathbb{R}^p \to \mathbb{R}^q$ with matrix*

$$\left(\frac{\partial f_i}{\partial x_j}(a) \right)_{ij} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(a) & \frac{\partial f_1}{\partial x_2}(a) & \cdots & \frac{\partial f_1}{\partial x_p}(a) \\ \frac{\partial f_2}{\partial x_1}(a) & \frac{\partial f_2}{\partial x_2}(a) & \cdots & \frac{\partial f_2}{\partial x_p}(a) \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \frac{\partial f_q}{\partial x_1}(a) & \frac{\partial f_q}{\partial x_2}(a) & \cdots & \frac{\partial f_q}{\partial x_p}(a) \end{pmatrix}. \qquad (9.2.2)$$

If F is defined and differentiable at all points of an open set $U \subset \mathbb{R}^p$, then we say that F is differentiable on U. Its differential dF is then a function on U whose values are linear transformations from $\mathbb{R}^p$ to $\mathbb{R}^q$. Equivalently, its differential matrix dF is a $q \times p$ matrix whose entries are functions on U.

Example 9.2.6. Assuming that the function F of Example 9.2.4 is differentiable everywhere, find its differential matrix. Verify that, at $a = (1, 2)$, it is the matrix A of the example.

Solution The coordinate functions for F are given by $f_1(x, y) = x^2 + y^2$ and $f_2(x, y) = xy$. The point a in this example is $a = (1, 2)$. The partial derivatives of f_1 and f_2 are

$$\frac{\partial f_1}{\partial x} = 2x, \quad \frac{\partial f_1}{\partial y} = 2y$$

$$\frac{\partial f_2}{\partial x} = y, \quad \frac{\partial f_2}{\partial y} = x.$$

Thus, the differential matrix at a general point (x, y) is

$$\begin{pmatrix} 2x & 2y \\ y & x \end{pmatrix}$$

At the particular point $a = (1, 2)$, this is

$$\begin{pmatrix} 2 & 4 \\ 2 & 1 \end{pmatrix}.$$

This is, indeed, the matrix A of Example 9.2.4.

A Condition for Differentiability

Since the vector function in (9.2.1) has limit 0 if and only if each of its coordinate functions has limit 0, we have the following theorem.

Theorem 9.2.7. *If $D \subset \mathbb{R}^p$ and $F = (f_1, \cdots, f_q) : D \to \mathbb{R}^q$ is a function, then F is differentiable at $a \in D$ if and only if, for each i, the coordinate function f_i is differentiable at a. In this case, the differential matrix dF is the matrix whose ith row is the differential df_i of the coordinate function f_i.*

This result allows us to reduce the proof of following theorem to the case $q = 1$.

Theorem 9.2.8. *Let $F = (f_1, \cdots, f_q) : U \to \mathbb{R}^q$ be a function defined on an open subset U of $\mathbb{R}^p$. If each first order partial derivative of each coordinate function f_i exists on U, then F is differentiable at each point of U where these partial derivatives are all continuous. Thus, if F is C^1 on all of U, then F is differentiable on all of U.*

Proof. By the previous theorem, it is enough to prove that each of the coordinate functions of F is differentiable at the point in question. Hence, it is enough to prove the theorem in the case $q = 1$. To complete the proof, we will prove the following statement by induction on p: If f is a real valued function defined on an open set $U \subset \mathbb{R}^p$ and each first order partial derivative of f exists on U, then f is differentiable at each point of U where all of these partial derivatives are continuous.

If $p = 1$, then the hypothesis implies, in particular, that f has a derivative at each point of U. For a function of one variable, this means the function is differentiable at each point of U. This completes the base case of the induction argument.

We now assume our statement is true for functions of p variables and let f be a function of $p+1$ variables. We write points of $\mathbb{R}^{p+1}$ in the form (x, y) with $x \in \mathbb{R}^p$ and $y \in \mathbb{R}$. For some $a = (a_1, \cdots, a_p) \in \mathbb{R}^p$ and $b \in \mathbb{R}$ we suppose (a, b) is a point of U at which the first order partial derivatives of f are all continuous.

If $h = (h_1, \cdots, h_p) \in \mathbb{R}^p$ and $k \in \mathbb{R}$, then

$$
\begin{aligned}
f(a+h, &b + k) - f(a, b)) \\
&= f(a + h, b) - f(a, b) + f(a + h, b + k) - f(a + h, b).
\end{aligned}
$$

If we set $g(x) = f(x, b)$ for x in an appropriate neighborhood of a in $\mathbb{R}^p$ and use the Mean Value Theorem in the last variable on the last two terms above, then this becomes

$$f(a + h, b + k) - f(a, b) = g(a + h) - g(a) + \frac{\partial f}{\partial y}(a + h, c)k, \qquad (9.2.3)$$

for some c between b and $b + k$.

Since g is a function of p variables which satisfies the hypotheses of the theorem, g is differentiable at a by our induction assumption. Hence, $dg(a)$ exists and

$$\lim_{h \to 0} \frac{g(a + h) - g(a) - dg(a)h}{||h||} = 0.$$

Because $||h|| \leq ||(h, k)||$ this implies

$$\lim_{(h,k) \to 0} \frac{g(a + h) - g(a) - dg(a)h}{||(h, k)||} = 0. \qquad (9.2.4)$$

Since $\dfrac{\partial f}{\partial y}$ is continuous at (a, b), $|k| \leq ||(h, k)||$, and $(a + h, c) \to (a, b)$ as $(h, k) \to (0, 0)$, we also have

$$\lim_{(h,k) \to 0} \frac{1}{||(h, k)||} \left(\frac{\partial f}{\partial y}(a + h, c) - \frac{\partial f}{\partial y}(a, b) \right) k = 0. \qquad (9.2.5)$$

Let v be the vector whose first p components are the components of $dg(a)$ and whose last component is $\dfrac{\partial f}{\partial y}(a, b)$. Then, by (9.2.3),

$$f(a+h, b + k) - f(a, b) - v \cdot (h, k)$$
$$= g(a + h) - g(a) - dg(a)h + \left(\frac{\partial f}{\partial y}(a + h, c) - \frac{\partial f}{\partial y}(a, b) \right) k, \qquad (9.2.6)$$

On combining (9.2.4), (9.2.5), and (9.2.6), we conclude that

$$\lim_{(h,k) \to (0,0)} \frac{f(a + h, b + k) - f(a, b) - v \cdot (h, k)}{||(h, k)||} = 0,$$

and, hence, that f is differentiable at (a, b) with differential v. This completes the induction and finishes the proof of the theorem. $\square$

Example 9.2.9. Show that the function $F : \mathbb{R}^2 \to \mathbb{R}^3$ defined by

$$F(x, y) = (x\,e^y, y\,e^x, xy)$$

is differentiable everywhere, and then find its differential matrix.

Solution: The first order partial derivatives of the coordinate functions of F exist and are continuous everywhere. Hence, F is differentiable everywhere by the previous theorem. Its differential matrix is

$$dF(x, y) = \begin{pmatrix} e^y & x\,e^y \\ y\,e^x & e^x \\ y & x \end{pmatrix}.$$

A Function Which is not Differentiable

The existence of the first order partial derivatives is not, by itself, enough to ensure that a function is differentiable. This is demonstrated by the next example.

Example 9.2.10. Show that the function f defined by

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

is not differentiable at $(0,0)$ even though its first order partial derivatives exist everywhere.

Solution: This is a rational function with a denominator which vanishes only at $(0,0)$. Hence, its first order partial derivatives exist everywhere except possibly at $(0,0)$. However f is identically 0 on both coordinate axes (that is, $f(x,0) = 0 = f(0,y)$. Hence, both $\dfrac{\partial f}{\partial x}$ and $\dfrac{\partial f}{\partial y}$ exist at $(0,0)$ and equal 0. However, f is clearly not differentiable at $(0,0)$, since it is not even continuous at this point (see Example 8.1.3).

Exercise Set 9.2

1. If $L : \mathbb{R}^p \to \mathbb{R}^p$ is a linear function, show that $dL = L$. In other words, if L has matrix A, then A is the differential matrix of the linear function $L(x) = Ax$.

2. Find the best affine approximation near $(0,0)$ to the function $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$F(x,y) = (xy - 2x + y + 1, x^2 + y^2 + x - 3y + 6).$$

3. If F is the function of the previous exercise, find the best affine approximation to F near $(1,-1)$.

4. Find the differential matrix for the function $G : R^+ \times R \to \mathbb{R}^3$ defined by

$$G(x,y) = (y \ln x, x\, e^y, \sin xy).$$

   Then find the best affine approximation to G at the point $(1, \pi)$.

5. Find the differential of the real valued function function $f(x,y,z) = xy^2 \cos xz$. Then find the best affine approximation to f at the point $(1,1,\pi/2)$.

6. Find the differential of the curve, $\gamma(t) = (\sin(2\pi t), \cos(2\pi t), t^2)$. Then find the best affine approximation to the curve γ at the point $t = 1$.

7. Prove that if f is a real valued function defined on an open interval containing a and if S is an affine function such that $f(a) = S(a)$ and

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

then $S(a + h) = f(a) + f'(a)h$.

8. Prove that if U is a neighborhood of 0 in $\mathbb{R}^p$ and if $F : U \to \mathbb{R}^q$ is a function such that $F(0) = 0$, then F is differentiable at 0 with $dF = 0$ if and only if $\lim_{x \to 0} ||F(x)||/||x|| = 0$.

9. Prove Theorem 9.2.3. That is, prove that if a function is differentiable at a point in its domain, then it is continuous at that point.

10. Does the function defined by

$$f(x, y) = \begin{cases} \dfrac{x^3}{x^2 + y^2} & \text{if} \quad (x, y) \neq (0, 0) \\ 0 & \text{if} \quad (x, y) = (0, 0) \end{cases}$$

have first order partial derivatives at every point of $\mathbb{R}^2$? Is this function differentiable at $(0, 0)$? Give reasons for your answers.

11. If $f : \mathbb{R}^p \to \mathbb{R}$ is differentiable at $a \in \mathbb{R}^p$, then show that, for each $h \in \mathbb{R}^p$, the function $g : \mathbb{R} \to \mathbb{R}$ defined by $g(t) = f(a + th)$ has a derivative at $t = 0$. Can you compute it in terms of $df(a)$ and h?

12. Prove that a function $F : \mathbb{R}^p \to \mathbb{R}^q$ is affine if and only if it is differentiable everywhere and its differential matrix is constant.

9.3   The Chain Rule

The differential of a function of several variables has properties similar to those of the derivative of a real valued function of a single variable. The simplest of these are stated in the following theorem, whose proof is left to the exercises.

Theorem 9.3.1. *Suppose F and G are functions defined on an open set $U \subset \mathbb{R}^p$, with values in $\mathbb{R}^q$, and c is a scalar. If F and G are differentiable at a point $x \in U$, then*

(a) cF is differentiable at x and $d(cF)(x) = cdF(x)$; and

(b) $F + G$ is differentiable at x and $d(F + G)(x) = dF(x) + dG(x)$.

A result which is more difficult to prove, but is of great importance is the chain rule for functions of several variables. The proof becomes considerably simpler if we reformulate the concept of differentiability in the following way.

An Equivalent Formulation of Differentiability

If f is a real valued function defined on an open interval containing the point $a \in \mathbb{R}$, then we can always express $f(a + h) - f(a)$ for h near but not equal to 0 in the following way:

$$f(a + h) - f(a) = q(h)h, \qquad\qquad (9.3.1)$$

where $q(h)$ is just the difference quotient

$$q(h) = \frac{f(a + h) - f(a)}{h}.$$

Of course, f is differentiable at a if and only if q has a limit as $h \to 0$. The derivative is then defined to be this limit. The function q becomes continuous at 0 if it is given the value $f'(a)$ at $h = 0$ and then (9.3.1) holds at $h = 0$ as well as at all nearby points. In fact, the differentiability of f at a is equivalent to the existence of a function q which satisfies (9.3.1) and is continuous at $h = 0$. This suggests the following reformulation of the definition of differentiability.

Theorem 9.3.2. *Let F be a function defined on an open set $U \subset \mathbb{R}^p$ with values in $\mathbb{R}^q$ and let a be a point of U. Then F is differentiable at a if and only if there is a $q \times p$ matrix valued function $Q(h)$, defined in a neighborhood of 0, such that Q is continuous at 0 and $F(a+h) - F(a)$ is the vector-matrix product*

$$F(a + h) - F(a) = Q(h)h$$

for all h in a neighborhood of 0. If this condition holds, then $dF(a) = Q(0)$.

Proof. Suppose a matrix Q with the required properties exists on some neighborhood V of 0. Then, for $h \in V$,

$$\frac{F(a + h) - F(a) - Q(0)h}{||h||} = \frac{Q(h)h - Q(0)h}{||h||} = \frac{(Q(h) - Q(0))h}{||h||}.$$

This expression has norm less than or equal to $||Q(h) - Q(0)||$ which converges to 0 as $h \to 0$, since Q is continuous at 0. Thus, F is differentiable and its differential matrix is $Q(0)$.

Conversely, suppose F is differentiable at a. If we set

$$\epsilon(h) = F(a + h) - F(a) - dF(a)h.$$

Then ϵ is a function on a neighborhood of 0 with values in $\mathbb{R}^q$ and

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

If, when written out in terms of coordinate functions, $\epsilon = (\epsilon_1, \epsilon_2, \cdots, \epsilon_q)$, and

$h = (h_1, h_2, \cdots, h_p)$, then we define a $q \times p$ matrix $\Delta(h)$ by

$$\Delta(h) = ||h||^{-2} \begin{pmatrix} \epsilon_1 h_1 & \epsilon_1 h_2 & \cdots & \epsilon_1 h_p \\ \epsilon_2 h_1 & \epsilon_2 h_2 & \cdots & \epsilon_2 h_p \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \epsilon_q h_1 & \epsilon_q h_2 & \cdots & \epsilon_q h_p \end{pmatrix}.$$

This is a matrix valued function of h, defined on a neighborhood of 0, except at 0 itself. Moreover, if we define this function to be 0 when $h = 0$, then it becomes continuous at $h = 0$, since

$$\frac{|\epsilon_i(h)h_j|}{||h||^2} \le \frac{||\epsilon(h)|| \, ||h||}{||h||^2} = \frac{||\epsilon(h)||}{||h||},$$

and this has limit 0 as $h \to 0$. Note also that if we apply the matrix $\Delta(h)$ to the vector h, the result is

$$\Delta(h)h = \epsilon(h),$$

Thus, if we set

$$Q(h) = dF(a) + \Delta(h),$$

then Q is continuous at $h = 0$, $Q(0) = dF(a)$, and

$$F(a + h) - F(a) = dF(a)h + \epsilon(h) = dF(a)h + \Delta(h)h = Q(h)h.$$

This completes the proof. $\square$

The Chain Rule

After the above reformulation of differentiability, the chain rule has a simple proof.

Theorem 9.3.3. *Let U and V be open subsets of $\mathbb{R}^r$ and $\mathbb{R}^p$. respectively, and let $G : U \to \mathbb{R}^p$ and $F : V \to \mathbb{R}^q$ be functions with $G(U) \subset V$. Suppose $a \in U$, G is differentiable at a, and F is differentiable at $b = G(a)$. Then $F \circ G$ is differentiable at a and*

$$d(F \circ G)(a) = dF(G(a))dG(a).$$

Proof. By the previous theorem, there are matrix valued functions Q_G and Q_F, defined in neighborhoods of 0 in $\mathbb{R}^r$ and $\mathbb{R}^p$, respectively, each continuous at 0, with $Q_F(0) = dF(b)$, $Q_G(0) = dG(a)$, and such that

$$G(a + h) - G(a) = Q_G(h)h \quad \text{and} \quad F(b + k) - F(b) = Q_F(k)k$$

for h and k in appropriate neighborhoods of 0. Then, since $G(a) = b$,

$$F \circ G(a + h) - F \circ G(a) = F(b + Q_G(h)h) - F(b) = Q_F(Q_G(h)h)Q_G(h)h.$$

Since Q_G and Q_F are both continuous at 0, we have

$$\lim_{h \to 0} Q_F(Q_G(h)h)Q_G(h) = Q_F(0)Q_G(0) = dF(b)dG(a) = dF(G(a))dG(a).$$

Thus, if we choose $Q_{F \circ G}(h)$ to be $Q_F(Q_G(h)h)Q_G(h)$, it satisfies the conditions of the previous theorem with F replaced by $F \circ G$ and, hence, by that theorem, $d(F \circ G)(a)$ exists and equals $dF(G(a))dG(a)$. ∎

Example 9.3.4. Let $f(x, y)$ be a real valued function of two variables and let

$$\phi(r, s, t) = f(r(s + t), r(s - t)).$$

Find $d\phi(1, 2, 1)$ if $\dfrac{\partial f}{\partial x}(3, 1) = 4$ and $\dfrac{\partial f}{\partial y}(3, 1) = -5$.

Solution: The function ϕ is just $f \circ G$, where $G : \mathbb{R}^3 \to \mathbb{R}^2$ is defined by

$$G(r, s, t) = (r(s + t), r(s - t)).$$

We have $G(1, 2, 1) = (3, 1)$ and

$$dG(1, 2, 1) = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.$$

Thus, $d\phi(1, 2, 1) = dF(G(1, 2, 1))dG(1, 2, 1)$ is

$$\left(\frac{\partial f}{\partial x}(3, 1), \quad \frac{\partial f}{\partial y}(3, 1) \right) \begin{pmatrix} 3 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix}$$

$$= \begin{pmatrix} 4, & -5 \end{pmatrix} \begin{pmatrix} 3 & 1 & 1 \\ 1 & 1 & -1 \end{pmatrix} = (7, -1, 9).$$

Example 9.3.5. If $F(x, y) = (f_1(x, y), f_2(x, y))$ is a differentiable function from $\mathbb{R}^2$ to $\mathbb{R}^2$ and we define $G : \mathbb{R}^2 \to \mathbb{R}^2$ by $G(s, t) = F(s^2 + t^2, s^2 - t^2)$, find an expression for the differential matrix of G in terms of the partial derivatives of f_1 and f_2.

Solution: The function G is $F \circ H$ where $H(s, t) = (s^2 + t^2, s^2 - t^2)$. The differential matrices of F and H are

$$dF = \begin{pmatrix} \dfrac{\partial f_1}{\partial x} & \dfrac{\partial f_1}{\partial y} \\ \dfrac{\partial f_2}{\partial x} & \dfrac{\partial f_2}{\partial y} \end{pmatrix} \text{ and } dH = \begin{pmatrix} 2s & 2t \\ 2s & -2t \end{pmatrix}.$$

By the chain rule,

$$dG(s, t) = d(F \circ H)(s, t) = dF(H(s, t))dH(s, t)$$

$$= \begin{pmatrix} 2s \left(\dfrac{\partial f_1}{\partial x} + \dfrac{\partial f_1}{\partial y} \right) & 2t \left(\dfrac{\partial f_1}{\partial x} - \dfrac{\partial f_1}{\partial y} \right) \\ 2s \left(\dfrac{\partial f_2}{\partial x} + \dfrac{\partial f_2}{\partial y} \right) & 2t \left(\dfrac{\partial f_2}{\partial x} - \dfrac{\partial f_2}{\partial y} \right) \end{pmatrix},$$

where the partial derivatives of f_1 and f_2 are to be evaluated at the point $H(s, t) = (s^2 + t^2, s^2 - t^2)$.

Differential of an Inner Product

The following theorem is a nice application of the chain rule.

Theorem 9.3.6. *Suppose F and G are functions defined in a neighborhood of a point $a \in \mathbb{R}^p$ and with values in $\mathbb{R}^q$. If F and G are both differentiable at a, then $F \cdot G$ is also differentiable at a and*

$$d(F \cdot G)(a) = G(a)dF(a) + F(a)dG(a),$$

where each of the products on the right is the matrix product of a $1 \times q$ times a $q \times p$ matrix.

Proof. Let $H : \mathbb{R}^{2q} \to \mathbb{R}$ be defined by

$$H(u,v) = u \cdot v,$$

where, if $u = (u_1, \cdots, u_q)$ and $v = (v_1, \cdots, v_q)$ are vectors in $\mathbb{R}^q$, then (u,v) denotes the vector $(u_1, \cdots, u_q, v_1, \cdots, v_q)$ in $\mathbb{R}^{2q}$.

Now $F \cdot G = H \circ (F, G)$, where (F, G) denotes the function with values in $\mathbb{R}^{2q}$ whose first q coordinate functions are the coordinate functions of F and whose last q coordinate functions are the coordinate functions of G.

The function H is differentiable everywhere because its coordinate functions $u_i v_i$ have continuous partial derivatives everywhere. That is,

$$\frac{\partial u_i v_i}{\partial u_i} = v_i, \quad \frac{\partial u_i v_i}{\partial v_i} = u_i,$$

and all other first order partial derivatives are zero. This means that its differential is the $1 \times 2q$ matrix

$$(v_1, \cdots, v_q, u_1, \cdots, u_q).$$

Since F and G are differentiable at a, the coordinate functions of both are all differentiable at a. This implies that the function (F, G) is differentiable at a, since each of its coordinate functions is a coordinate function of F or a coordinate function of G. Furthermore,

$$d(F, G)(a) = \begin{pmatrix} dF(a) \\ dG(a) \end{pmatrix},$$

where the matrix on the right has its first q rows the rows of $dF(a)$ and its last q rows the rows of $dG(a)$.

By the chain rule,

$$d(F \cdot G)(a) = dH(F(a), G(a))d(F, G)(a)$$
$$= (G(a), F(a)) \begin{pmatrix} dF(a) \\ dG(a) \end{pmatrix}$$
$$= G(a)dF(a) + F(a)dG(a).$$

$\square$

Dependent Variable Notation

A notation that is often used in connection with differentiation and specifically the chain rule is one which emphasizes the variables in a problem, some of which depend on others through functional relations, but which de-emphasizes the functions defining these relations. In this notation, a function F of p variables with values in $\mathbb{R}^q$ determines a vector of q dependent variables

$$u = (u_1, u_2, \cdots, u_q)$$

which depend on a vector of p variables

$$x = (x_1, x_2, \cdots, x_p)$$

through the relation $u = F(x)$. The differential matrix is then the matrix

$$\left(\frac{\partial u_i}{\partial x_j}\right)_{ij} = \begin{pmatrix} \dfrac{\partial u_1}{\partial x_1} & \dfrac{\partial u_1}{\partial x_2} & \cdots & \dfrac{\partial u_1}{\partial x_p} \\ \dfrac{\partial u_2}{\partial x_1} & \dfrac{\partial u_2}{\partial x_2} & \cdots & \dfrac{\partial u_2}{\partial x_p} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \dfrac{\partial u_q}{\partial x_1} & \dfrac{\partial u_q}{\partial x_2} & \cdots & \dfrac{\partial u_q}{\partial x_p} \end{pmatrix}.$$

where $\dfrac{\partial u_i}{\partial x_j}$ is understood to be the partial derivative $\dfrac{\partial f_i}{\partial x_j}$ of the ith coordinate function of F evaluated at a generic point x of the domain of F.

Now the variables x_j themselves may depend on a vector of variables

$$t = (t_1, t_2, \cdots, t_r)$$

through a function G. The differential matrix for this relationship would be the matrix

$$\left(\frac{\partial x_j}{\partial t_k}\right)_{jk}.$$

Since the variables u_i depend on the variables x_j, which in turn depend on the variables t_k, the variables u_i also depend on the variables t_k (through the function $F \circ G$), and the differential matrix for this relationship is denoted

$$\left(\frac{\partial u_i}{\partial t_k}\right)_{ik}.$$

Using this notation, the chain rule becomes

$$\left(\frac{\partial u_i}{\partial t_k}\right)_{ik} = \left(\frac{\partial u_i}{\partial x_j}\right)_{ij} \left(\frac{\partial x_j}{\partial t_k}\right)_{jk}, \tag{9.3.2}$$

where the expression on the right is the product of the indicated matrices. This product will involve the variables x_j as well as the variables t_k and it is important to remember that the x_js are themselves functions of the variables t_k.

A Change of Variables

Example 9.3.7. If $u = f(x, y)$ expresses the variable u as a function of Cartesian coordinates (x, y) on an open subset of the plane, what is the relationship between the differential matrix of u as a function of (x, y) and its differential matrix as a function of the corresponding polar coordinates (r, θ), where $x = r \cos \theta$ and $y = r \sin \theta$.

Solution: The change of coordinate transformation $(x, y) = (r \cos \theta, r \sin \theta)$ has differential matrix

$$\begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}.$$

Thus,

$$\begin{pmatrix} \dfrac{\partial u}{\partial r} & \dfrac{\partial u}{\partial \theta} \end{pmatrix} = \begin{pmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial y} \end{pmatrix} \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix},$$

or

$$\frac{\partial u}{\partial r} = \cos \theta \frac{\partial u}{\partial x} + \sin \theta \frac{\partial u}{\partial y}$$

$$\frac{\partial u}{\partial \theta} = -r \sin \theta \frac{\partial u}{\partial x} + r \cos \theta \frac{\partial u}{\partial y}.$$

Exercise Set 9.3

1. If F is a function from an open subset U of $\mathbb{R}^p$ to $\mathbb{R}^q$ which is differentiable at a and if B is an $r \times q$ matrix, then show that $d(BF)(a) = B dF(a)$. Here, $BF(x)$ is the matrix B applied to the vector $F(x)$ and $B dF(a)$ is the product of the matrix B and the matrix $dF(a)$.

2. If $f(x, y)$ is a differentiable function of $(x, y) \in \mathbb{R}^2$, and $g(t) = f(tx, ty)$, for all $t \in \mathbb{R}$, find $g'(1)$ in terms of the partial derivatives of f.

3. An n-homogeneous function on $\mathbb{R}^2$ is a function that satisfies $f(tx, ty) = t^n f(x, y)$ for all $t \in \mathbb{R}$ and $(x, y) \in \mathbb{R}^2$. Show that a differentiable function on $\mathbb{R}^2$ is n-homogeneous if and only it satifies the differential equation

$$x \frac{\partial f}{\partial x} + y \frac{\partial f}{\partial y} = nf$$

at each $(x, y) \in \mathbb{R}^2$.

4. If f is a differentiable function on $\mathbb{R}$ and $g(x, y) = f(xy)$, show that

$$x \frac{\partial g}{\partial x} - y \frac{\partial g}{\partial y} = 0.$$

5. If f and g are twice differentiable functions on $\mathbb{R}$ and

$$h(x, y) = f(x - y) + g(x + y),$$

show that h satisfies the wave equation:

$$\frac{\partial^2 h}{\partial x^2} - \frac{\partial^2 h}{\partial y^2} = 0.$$

6. If u is a variable which is a differentiable function of (x, y) in an open set $U \subset \mathbb{R}^2$, if x and y are differentiable functions of $(s, t) \in V$ for an open set $V \subset \mathbb{R}^2$, and if $(x, y) \in U$ whenever $(s, t) \in V$, then use the chain rule to obtain expressions for $\dfrac{\partial u}{\partial s}$ and $\dfrac{\partial u}{\partial t}$ on V in terms of the partial derivatives of u with respect to x and y and the partial derivatives of x and y with respect to s and t.

7. Do the preceding exercise in the special case where

$$x = as + bt \quad \text{and} \quad y = cs + dt.$$

for some constants a, b, c, d.

8. If $F(x, y) = (f_1(x, y), f_2(x, y))$ is a differentiable function from $\mathbb{R}^2$ to $\mathbb{R}^2$ and we define $G : \mathbb{R}^2 \to \mathbb{R}^2$ by $G(s, t) = F(st, s+t)$, find an expression for the differential matrix of G in terms of the partial derivatives of f_1 and f_2.

9. If (x, y, z) are the Cartesian coordinates of a point in $\mathbb{R}^3$ and the spherical coordinates of the same point are r, θ, ϕ, then

$$x = r \cos\theta \sin\phi, \quad y = r\sin\theta \sin\phi, \quad z = r\cos\phi.$$

Let u be a variable which is a differentiable function of (x, y, z) on $\mathbb{R}^3$. Find a formula for the partial derivatives of u with respect to r, θ, ϕ in terms of its partial derivatives with respect to x, y, z.

10. Suppose U and V are open subsets of $\mathbb{R}^p$ and $F : U \to V$ has an inverse function $G : V \to U$. This means $F \circ G(y) = y$ for all $y \in V$ and $G \circ F(x) = x$ for all $x \in U$. Show that, if F is differentiable on U and G is differentiable on V, then $dF(x)$ is non-singular at each $x \in U$, and for each $x \in U$,
$$dF(x)^{-1} = dG(y) \quad \text{where} \quad y = F(x).$$

11. Show that if F is differentiable function on an open set $U \subset \mathbb{R}^p$ with values in $\mathbb{R}^q$, then the real valued function $||F(x)||^2$ on U has zero differential at x if and only if the vector $F(x)$ is orthogonal to each of the columns of $dF(x)$.

12. Prove Theorem 9.3.1.

13. If $f(x, y) = x^2 + y^2$ find a 1×2 matrix valued function Q which satisfies the conclusion of Theorem 9.3.2 for f.

14. In the proof of Theorem 9.3.3, the following fact is used twice: If $A(h)$ is a $q \times p$ matrix whose entries are functions of $h \in \mathbb{R}^p$ and if $A(h)$ is continuous at $h = 0$, then $\lim_{h \to 0} A(h)h = 0$, where $A(h)h$ is the result of the matrix $A(h)$ acting via vector-matrix product on the vector h. Prove that this limit is 0, as claimed.

9.4   Applications of the Chain Rule

The Gradient

The case $q = 1$ is of special interest in this discussion. In this case, we are dealing with a real valued function f on a domain $D \subset \mathbb{R}^p$. At any point x where the function f is differentiable, its differential matrix is a $1 \times p$ matrix – that is, a row vector

$$df = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_p} \right),$$

The resulting vector is called the *gradient* of f at x. It is sometimes denoted ∇f and sometimes denoted $\operatorname{grad} f$.

If $f(x_1, \cdots, x_p)$ is the function f with its argument written out in terms of coordinates, then a notation often used for df is

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \cdots \frac{\partial f}{\partial x_p} dx_p. \tag{9.4.1}$$

The interpretation of this is as follows: It is understood that df and the partial derivatives in this equation are evaluated at some generic point x of the domain of f. For each j, dx_j is the differential of the jth coordinate function x_j on $\mathbb{R}^p$. As such, it is the linear transformation from $\mathbb{R}^p$ to $\mathbb{R}$ which sends a vector $(v_1, \cdots, v_p) \in \mathbb{R}^p$ to its jth component v_j. As a row vector, it is the vector which has 1 as jth component and 0 for all other components. Earlier we called this vector e_j, but in the context of differentials it is common to call it dx_j. Equation 9.4.1 expresses the fact that, for each function f as above, df at a given point is a linear combination of the basis elements dx_j with the coefficients being the corresponding partial derivatives of f at that point.

Example 9.4.1. If $f(x, y, z) = z^2 + \sin xy$, find the gradient of f at a generic point (x, y, z) and at the particular point $(1, 0, 3)$.

Solution: At (x, y, z) the gradient of f is

$$df = (y \cos xy, x \cos xy, 2z).$$

At $(x, y, z) = (1, 0, 3)$ this is the vector $(0, 1, 6)$. In terms of the basis vectors dx, dy, dz, we have

$$df = y \cos xy \, dx + x \cos xy \, dy + 2z \, dz,$$

which, at $(x, y, z) = (1, 0, 3)$ is $dy + 6 \, dz$.

Directional Derivatives

We specify a *direction* in $\mathbb{R}^p$ by specifying a unit vector (vector of length 1) that points in this direction. For example, in $\mathbb{R}^2$ we may specify a direction by specifying an angle θ relative to the positive x axis, but this is equivalent to specifying the unit vector $(\cos\theta, \sin\theta)$ which points in this direction.

Given a function f, defined on a neighborhood of a point $a \in \mathbb{R}^p$, each first order partial derivative of f at a is defined by restricting f to a line through a parallel to one of the coordinate axes and differentiating the resulting function of one variable. However, there is nothing special about the coordinate axes. We may restrict f to a line in any direction through a and differentiate the resulting function of one variable. This leads to the concept of directional derivative.

Definition 9.4.2. Suppose f is a function defined in a neighborhood of $a \in \mathbb{R}^p$ and and u is a unit vector in $\mathbb{R}^p$. The *directional derivative* of f at a, in the direction u, is defined to be

$$D_u f(a) = \frac{d}{dt} f(a + tu)|_{t=0}$$

If f happens to be differentiable at a, then its directional derivatives all exist and are easily calculated.

Theorem 9.4.3. *Suppose f is a function defined in a neighborhood of $a \in \mathbb{R}^p$ and differentiable at a. If u is a unit vector in $\mathbb{R}^p$, then the directional derivative $D_u f(a)$ exists and*

$$D_u f(a) = df(a)u.$$

Proof. If $g : \mathbb{R} \to \mathbb{R}^p$ is defined by $g(t) = a + tu$, then $dg(t) = g'(t) = u$ and $D_u f(a) = d(f \circ g)(0)$. The chain rule implies that this exists and is equal to $df(a)dg(0) = df(a)u$. □

The directional derivative $D_u f(a)$ represents the rate of change of f as we pass through a in the direction specified by u. If this is positive, then it represents the rate of *increase* of f in the u direction as we pass through a.

The proof of the following theorem is left to the exercises.

Theorem 9.4.4. *Suppose f is a real valued function which is defined and differentiable in a neighborhood of $a \in \mathbb{R}^p$, and suppose that $df(a) \neq 0$. Then the gradient $df(a)$ points in the direction of greatest increase for f at a – that is, $D_u f(a)$ has its maximum value when the unit vector u is a positive scalar multiple of $df(a)$.*

Example 9.4.5. If $f(x, y) = 2 - x^2 - y^2$, find the direction of greatest increase of f at $(1, 1)$ and the rate of increase of f in this direction at $(1, 1)$.

Solution: The gradient of f is

$$df(x, y) = (-2x, -2y).$$

At $(1, 1)$ this is

$$df(1, 1) = (-2, -2).$$

A unit vector which points in the same direction is $u = (-1/\sqrt{2}, -1/\sqrt{2})$. The directional derivative in the direction of u is

$$D_u f(1,1) = df(1,1) \cdot u = \sqrt{2} + \sqrt{2} = 2\sqrt{2}.$$

The Derivative of a Curve

Another special case of importance in the study of differentials is the case of a curve in $\mathbb{R}^q$ – that is, a function

$$\gamma(t) = (\gamma_1(t), \gamma_2(t), \cdots, \gamma_q(t)),$$

defined on an interval $I \subset \mathbb{R}$, with values in $\mathbb{R}^q$. In this case, the differential matrix $d\gamma$, at an interior point of I is a $q \times 1$ matrix – that is, a column vector. This is the column vector obtained by transposing the vector

$$\gamma'(t) = (\gamma_1'(t), \gamma_2'(t), \cdots, \gamma_q'(t)).$$

of derivatives of the coordinate functions of γ.

If $a \in I$, the best affine approximation to $\gamma(t)$ for t near a is the function

$$\tau(t) = \gamma(a) + \gamma'(a)(t - a).$$

Assuming $\gamma'(a) \neq 0$, this is a parametric equation for a line through $b = \gamma(a)$ which is parallel to the vector $\gamma'(a)$. If one more restriction on the curve γ is met, this line will be called the *tangent line* to the curve at $\gamma(a)$.

The additional restriction needed on γ is that a is the only point on the interval I at which γ has the value b. Otherwise, the curve crosses itself at b and the tangent line to the curve at b is not well defined – there is a different tangent line for each branch of the curve passing through b (see Figure 9.1). In this case, we will say that b is a *crossing point* for γ. Crossing points can be eliminated by replacing the interval I with a smaller open interval, containing a, but no other points at which γ has the value $\gamma(a)$. In our continuing discussion of curves and their tangent lines, we will assume that $\gamma(a)$ is not a crossing point of γ. This assumption and the assumption that $\gamma'(a) \neq 0$ ensure that γ has a well defined tangent line at $\gamma(a)$.

Note that each point $\tau(t)$ which is on the tangent line and sufficiently close to $\gamma(a)$ determines a parameter value $t \in I$ and this, in turn, determines a point $\gamma(t)$ on the curve. The two points $\gamma(t)$ and $\tau(t)$ differ from one another by

$$\gamma(t) - \gamma(a) - \gamma'(a)(t - a)$$

and the norm of this vector approaches 0 faster than $t - a$ approaches 0 as $t \to a$. This justifies the claim that the curve γ and the line τ are tangent at the point $\gamma(a)$. Note, however, that this line of reasoning is only valid if $\gamma'(a) \neq 0$, since, otherwise, τ is constant and fails to determine a non-degenerate line.

If $\gamma'(a) \neq 0$, the vector

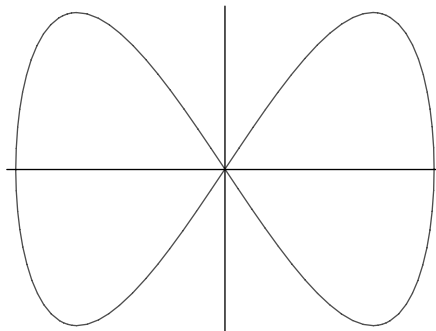$$T(a) = \frac{\gamma'(a)}{||\gamma'(a)||}$$

Figure 9.1: Curve With a Crossing Point

is a unit vector (a vector of length one) which is parallel to the tangent line at a. It is called the *tangent vector* to the curve at $\gamma(a)$.

The vector $\gamma'(a)$ is sometimes called the *velocity vector* of the curve at $\gamma(a)$, since it does represent velocity in the case where the curve is describing the motion of a body through space.

Example 9.4.6. The parameterized curve $\gamma(t) = (\cos t, \sin 2t)$, $0 < t < 2\pi$, passes through the origin. At the origin, find its velocity vector, tangent vector, and tangent line. Do the same problem if the domain of γ is restricted to $(0, \pi)$.

Solution: The origin is a crossing point for this curve (see Figure 9.1). The curve passes through the origin when $t = \pi/2$ and when $t = 3\pi/2$. Thus, there is no well defined velocity vector, tangent vector, or tangent line. If we restrict the domain of γ to the interval $(0, \pi)$, then the effect is to choose one branch of the curve and the crossing is eliminated. Then the curve passes through $(0, 0)$ only at $\pi/2$. We have

$$\gamma'(t) = (-\sin t, 2\cos 2t) \quad \text{and} \quad \gamma'(\pi/2) = (-1, -2).$$

Hence, the velocity vector at $(0, 0)$ is $\gamma'(\pi/2) = (-1, -2)$, the tangent vector at this point is $\dfrac{\gamma'(\pi/2)}{||\gamma'(\pi/2)||} = \left(\dfrac{-1}{\sqrt{5}}, \dfrac{-2}{\sqrt{5}}\right)$ and a parametric equation for the tangent line to this curve at $(0, 0)$ is

$$\tau(t) = (0, 0) + (t - \pi/2)(-1, -2) = (\pi/2 - t, \pi - 2t).$$

If we define the domain of γ to be $(\pi, 2\pi)$, then we are choosing the other branch of the curve – the one which passes through $(0, 0)$ at $t = 3\pi/2$. We leave the problem of finding the tangent line to the curve at this point to the exercises.

Higher Dimensional Tangent spaces

The following discussion is a higher dimensional version of the above discussion of curves and tangent lines. Suppose $p < q$, $U \subset \mathbb{R}^p$ is open, and $F : U \to \mathbb{R}^q$

is a smooth function. Since dF is a $q \times p$ matrix at each point of U and $p < q$, the maximal possible rank of dF is p. Suppose $a \in U$ is a point at which dF has rank p. Then the function

$$\Phi(x) = F(a) + dF(a)(x - a) \tag{9.4.2}$$

is an affine function of rank p (Definition 8.5.8). This implies that its image is a p-dimensional affine subspace of $\mathbb{R}^q$ (a translate of a p-dimensional linear subspace). Each point in this subspace which is sufficiently near $F(a)$ is $\Phi(x)$ for some $x \in U$ and, for such a point, there is a corresponding point $F(x)$ in the image of F. Now Φ is the best affine approximation to F near a and so the norm of

$$F(x) - \Phi(x) = F(x) - F(a) - dF(a)(x - a)$$

approaches 0 faster than $||x - a||$ approaches 0 as $x \to a$. This justifies calling the image of Φ the *tangent space* to the image of F at $F(a)$. At least, this is the case if a is the only point in U at which F has the value $F(a)$ (so that $F(a)$ is not a *crossing point* of F). The situation described in this discussion is important enough to warrant a definition.

A function F, defined on U, is *one to one* if there are no two distinct points of U at which F has the same value.

Definition 9.4.7. With $p < q$, let U be an open subset of $\mathbb{R}^p$ and $F : U \to \mathbb{R}^q$ be a one to one smooth function on U such that $dF(a)$ has rank p at each point $a \in U$. Then we will call the image S of F a smoothly parameterized p-surface in $\mathbb{R}^q$ and we will say that F is a smooth parameterization of S.

We define the tangent space of S at each $b = F(a) \in S$ to be the affine subspace of $\mathbb{R}^q$ which is the image of the function Φ of (9.4.2).

In the case where $p = q - 1$, a p-surface in $\mathbb{R}^q$ is called a *hypersurface* in $\mathbb{R}^q$ and its tangent space at $b = F(a)$ is its tangent *hyperplane* at b. If $q = 3$ and $p = 2$, then a 2-surface in $\mathbb{R}^3$ is just a surface and its tangent space at b is its tangent plane at b.

Example 9.4.8. . With $a = r_0 \cos \theta_0, b = r_0 \sin \theta_0$, and $r_0 > 0$, find the tangent plane at (a, b, r_0) to the cone in $\mathbb{R}^3$ parameterized by the function G defined by

$$G(r, \theta) = (r \cos \theta, r \sin \theta, r).$$

Is there a point on the cone where the tangent plane is not defined?

Solution: The differential dG at (r_0, θ_0) is

$$\begin{pmatrix} \cos \theta_0 & -r_0 \sin \theta_0 \\ \sin \theta_0 & r_0 \cos \theta_0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a/r_0 & -b \\ b/r_0 & a \\ 1 & 0 \end{pmatrix}.$$

If $r_0 \neq 0$, this matrix has rank 2. It defines a parameterized plane by

$$\Phi(r, \theta) = \begin{pmatrix} a \\ b \\ r_0 \end{pmatrix} + \begin{pmatrix} a/r_0 & -b \\ b/r_0 & a \\ 1 & 0 \end{pmatrix} \begin{pmatrix} r - r_0 \\ \theta - \theta_0 \end{pmatrix} \quad \text{or}$$

$$\Phi(r, \theta) = (ar/r_0 - b(\theta - \theta_0), \, br/r_0 + a(\theta - \theta_0), \, r).$$
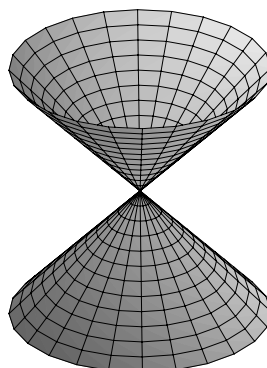
Figure 9.2: Cone in $\mathbb{R}^3$

There is no tangent plane to the curve at the origin. The differential of G at this point has rank 1 rather than rank 2 and the origin is a crossing point, which means that G does not satisfy the conditions of Definition 9.4.7. In fact, it is apparent from Figure 9.2 that there is no parametization of the cone in a neighborhood of the origin that will make it a smooth p-surface and no reasonable candidate for a tangent plane.

Level Sets

If $F : U \to \mathbb{R}^d$ is a function defined on an open subset U of $\mathbb{R}^q$, then a *level set* for F is a set of the form

$$S = \{y \in U : F(y) = c\}$$

where c is a constant vector in $\mathbb{R}^d$. By subtracting c from F, we can always arrange things so that S is the subset of U defined by the equation $F(y) = 0$.

Under these circumstances, it is often the case that locally (meaning near a given point $b \in S$) S can be represented as a smoothly parameterized surface of some dimension and its tangent space can be realized as the set of solutions y to the equation

$$dF(b)(y - b) = 0.$$

We will learn more about when this is true in the last section of this chapter. For now, we settle for a couple of preliminary results.

Theorem 9.4.9. *With F as above, let V be an open subset of $\mathbb{R}^p$ and $G : V \to \mathbb{R}^q$ a smooth function such that $G(V)$ is contained in a level set of F. Then*

$$dF(y)dG(x) = 0, \quad where \quad y = G(x),$$

for each $x \in V$.

Proof. If the image of G lies in a level set of F, then there is a constant $c \in \mathbb{R}^d$ such that

$$(F \circ G)(x) = c \quad \text{for all} \quad x \in V.$$

Then, by the chain rule,

$$0 = d(F \circ G)(x) = dF(G(x))dG(x).$$

$\square$

Example 9.4.10. Show that a curve γ in $\mathbb{R}^p$ of constant norm, $||\gamma(t)||$, has its tangent vector orthogonal to its position vector at each point.

Solution: If $||\gamma(t)||$ is constant, then so is $||\gamma(t)||^2$. This means that γ has its image in a level set of the function $f(x) = ||x||^2 = x \cdot x$. By the previous theorem, $df(x)d\gamma(t) = 0$ if $x = \gamma(t)$ is a point on the curve. This means that the velocity vector $\gamma'(t)$ is orthogonal to the gradient $2x$ of the function f at each point $x = \gamma(t)$ of the curve (see Exercise 9.4.6). Hence, $\gamma'(t)$ is orthogonal to $\gamma(t)$ at each t. Since the tangent vector $T(t) = \gamma'(t)/||\gamma'(t)||$ is a scalar times $\gamma'(t)$, it is also orthogonal to the position vector $\gamma(t)$ for each t.

How smooth is a level set for a smooth function $F : U \to \mathbb{R}^d$? Does it have a tangent space at some or all of its points? If so, does it resemble a curved version of its tangent space?

By Definition 9.4.7, In order for a level set S for F to have a tangent space at a point $b \in S$, there must be a neighborhood of b in which S is a smoothly parameterized p-surface. That is, near b, S must be the image of a smooth function $G : V \to \mathbb{R}^q$, with V an open subset of $\mathbb{R}^p$, and the rank of dG equal to p (the maximal rank possible) at each $a \in V$. Then the image of the affine function $\Phi(x) = b + dG(a)(x - a)$ is a p dimensional affine subspace of $\mathbb{R}^q$ (The tangent space to S at $b = G(a)$). Also, by the previous theorem

$$0 = dF(b)dG(a)(x - a) = dF(b)(\Phi(x) - b)$$

This means that the image of $\Phi - b$ is a linear subspace of $K = \ker dF(b)$. Hence, K has dimension at least p and it has dimension exactly p if and only if the image of $\Phi - b$ is equal to K. The dimension of K is p if and only if the rank of $dF(b)$ is $q - p$. Hence, we have proved:

Theorem 9.4.11. *With F as above and S a level set of F containing the point b, if in some neighborhood of b the space S is a smoothly parameterized p surface, and if $dF(b)$ has rank $q-p$ then the tangent space to S at b is the set of solutions y to the equation $dF(b)(y - b) = 0$. If the rank of $dF(b)$ is less than $q - p$, then the set of solutions to this equation contains the tangent space to S at b as a proper subset.*

Example 9.4.12. If $f(x, y, z) = x^2 + y^2 - z^2$ and $S = \{(x, y, z) : f(x, y, z) = 0\}$, show that at every point (a, b, c) on S, except at the origin, S is a smoothly parameterized 2-surface with tangent space defined in terms of the kernel of df

as in the previous theorem. Give the resulting equation for the tangent space. Then show that all of this fails at the origin.

Solution: The surface S is the same as the parameterized surface of Example 9.4.8 and Figure 9.2. By that example, S is a smoothly parameterized 2 surface near each such point except the origin. At $(a, b, c) \neq (0, 0, 0)$, df is $(2a, 2b, 2c)$. This has rank $1 = 3 - 2$. Therefore, by the previous theorem, S has a tangent space given by

$$2a(x - a) + 2b(y - b) + 2c(z - c) = 0.$$

At 0 df is the 0 matrix. Hence, the kernel of $df(0)$ is all of $\mathbb{R}^3$. Since S is the cone of Example 9.4.8, it is a 2 dimensional surface and it does not seem reasonable for it to have a 3-dimensional tangent space at a point. The problem is that S is not a smoothly parameterized surface in a neighborhood of the origin and, hence, does not have a tangent space there in the sense we are using the term in this text.

When can a level set of a function $F : U \to \mathbb{R}^d$ be represented as a smoothly parameterized p-surface where $q - p$ is the rank of $dF(b)$? That is the subject of the implicit function theorem discussed in the last section of this chapter. At this point, it is not clear that a level set of a smooth function F has a smooth parameterization near any of its points.

For some level sets the construction of a smooth parameterization of the right dimension is easy. This is true of a level set which arises as the graph of a function, as the next example shows.

Example 9.4.13. Show that if g is a smooth real valued function defined on $\mathbb{R}^2$, then each level set of the function $f : \mathbb{R}^3 \to \mathbb{R}$ defined by $f(x, y, z) = z - g(x, y)$ may be represented as a parameterized 2-surface.

Solution: Choose $G(x, y) = (x, y, g(x, y) + c)$. This is a smooth function from $\mathbb{R}^2$ to $\mathbb{R}^3$ with differential of rank 2 at each point and image equal to the level set $S = \{(x, y, z) : f(x, y, z) = c\}$.

Exercise Set 9.4

1. If $f(x, y, z) = x \sin z + y \cos z$ at each $(x, y, z) \in \mathbb{R}^3$, then find the gradient df of f at any point (x, y, z). What is $df(1, 2, \pi/4)$?

2. For the function $f(x, y) = x^2 + y^3 + xy$, find the gradient at the point $(1, 1)$, the direction of greatest ascent of f at this point, and a direction in which the rate of increase of this function is 0 (the answers to the last two questions should be unit vectors).

3. Find a parametric equation for the tangent line to the curve

$$\gamma(t) = (t^3, 1/t, e^{2t-2})$$

at the point where $t = 1$.

4. For the curve γ of Example 9.4.6, find a parametric equation of the tangent line to this curve at $(0,0)$ if the domain of $\gamma(t)$ is $\{t : \pi < t < 2\pi\}$.

5. Prove Theorem 9.4.4

6. Show that the gradient at $x \in \mathbb{R}^p$ of the function $g(x) = x \cdot x$ is the vector $2x$.

7. Let $\gamma : \mathbb{R} \to \mathbb{R}^p$ be a curve which passes through the origin in $\mathbb{R}^p$ at a point where its velocity vector is non-zero (that is, assume $\gamma(t_0) = 0$ and $\gamma'(t_0) \neq 0$ at some point $t_0 \in \mathbb{R}$). Prove that there is an interval I centered at t_0 such that $\|\gamma(t)\|$ is decreasing for $t < t_0$ and increasing for $t > t_0$. Hint: $\|\gamma\|$ is increasing (decreasing) wherever $\|\gamma\|^2 = \gamma \cdot \gamma$ is increasing (decreasing).

8. Find the tangent space at $(2,4,1)$ for the parameterized surface in $\mathbb{R}^3$ parameterized by the function $G : U \to \mathbb{R}^3$, where

$$U = \{(u,v) \in \mathbb{R}^2 : u > 0, v > 0\} \quad \text{and} \quad G(u,v) = (uv, u^2, v^2).$$

9. If a surface in $\mathbb{R}^3$ is defined by the equation $z = g(x,y)$, where g is a differentiable function of (x,y) in an open set U, find the equation for the tangent plane to this surface at a point (a,b,c) on the surface.

10. Find an equation for the tangent plane to the surface $z = x^2 \sin y + 2x$ at the point $(1,0,2)$. Also find parametric equations for a line which passes through this point and is perpedicular to the tangent plane.

11. Find the equation for the tangent plane to the cone $z = x^2 + y^2$ at the point $(1,2,5)$.

12. Show that for each point (a,b,c) on the surface $x^2 + y^2 + z^2 = 1$, there is a neighborhood of (a,b,c) in which the surface may be represented as a smoothly parameterized 2-surface. Hence, there is a tangent plane to this surface at every point.

13. Find an equation for the tangent plane to the surface of the previous problem at each point (a,b,c) on the surface.

14. Find an equation for the tangent plane to the surface $x^2 + y^2 - z^2 = 1$ at each point (a,b,c) on the surface.

9.5 Taylor's Formula

In this section we discuss Taylor's formula in several variables and some of its applications.

The Formula

If a and x are points of $\mathbb{R}^p$, then a parameterized line passing through a and x is given by

$$\gamma(t) = a + t(x - a)$$

Note $\gamma(0) = a$ and $\gamma(1) = x$. The *line segment* joining a to x is the closed interval $[a, x]$ on this line defined by

$$[a, x] = \{a + t(x - a) : t \in [0, 1]\}.$$

Let f be a real valued function defined on an open subset $U \subset \mathbb{R}^p$ and suppose that all partial derivatives of f through degree n exist on U and are themselves differentiable on U. If $a, x \in U$ and the line segment joining a to x is contained in U, then we set $h = x - a$ and define a function g on an open interval I containing $[0, 1]$ by

$$g(t) = f(a + th).$$

The function g is $n + 1$ times differentiable on I (by the chain rule) and so g satisfies Taylor's Formula (Theorem 6.5.3):

$$g(t) = g(0) + g'(0)t + \frac{g''(0)}{2}t^2 + \cdots + \frac{g^{(n)}(0)}{n!}t^n + R_n(t), \qquad (9.5.1)$$

Where

$$R_n(t) = \frac{g^{(n+1)}(s)}{(n+1)!}t^{n+1} \qquad (9.5.2)$$

for some s between 0 and t.

Since $g(1) = f(a+h)$, to get a formula for $f(a+h)$ we need only set $t = 1$ in the above formula and then find expressions for the functions $g^k(0)$ and $g^{(n))}(c)$ in terms of f and its derivatives. This is not difficult for the first few terms:

$$g(0) = f(a)$$

$$g'(0) = df(a)h = \sum_{j=1}^{p} \frac{\partial f}{\partial x_j}(a)h_j \qquad (9.5.3)$$

$$g''(0) = h \cdot d^2 f(a)h = \sum_{i=1}^{p}\sum_{j=1}^{p} \frac{\partial^2 f}{\partial x_i \partial x_j}(a)h_i h_j$$

Here we have used $d^2 f(a)$ to stand for the matrix

$$\left(\frac{\partial^2 f}{\partial x_i \partial x_j}(a) \right)_{ij}.$$

If we apply this matrix to h, the result is a vector of length p and we may take the inner product of h with this vector. The result is the formula for $g''(0)$ in (9.5.3).

The kth derivative of g at 0 is

$$g^{(k)}(0) = \sum_{i_1=1}^{p} \cdots \sum_{i_k=1}^{p} \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(a) h_{i_1} \cdots h_{i_k}. \tag{9.5.4}$$

We may think of this as a k dimensional array (a tensor of rank k)

$$d^k f(a) = \left(\frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(a) \right),$$

applied k times to the vector h. Here applying a tensor of rank k to a vector h yields a tensor of rank $k-1$ in the same way applying a matrix (tensor of rank 2) to a vector produces a vector (a tensor of rank 1). Thus, applying the tensor $d^k f(a)$ to the vector h produces the tensor of rank $k-1$:

$$d^k f(a) h = \left(\sum_{i_k=1}^{p} \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(a) h_{i_k} \right).$$

This has rank $k-1$ because we have summed over the index i_k, and so the result is no longer a function of this index. If we repeat this k times, we obtain the number (tensor of rank 0) expressed in (9.5.4). This is the result of applying $d^k f(a)$ a total of k times to the vector h and, hence, we will denote it by $d^k f(a) h^k$. Note, in particular, that $d^2 f(a) h^2$ is just $h \cdot d^2 f(a) h$.

If we use this notation for the derivatives of g in (9.5.1) and (9.5.2) the result is :

$$f(a+h) = f(a) + df(a)h + \frac{1}{2} d^2 f(a) h^2 + \cdots + \frac{1}{n!} d^n f(a) h^n + R_n, \tag{9.5.5}$$

where

$$R_n = \frac{1}{(n+1)!} d^{n+1} f(c) h^{n+1}, \tag{9.5.6}$$

for some point c on the line segment joining a to $a+h$. This is Taylor's formula in several variables. Expressed in terms of the variable $x = a + h$ (so that $h = x - a$), this becomes the formula of the following theorem.

Theorem 9.5.1. *Let f be a real valued function defined on an open set $U \subset \mathbb{R}^p$ and suppose all partial derivatives of f through degree n exist and are differentiable on U. If $a, x \in U$ and U contains the line segment $[a, x]$, then*

$$f(x) = f(a) + df(a)(x-a) + \frac{1}{2} d^2 f(a)(x-a)^2 + \cdots + \frac{1}{n!} d^n f(a)(x-a)^n + R_n,$$

where

$$R_n = \frac{1}{(n+1)!} d^{n+1} f(c)(x-a)^{n+1},$$

for some point c on the line segment $[a, x]$.

Example 9.5.2. Find the degree $n = 2$ Taylor's formula for $f(x, y) = \ln(x+y)$ at the point $a = (0, 1)$.

   Solution: We will need expressions for all partial derivatives of f through degree 3. However, these are easy to calculate because each nth order partial derivative of f is just the nth derivative of ln evaluated at $x+y$. Thus, $f(0, 1) = 0$, all first order partial derivatives of f are $(x + y)^{-1}$, which is 1 at $(0, 1)$. The second degree partial derivatives are all equal to $-(x + y)^{-2}$, which is -1 at $(x, y) = (0, 1)$. Each third degree partial derivative is $2(x + y)^{-3}$. Thus, the degree 2 Taylor formula for f is

$$\ln(x + y) = (1, 1) \begin{pmatrix} x \\ y - 1 \end{pmatrix}$$

$$-\frac{1}{2}(x, y - 1) \cdot \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y - 1 \end{pmatrix} + R_2$$

$$= x + y - 1 - \frac{1}{2}(x + y - 1)^2 + R_2,$$

where

$$R_2 = \frac{1}{3c^3}(x + y - 1)^3,$$

for some c between 1 and $x + y$. Here the expression in parentheses is the result of applying the rank three tensor which is 1 in every entry three times to the vector $(x, y - 1)$. The result is $(x + y - 1)^3$.

The Mean Value Theorem

The Mean Value Theorem for a real valued function on an open subset of $\mathbb{R}^p$ is a special case of Taylor's formula. In fact, if we apply Theorem 9.5.1 in the case $n = 0$, it yields:

$$f(x) = f(a) + R_0,$$

where

$$R_0 = df(c)(x - a)$$

for some c on the line segment joining a to x. Thus, we have proved,

Theorem 9.5.3. *If f is a differentiable real valued function on $B_r(a) \subset \mathbb{R}^p$, then for $x \in B_r(a)$ we have*

$$f(x) - f(a) = df(c)(x - a)$$

for some point c on the line segment joining a to x.

   As is the case for functions of one variable, the several variable mean value theorem has a host of applications. We point out two of these in the following corollaries, the proofs of which are left to the exercises.

Definition 9.5.4. A subset $A \subset \mathbb{R}^p$ is said to be *convex* if, for each pair of points $x, y \in A$, the line segment $[x, y]$ is also contained in A.
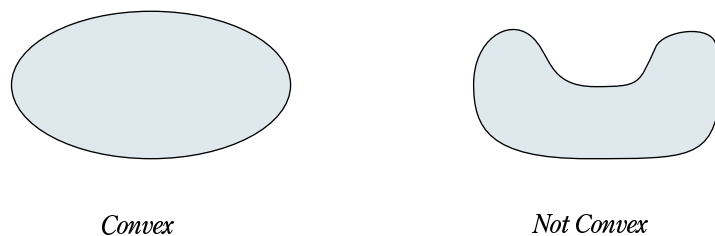
Convex   Not Convex

Figure 9.3: Convex and Nonconvex Sets

Figure 9.3 illustrates examples of a convex set and a set which is not convex.

Corollary 9.5.5. *Suppose U is an open convex set and f is a differentiable real valued function on U. If there is a number $M > 0$ such that $||df(x)|| \leq M$ for all $x \in U$, then*

$$|f(x) - f(y)| \leq M||x - y||$$

for all $x, y \in U$.

Corollary 9.5.6. *Let U be a connected open subset of $\mathbb{R}^p$ and f a differentiable function on U. If $df(x) = 0$ for all $x \in U$, then f is a constant.*

Max and Min

We know that if f is a real valued function of one variable, defined on an interval I, which has a local maximum or minimum at an interior point a of I, then either $f'(a)$ fails to exist or $f'(a) = 0$. We now discuss the several variable analogue of this result.

A function defined on a subset $D \subset \mathbb{R}^p$ is said to have a *local maximum* at $a \in D$ if there is a ball $B_r(a)$, centered at a, such that

$$f(x) \leq f(a) \quad \text{for all} \quad x \in D \cap B_r(a).$$

If a is an interior point of D, then r may be chosen so that $D_r(a) \subset D$ and then this inequality holds for all $x \in B_r(a)$. The concept of local minimum is defined in the same way, but with the inequality reversed.

Theorem 9.5.7. *If f is a function defined on $D \subset \mathbb{R}^n$ and if f has a local maximum or a local minimum at an interior point $a \in D$ at which f is differentiable, then $df(a) = 0$.*

Proof. Given any unit vector u, the function $g(t) = f(a + tu)$ is defined for all real numbers t in an open interval containing 0 and it has a local maximum (or minimum) at $t = 0$. By the chain rule, g is differentiable at 0 and its derivative at 0 is the directional derivative $df(a) \cdot u$ of f at a in the direction u. Since the derivative of g at 0 must be 0, we conclude that $df(a) \cdot u = 0$ for all unit vectors u and, hence, $df(a) = 0$. □

This theorem does not tell us that a function must have a local max or min at a point where df is 0. However, for functions of one variable, the second derivative test does give conditions that ensure that a local max or a local min occurs at a.

The second derivative test for functions of one variable says that if f is a real valued function on an interval I, then f has a local maximum at $a \in I$ if $f'(a) = 0$ and $f''(a) < 0$. It has a local minimum at a if $f'(a) = 0$ and $f''(a) > 0$. The analogue of this in several variables will be presented below, but it requires the concept of a *positive definite* matrix.

Definition 9.5.8. A $p \times p$ matrix A is said to be *positive definite* if $h \cdot Ah > 0$ for every non-zero vector $h \in \mathbb{R}^p$. It is negative definite if $h \cdot Ah < 0$ for every non-zero vector $h \in \mathbb{R}^p$.

Note that, in checking to see if a matrix is positive definite, we only need to check that $u \cdot Au > 0$ for every unit vector u in $\mathbb{R}^p$. This is because, if h is any non-zero vector, then $u = h/||h||$ is a unit vector and $h \cdot Ah = ||h||^2 u \cdot Au$, which is positive if and only if $u \cdot Au$ is positive.

It turns out that if a matrix is positive definite, then all nearby matrices are also positive definite. We will prove this using the concept of *operator norm* for a matrix (Definition 8.4.9). Recall that $||Ax|| \le ||A||||x||$ if x is a vector in $\mathbb{R}^p$, A is a $p \times p$ matrix, and $||A||$ is the operator norm of A.

Lemma 9.5.9. *If A is a positive definite $p \times p$ matrix, then there is is a positive number m such that if B is any $p \times p$ matrix with $||B - A|| < m/2$, then $u \cdot Bu \ge m/2$ for all unit vectors $u \in \mathbb{R}^p$ and, hence, B is also positive definite.*

Proof. The set of all unit vectors u is a closed bounded subset of $\mathbb{R}^p$. It is, therefore, compact. The function $g(u) = u \cdot Au$ is a continuous real valued function on this set and, hence, by Corollary 8.2.5, it takes on a minimum value m. Since $u \cdot Au > 0$ for all such u, we conclude that $m > 0$. Now it follows from the Cauchy-Schwarz inequality that

$$u \cdot (A - B)u \le ||u|| \, ||(A - B)u|| \le ||u||^2 ||A - B)|| = ||A - B||.$$

This implies

$$u \cdot Bu = u \cdot Au - u \cdot (A - B))u \ge m - ||A - B|| \qquad (9.5.7)$$

for all unit vectors u. Hence, if $||A - B|| < m/2$, then $u \cdot Bu > m/2$ for all unit vectors u, which implies that B is positive definite. $\square$

Theorem 9.5.10. *Let f be a real valued function defined on a neighborhood of $a \in \mathbb{R}^p$. Suppose the second order partial derivatives of f exist in this neighborhood and are continuous at a. If $df(a) = 0$ and $d^2 f(a)$ is positive definite, then f has a local minimum at a. If $df(a) = 0$ and $d^2 f(a)$ is negative definite, then f has a local maximum at a.*

Proof. We use Taylor's formula with $n = 1$. Since, $df(a) = 0$, It tells us that there is an $r > 0$ such that, for each $h \in B_r(0)$,

$$f(a + h) = f(a) + h \cdot d^2 f(c)h, \qquad (9.5.8)$$

for some c on the line segment joining a to $a + h$.

Assume $d^2 f(a)$ is positive definite. By the previous lemma, there is an $m > 0$ such that if

$$||d^2 f(a) - d^2 f(c)|| < m/2, \qquad (9.5.9)$$

then $d^2 f(c)$ is also positive definite.

Since the second order partial derivatives of f are continuous at a and since $||c - a|| \leq ||h||$, it follows from Theorem 8.4.11 that we can ensure (9.5.9) holds by choosing $||h||$ sufficiently small. Hence, there is an $\delta > 0$, with $\delta \leq r$, such that $||h|| < \delta$ implies that $d^2 f(c)$ is positive definite for all c on the line segment joining a to h. By 9.5.8, this implies that $f(a + h) > f(a)$. Thus, f has a local minimum at a in this case.

The case where $d^2 f(a)$ is negative definite follows from the above by simply applying the above result to $-f$. $\qquad \square$

Max/Min for Functions of 2 Variables

Let f be a function of 2 variables with second order partial derivatives which are defined in a neighborhood of $(x_0, y_0) \in \mathbb{R}^2$ and continuous at this point. The matrix $d^2 f$ has the form

$$\begin{pmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial y \partial x} & \dfrac{\partial^2 f}{\partial y^2} \end{pmatrix}.$$

Since the second order partial derivatives are continuous at (x_0, y_0), the cross partials are equal and so this matrix is symmetric (meaning it is its own transpose) at (x_0, y_0). There is a simple criteria for a symmetric 2×2 matrix to be positive definite. This is described in the next theorem, the proof of which is left to the exercises.

Theorem 9.5.11. *Let* $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ *be a symmetric* 2×2 *matrix and let* $\Delta = ac - b^2$ *be its determinant. Then*

(a) *A is positive definite if and only if* $\Delta > 0$ *and* $a > 0$;

(b) *A is negative definite if and only if* $\Delta > 0$ *and* $a < 0$;

(c) *if* $\Delta < 0$, *then there are vectors* $u, v \in \mathbb{R}^2$ *with* $u \cdot Au > 0$ *and* $v \cdot Av < 0$.

For a function f on $\mathbb{R}^2$, a point where the expression $u \cdot d^2 f(a)u$ is positive for some unit vectors u and negative for others is called a *saddle point*. At such
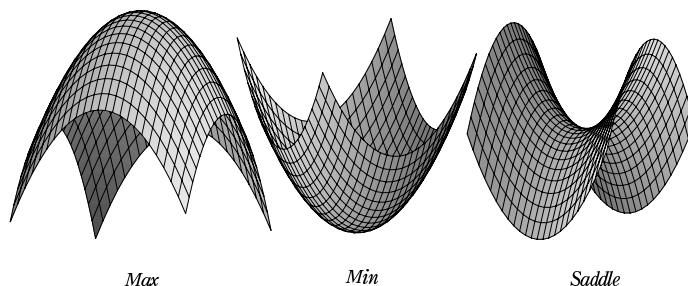
Figure 9.4: Surfaces With Max, Min, and Saddle Points.

a point, there will exist lines through a along which f has a local maximum at a and other lines through a along which f has a local maximum at a.

The previous theorem has the following corollary, the proof of which is also left to the exercises.

Corollary 9.5.12. *Let f be a function of 2 variables with second order partial derivatives which are defined in a neighborhood of $(x_0, y_0) \in \mathbb{R}^2$ and continuous at this point. Let $\Delta = \dfrac{\partial^2 f}{\partial x^2}\dfrac{\partial^2 f}{\partial y^2} - \left(\dfrac{\partial^2 f}{\partial x \partial y}\right)^2$ evaluated at (x_0, y_0). Then*

(a) f has a local minimum at (x_0, y_0) if $\Delta > 0$ and $\dfrac{\partial^2 f}{\partial x^2} > 0$ at (x_0, y_0);

(b) f has a local maximum at (x_0, y_0) if $\Delta > 0$ and $\dfrac{\partial^2 f}{\partial x^2} < 0$ at (x_0, y_0);

(c) if $\Delta < 0$, then f has a saddle point at x_0, y_0.

Example 9.5.13. Find all points where the function $f(x, y) = x^2 + xy + y^2 - 2x - 4y + 1$ has a local maximum and all points where it has a local minimum.

Solution: We have $df(x, y) = (2x + y - 2, x + 2y - 4)$. Thus, the only point at which $df(x, y) = 0$ is the point $a = (0, 2)$. This is the only possible point at which a local max or min can occur. The second differential $d^2 f(x, y)$ is the constant matrix

$$d^2 f(x, y) = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

This has determinant $\Delta = 3$. By the previous corollary, we conclude that $(0, 2)$ is a point at which a local minimum occurs and there is no local maximum.

Example 9.5.14. Find all points where the function

$$f(x, y) = x^2 + 3xy + y^2 - x - 4y + 5$$

has a local maximum, minimum, or saddle.

Solution: We have $df(x,y) = (2x + 3y - 1, 3x + 2y - 4)$. Thus, the only point at which $df(x,y) = 0$ is the point $a = (2,-1)$. This is the only possible point at which a local max or min can occur. The second differential $d^2f(x,y)$ is the constant matrix

$$d^2 f(x,y) = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}.$$

This has determinant $\Delta = -5$. Thus, $(2,-1)$ is a saddle point for f.

Lagrange Multipliers

Suppose U is an open subset of $\mathbb{R}^q$ and $f : U \to \mathbb{R}$ and $G : U \to \mathbb{R}^d$ are differentiable functions. The subject of Lagrange multipliers concerns the problem of finding points of local maximum or local minimum of f subject to the constraint that $G(x) = 0$. That is, we wish to find the points of local maximum and local minimum of f considered as a function on the level set $G(x) = 0$ for G. The following theorem applies to this problem. Its proof uses a corollary of the Implicit Function Theorem which will be proved at the end of this chapter.

Theorem 9.5.15. *With U, F and G as above, suppose that dG has rank d on U and S is the level set $S = \{x \in U : G(x) = 0\}$. If b is a point of relative max or min for f on S, then there is a linear transformation $\Lambda : \mathbb{R}^d \to \mathbb{R}$ such that $df(b) = \Lambda dG(b)$.*

Proof. In Corollary 9.7.3 we will prove that, under the above conditions, S is a smoothly parameterized p surface in a neighborhood of each point of S. We will assume this result here and we may as well assume that U is the neighborhood. Then there is an open subset V of $\mathbb{R}^p$ such that S is the image of a one-to-one differentiable function $H : V \to U$ with Rank $dH = p$ on V. Furthermore, $dG(H(a)) \circ dH(a) = 0$ for each $a \in V$. Thus, if $a \in V$ and $b = H(a)$, then the kernel of $dG(b)$ contains the image of $dH(a)$. However, the kernel of $dG(b)$ has dimension $q - d = p$ as does the image of $dH(a)$. It follows that the two subspaces of $\mathbb{R}^q$ are equal.

Since f has a local max or min on S at b, $f \circ H$ has a local max or min on V at a. This implies $df(b)dH(a) = d(f \circ H) = 0$. Since $dG(b)$ has rank d, its image is all of $\mathbb{R}^d$. Thus, for each $y \in \mathbb{R}^d$, there is an $x \in \mathbb{R}^q$ such that $dG(b)x = y$. We then set $\Lambda(y) = df(x)$. If x_1 is another vector in $\mathbb{R}^q$ with $dG(b)x_1 = y$, then $x - x_1 \in \ker dG(b) = \Im dH(b)$ and so $df(b)(x - x_1) = 0$. This means $df(b)x$ is the same vector no matter which vector x is chosen with $dG(b)x = y$. Thus, $\Lambda(y)$ is well defined by the condition

$$\Lambda(y) = df(x) \quad \text{whenever} \quad dG(b)x = y. \tag{9.5.10}$$

For vectors $y_1, y_2 \in \mathbb{R}^d$ we may choose x_1, x_2 such that $dG(b)x_i = y_i$. Then, $dG(x_1 + x_2) = dG(x_1) + dG(x_2) = y_1 + y_2$ and so

$$\Lambda(y_1 + y_2) = df(x_1 + x_2) = df(x_1) + df(x_2) = \Lambda(y_1) + \Lambda(y_2).$$

A similar argument shows that $\Lambda(kx) = k\Lambda(x)$ if k is a scalar. Thus, Λ is a linear transformation. By (9.5.10) Λ satisfies $df(b) = \Lambda dG(b)$. $\square$

The above result looks less mysterious if we write it out in terms of the coordinate functions of G. If $G = (g_1, \cdots, g_d)$, then S is the surface of vectors $x \in \mathbb{R}^q$ which satisfy the constraints

$$g_1(x) = 0, \cdots, g_d(x) = 0. \tag{9.5.11}$$

The theorem says that, if b is a point of S on which f has a local max or min on S, then there is a vector $\Lambda = (\lambda_1, \cdots, \lambda_d)$ such that

$$\frac{\partial f}{\partial x_k}(b) = \sum_{j=1}^{d} \lambda_j \frac{\partial f_j}{\partial x_k}(b) \quad \text{for} \quad k = 1, \cdots q. \tag{9.5.12}$$

Thus, to find candidates for points on S where a local max or min could occur, one should simultaneously solve the equations (9.5.11) and (9.5.12) for $x_1, \cdots, x_q, \lambda_1, \cdots, \lambda_d$. Note, this system of equations has $d + q$ equations and $d+q$ unknowns. The components $\lambda_1, \cdots, \lambda_d$ of Λ are called *Lagrange multipliers*.

Example 9.5.16. Find where the function $f(x, y, z) = 2xy + z$ attains its maximum and minimum values on $S = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 + z^2 = 1\}$.

   Solution: Since the unit sphere in $\mathbb{R}^3$ is compact and f is continuous, there are points on S where f attains its maximum and minimum values. We use the method of Lagrange multipliers, as described in the previous theorem to obtain candidates for these points. Here, $d = 1$ and $q = 3$ in (9.5.11) and (9.5.12).

   With $g(x, y, z) = x^2 + y^2 + z^2 - 1$, we must solve the system of equations:

$$g(x, y, z) = 0, \quad \frac{\partial f}{\partial x} = \lambda \frac{\partial g}{\partial x}, \quad \frac{\partial f}{\partial y} = \lambda \frac{\partial g}{\partial y}, \quad \frac{\partial f}{\partial z} = \lambda \frac{\partial g}{\partial z}.$$

These are the equations

$$x^2 + y^2 + z^2 = 1, \quad 2x = 2\lambda y, \quad 2y = 2\lambda x, \quad 1 = 2\lambda z.$$

The second and third equations yield $x = \lambda^2 x$ and $y = \lambda^2 y$. These hold if and only if $x = y = 0$ or $\lambda = \pm 1$. But $\lambda = \pm 1$ implies $x = \pm y$ and, together with the fourth equation, implies $z = \pm 1/2$. This, and the first equation imply $x = \pm\sqrt{3/8}$, $y = \pm\sqrt{3/8}$. Thus, the solutions of the above system of equations are

$$(0, 0, 1), \quad \left(\sqrt{3/8}, \sqrt{3/8}, 1/2\right), \quad \left(-\sqrt{3/8}, -\sqrt{3/8}, 1/2\right),$$

$$\left(-\sqrt{3/8}, \sqrt{3/8}, -1/2\right), \quad \text{and} \quad \left(\sqrt{3/8}, -\sqrt{3/8}, -1/2\right).$$

The values of f at these five points are, respectively, $1, 5/4, 5/4, -5/4, -5/4$. Thus, f has maximum value $5/4$ on S which is attained at $\left(\sqrt{3/8}, \sqrt{3/8}, 1/2\right)$, and at $\left(-\sqrt{3/8}, -\sqrt{3/8}, 1/2\right)$, while the minimum value is $-5/4$ attained at $\left(-\sqrt{3/8}, \sqrt{3/8}, -1/2\right)$ and $\left(\sqrt{3/8}, -\sqrt{3/8}, -1/2\right)$.

Exercise Set 9.5

1. Find the degree $n = 2$ Taylor's formula for $f(x, y) = x^2 + xy$ at the point $a = (1, 2)$.

2. Find the degree $n = 2$ Taylor's formula for $f(x, y) = e^{xy}$ at the point $a = (0, 0)$.

3. Suppose $a \in \mathbb{R}^p$ and f is a real valued function whose second order partial derivatives all exist and are continuous on $B_r(a)$. Also, suppose that the operator norm $\|d^2 f(x)\|$ of the matrix $d^2 f(x)$ is bounded by M on $B_r(a)$. Prove that

$$|f(x) - f(a) - df(a)(x - a)| \leq \frac{M}{2}\|x - a\|^2$$

for all $x \in B_r(a)$.

4. Prove Corollary 9.5.5.

5. Prove Corollary 9.5.6.

6. Show that the following form of the Mean Value Theorem is not true: If $F : \mathbb{R}^2 \to \mathbb{R}^2$ is a differentiable function and $a, b \in \mathbb{R}^2$, then there is a c on the line segment joining a to b such that $F(b) - F(a) = dF(c)(b - a)$. The problem here is that F is vector valued, not real valued.

7. Show that the following version of the Mean Value Theorem for vector valued functions is true: If U is an open set in $\mathbb{R}^p$ containing the line segment joining a to b and if $F : U \to \mathbb{R}^q$ is a differentiable function on U, then, for each vector $u \in \mathbb{R}^q$, there is a point c on the line segment joining a to b such that

$$u \cdot (F(b) - F(a)) = u \cdot dF(c)(b - a).$$

8. Find all points of relative maximum and relative minimum and all saddle points for
$$f(x, y) = 1 - 2x^2 - 2xy - y^2.$$

9. Find all points of relative maximum and relative minimum and all saddle points for
$$f(x, y) = y^3 + y^2 + x^2 - 2xy - 3y.$$

10. Prove Theorem 9.5.11.

11. Prove Corollary 9.5.12.

12. Show that it is possible for a function to have a relative minimum or maximum or a saddle at a point where both df and $d^2 f$ are 0.

13. Use the Lagrange multiplier method to find the maximal and minimal values of $f(x, y, z) = x - 2y + 3z$ on the sphere $x^2 + y^2 + z^2 = 1$.

9.6   The Inverse Function Theorem

If f is a real valued function of one variable which is $\mathcal{C}^1$ on an open interval containing a and if $f'(a) \neq 0$, then $f'(a)$ is either positive or negative. Because f' is continuous, $f'(x)$ will have the same sign as $f'(a)$ for all x in some neighborhood of a. This implies that f is strictly monotone in a neighborhood of a and, hence, has an inverse function. This inverse function is differentiable at $b = f(a)$ and $(f^{-1})'(b) = 1/f'(a)$ (Theorem 4.2.9). In this section we will prove an analogous result for a vector valued function F of several variables.

The condition that $f'(a) \neq 0$ is replaced in several variables by the condition that $dF(a)$ is a non-singular matrix (a matrix for which there is an inverse matrix). The conculsion that f is strictly monotone in some open interval containing a is replaced by the conclusion that F is a one to one function in some neighborhood of a in $\mathbb{R}^p$.

A function $F : V \to W$ is *one to one* on V if whenever $x, y \in V$ and $x \neq y$ then $F(x) \neq F(y)$. It is *onto* W if every $u \in W$ is $F(x)$ for some $x \in V$.

Definition 9.6.1. With F as above, we will say that F has a smooth local inverse near a if there are neighborhoods V of a and W of $F(a)$ such that F is a one to one function from V onto W and the function $F^{-1} : W \to V$, defined by $F^{-1}(u) = x$ if $F(x) = u$, is smooth on W.

In what follows (until the proof of the Inverse Function Theorem is complete), U will be an open subset of $\mathbb{R}^p$, $F : U \to \mathbb{R}^p$ a smooth (that is, $\mathcal{C}^1$) function on U. We will prove that F has a smooth local inverse near any point $a \in U$ at which its differential is non-singular.

The proof involves three steps. Assuming $dF(a)$ is non-singular: (1) we show F is one-to-one in a neighborhood of a; (2) we show F maps this neighborhood onto an open set; (3) we show the resulting inverse function is smooth and we calculate its differential.

One to One

The next theorem shows that our function F is necessarily one to one on some open ball centered at a point where dF is non-singular. In fact, it shows much more.

Theorem 9.6.2. *If $a \in U$ and $dF(a)$ is non-singular, then there is an open ball $B_r(a)$, centered at a, and a positive number M such that :*

(a) the matrix $dF(x)$ is non-singular for all $x \in B_r(a)$;

(b) $||x - y|| \leq M||F(x) - F(y)||$ for all $x, y \in B_r(a)$,

(c) the function F is one to one on $B_r(a)$.

Proof. Let B be an inverse matrix for $dF(a)$. Then $d(BF)(a) = BdF(a) = I$, where I is the $p \times p$ identity matrix (Exercise 9.3.1).

Let $G(x) = BF(x)$. Note that $dG(a) = I$, which is positive definite (since $u \cdot Iu = ||u||^2 = 1$ for every unit vector $u \in \mathbb{R}^p$). Hence, by Lemma 9.5.9, there is an $m > 0$ such that $dG(x)$ is also positive definite and, in fact,

$$m/2 \leq u \cdot dG(x)u \quad \text{whenever} \quad ||dG(x) - dG(a)|| < m/2$$

and u is a unit vector in $\mathbb{R}^p$.

The partial derivatives of the coordinate functions of F are all continuous and so the same thing is true of G. If follows from Theorem 8.4.11 that, given $m > 0$, there is an r such that $B_r(a) \subset U$ and

$$||dG(x) - dG(a)|| < m/2 \quad \text{whenever} \quad ||x - a|| < r.$$

Thus,

$$u \cdot dG(x)u \geq m/2 \tag{9.6.1}$$

for all $x \in B_r(a)$ and all unit vectors $u \in \mathbb{R}^p$. In particular, $dG(x)$ is positive definite and, hence, non-singular, for all $x \in B_r(a)$. Since $dF(x) = B^{-1}dG(x)$, this matrix is also non-singular for all $x \in B_r(a)$. This proves part (a).

Given two distinct points $x, y \in B_r(a)$, we set $k = ||y - x|| \neq 0$ and $u = (y - x)/k$. Then u is a unit vector and the function ϕ, defined by,

$$\phi(t) = u \cdot G(x + tu).$$

is a real valued differentiable function on an open interval containing $[0, k]$.

By the Mean Value Theorem, there is an $s \in [0, k]$ at which

$$k\phi'(s) = \phi(k) - \phi(0).$$

By the chain rule, $k\phi'(s) = ku \cdot dG(x + su)u$ and $\phi(k) - \phi(0) = u \cdot (G(y) - G(x))$. Thus,

$$ku \cdot dG(c)u = u \cdot (G(y) - G(x)),$$

where $c = x + su$. Then, by (9.6.1),

$$\begin{aligned} mk/2 \leq ku \cdot dG(c)u &= u \cdot (G(y) - G(x)) \\ &\leq ||u|||||G(y) - G(x)|| \leq ||B|||||F(y) - F(x)||, \end{aligned} \tag{9.6.2}$$

which, since $k = ||y - x||$, implies

$$||y - x|| \leq \frac{2||B||}{m}||F(x) - F(y)||.$$

This concludes the proof of part (b) if we set $M = 2||B||/m$.

Part (c) – that F is one to one on $B_r(a)$ – follows immediately from part (b) which shows that, for $x, y \in B_r(a)$, $x = y$ whenever $F(x) = F(y)$. $\square$

Open Mapping Theorem

An open map is a function F such that $F(V)$ is open whenever V is open.

Theorem 9.6.3. *With F as above, if dF is non-singular at every point of an open subset V of U, then $F : V \to \mathbb{R}^p$ is an open map.*

Proof. Given $a \in V$, set $b = F(a)$. We will show that $F(V)$ contains an open ball centered at b. If we can do this for every $a \in V$, then $F(V)$ is open. The same argument can be applied to every open subset of V and, hence, we may conclude that F is an open map.

The fact that $dF(a)$ is non-singular implies there is a open ball $B_r(a) \subset V$ for which the conclusions of the previous theorem hold. We will show that the image of this ball contains an open ball $B_\delta(b)$

Let r_1 be a positive number less than r. Then part (b) of the previous theorem implies that there is a positive number M such that

$$||x - y|| \leq M||F(x) - F(y)|| \quad \text{for all} \quad x, y \in \overline{B}_{r_1}(a).$$

Since $b = F(a)$, this implies, in particular, that

$$||F(x) - b|| \geq \frac{r_1}{M} \quad \text{whenever} \quad ||x - a|| = r_1. \qquad (9.6.3)$$

We set $\delta = \dfrac{r_1}{2M}$ and let v be any element of $B_\delta(b)$. If

$$g(x) = ||F(x) - v|| \quad \text{for} \quad x \in \overline{B}_{r_1}(a),$$

then our objective is to show that $g(u) = 0$ for some u in this ball.

We will first show that g takes on its minimum value at an interior point of $\overline{B}_{r_1}(a)$. It does take on a minimum value, since g is a continuous function on the compact set $\overline{B}_{r_1}(a)$ (Corollary 8.2.5). Thus, we need to show that it does not take on this minimum at a boundary point of $\overline{B}_{r_1}$.

If x is a boundary point of $\overline{B}_{r_1}$, then $||x - a|| = r_1$ and (9.6.3) applies. Also, $v \in B_\delta(b)$ means $||b - v|| < \dfrac{r_1}{2M}$. Thus,

$$g(x) = ||F(x) - v|| \geq ||F(x) - b|| - ||b - v|| \geq \frac{r_1}{2M} = \delta$$

on the boundary of $\overline{B}_{r_1}$.

Since $g(a) = ||F(a) - v|| = ||b - v|| < \delta$, the function $g(x)$ does not achieve its minimum value on the boundary of $\overline{B}_{r_1}(a)$. Hence, it must achieve its minimum value at a point u in the open ball $B_{r_1}(a)$. Then $g^2(x) = (F(x) - v) \cdot (F(x) - v)$ has a local minimum at u and, hence, its differential vanishes at u, by Theorem 9.5.7. By Theorem 9.3.6, its differential is $2(F(x) - v)dF(x)$. This expression vanishes at u if and only if $F(u) - v$ is orthogonal to all the columns of $dF(u)$. Since $dF(u)$ is non-singular, by Theorem 9.6.2 part (a), this can happen only if $F(u) - v = 0$. Hence, we have shown that each $v \in B_\delta(b)$ is the image under F of some $u \in B_r(a)$, as required. $\square$

The Inverse Function and its Differential

With F as above, if F is one-to-one with a non-singular differential on an open subset V of U then $\phi(V) = W$ is also open, by the previous theorem. In this situation, F has an inverse function $F^{-1} : W \to V$ defined by the condition that, for each $y \in W$, $F^{-1}(y)$ is the unique $x \in V$ such that $F(x) = y$.

Theorem 9.6.4. *With F, V and W as above, the inverse function $F^{-1} : W \to V$ is a smooth function on W with differential given by*

$$dF^{-1}(b) = (dF(a))^{-1} = (dF(F^{-1}(b)))^{-1} \qquad (9.6.4)$$

for each $b \in W$. Here $a = F^{-1}(b) \in V$.

Proof. Given $b \in W$ and $a = F^{-1}(a)$, we choose r as in Theorem 9.6.2 and we choose it small enough that $B_r(a) \subset V$. Then $F(B_r(a))$ is also open, by the previous theorem.

If $y \in F(B_r(a))$ and and $x = F^{-1}(y)$, then $x \in B_r(a)$. By the choice of r, the inequality in part (b) of Theorem 9.6.2 holds for x and a and says that

$$||F^{-1}(y) - F^{-1}(b)|| = ||x - a|| \leq M||y - b||.$$

This implies that F^{-1} is continuous at b. We calculate the differential of F^{-1} at b as follows:

The fact that F is differentiable at a means that if we set

$$\epsilon(x) = F(x) - F(a) - dF(a)(x - a), \qquad (9.6.5)$$

then

$$\lim_{x \to a} \frac{\epsilon(x)}{||x - a||} = 0.$$

If we apply the matrix $(dF(a))^{-1}$ to both sides of (9.6.5) and use $a = F^{-1}(b)$, $x = F^{-1}(y)$, the result is

$$dF(a)^{-1}\epsilon(y) = (dF(a))^{-1}(y - b) - (F^{-1}(y) - F^{-1}(b)),$$

or

$$F^{-1}(y) - F^{-1}(b) - dF(a)^{-1}(y - b) = -dF(a)^{-1}\epsilon(x).$$

If we set $K = ||(dF(a))^{-1}||$, then

$$\frac{||F^{-1}(y) - F^{-1}(b) - (dF(a))^{-1}(y - b)||}{||y - b||} \leq \frac{K||\epsilon(x)||}{||y - b||} \leq \frac{KM||\epsilon(x)||}{||x - a||}.$$

Since F^{-1} is continuous at b, $x = F^{-1}(y)$ approaches $a = F^{-1}(b)$ as y approaches b. Thus, the right side of the above inequality approaches 0 as $y \to b$. By definition, this means that F^{-1} is differentiable at b and

$$dF^{-1}(b) = (dF(a))^{-1} = (dF(F^{-1}(b)))^{-1}.$$

The partial derivatives of the coordinate functions of F^{-1} are the entries of its differential matrix dF^{-1}, which we just concluded is given by (9.6.4). Since, F^{-1} is continuous on W, the entries of $dF(x)$ (the partial derivatives of the coordinate functions of F) are continuous on V, and the determinant of $dF(x)$ is continuous and non-vanishing on V, we conclude that the partial derivatives of the coordinate functions of F^{-1} are continuous on W. This means that F^{-1} is $\mathcal{C}^1$, as claimed. This completes the proof. $\square$

The Inverse Function Theorem

The proof of the Inverse Function Theorem is now just a matter of combining the previous three theorems.

Theorem 9.6.5. *Let U be an open subset of $\mathbb{R}^p$ and $F : U \to \mathbb{R}^p$ a smooth function. If $a \in U$ and $\det dF(a) \neq 0$, then F has a smooth local inverse function near a, with differential given by (9.6.4).*

Proof. By Theorem 9.6.2, F is one-to-one with a non-singular differential in an open ball $B_r(a)$. By Theorem 9.6.3, the image of $B_r(a)$ under F is an open set W. Then F has an inverse function $F^{-1} : W \to B_r(a)$ and, by Theorem 9.6.4, the inverse function is smooth with differential as claimed. $\square$

Example 9.6.6. Find all points $a = (r, \theta) \in \mathbb{R}^2$ such that the polar change of coordinates function

$$F(r, \theta) = (r \cos \theta, r \sin \theta)$$

has a smooth local inverse near a. Find the inverse and its differential near one such point

Solution: The differential of F is

$$dF(r, \theta) = \begin{pmatrix} \cos \theta & -r \sin \theta \\ \sin \theta & r \cos \theta \end{pmatrix}.$$

The determinant of this matrix is r, and so dF is non-singular everywhere except at $r = 0$. By the previous theorem, this implies that F has a smooth local inverse near each $a = (r, \theta)$ with $r \neq 0$.

If we choose the point $a = (1, 0)$, then $F(a) = (1, 0)$. If V is the neighborhood of a defined by

$$V = \{(r, \theta) : r > 0, \ -\pi/2 < \theta < \pi/2\},$$

and W is the neighborhood of $b = F(a)$ defined by

$$W = \{(x, y) : x > 0\},$$

then

$$F^{-1}(x, y) = \left(\sqrt{x^2 + y^2}, \tan^{-1}(y/x) \right) \qquad (9.6.6)$$

defines the inverse function $F^{-1} : W \to V$.

The inverse matrix $(dF(r,\theta))^{-1}$ of the differential matrix $dF(r,\theta)$ of F is

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -r^{-1}\sin\theta & r^{-1}\cos\theta \end{pmatrix}.$$

By the previous theorem, this is the differential of the inverse function F^{-1} at the point $(x,y) = F(r,\theta)$. If express r and θ in terms of x and y using (9.6.6), we obtain

$$dF^{-1}(x,y) = \begin{pmatrix} \dfrac{x}{\sqrt{x^2+y^2}} & \dfrac{y}{\sqrt{x^2+y^2}} \\ -\dfrac{y}{x^2+y^2} & \dfrac{x}{x^2+y^2} \end{pmatrix}.$$

Note that the function F of the above example is definitely not one to one on all of $\mathbb{R}^2$ or on $\{(r,\theta) \in \mathbb{R}^2 : r \neq 0\}$ and so, as a function with either of these sets as domain, it does not have an inverse function. It is only when we restrict the domain of F to a set like the set V in the above example that it has an inverse function. What are some other sets V with the property that the restriction of F to the set V has an inverse function? This question is left to the exercises.

Exercise Set 9.6

1. According to the Inverse Function Theorem, near which points of $\mathbb{R}$ does the sin function have a smooth local inverse function? According to this theorem, what is the derivative of the inverse function when it exists?

2. Show that the function $F : \mathbb{R}^2 \to \mathbb{R}^2$ defined by $F(x,y) = (x^2 + y^2, xy)$ has a smooth local inverse near points (x,y) where $x \neq \pm y$. On the set $\{(x,y) : -x < y < x\}$ find the inverse function F^{-1} and identify its domain. Calculate the differential of this inverse function (1) directly, and (2) by using the Inverse Function Theorem. Verify that the two methods give the same answer.

3. Near which points of $\mathbb{R}^3$ does the spherical change of coordinates function

$$F(\rho,\theta,\phi) = (\rho\cos\theta\sin\phi, \rho\sin\theta\sin\phi, \rho\cos\phi)$$

have a smooth local inverse? What is the differential of the local inverse at those points, where it exists? To avoid tedious computation, express this in terms of (r,θ,ϕ) rather than in terms of the image variables $(x,y,z) = F(r,\theta,\phi)$.

4. Show that the system of equations

$$x = u^4 - u + uv + v^2$$
$$y = \cos u + \sin v$$

can be solved for (u,v) as a smooth function F of (x,y), in some neighborhood of $(0,0)$, in such a way that $(u,v) = (0,0)$ when $(x,y) = (0,1)$. What is the differential of the resulting function F at $(0,1)$?

5. Find a smooth local inverse function near $(1, \pi/2)$ for the function F of Example 9.6.6.

6. Find a smooth local inverse function near $(1, 2\pi)$ for the function F of Example 9.6.6. Note that this is different from the inverse function found in the example, even though the point $b = F(a)$ is the same in both cases.

7. Show that if U is a convex open subset of $\mathbb{R}^p$ and $F : U \to \mathbb{R}^p$ is a $\mathcal{C}^1$ function on U with a differential dF which is positive definite at every point of U, then F is one to one. Hint: examine the role played by the function ϕ in the proof of Theorem 9.6.2.

8. Show by example that the result of the previous problem is not true if U is only assumed to be connected, rather than convex. Hint: try the function $F(x, y) = (x^2 - y^2, 2xy)$ on $\mathbb{R}^2 \setminus \{0\}$.

9. Show that if $F = (f_1, f_2) : \mathbb{R}^3 \to \mathbb{R}^2$ is a $\mathcal{C}^1$ function and a is a point of $\mathbb{R}^3$ at which dF has rank 2, then there is a $\mathcal{C}^1$ function $f_3 : \mathbb{R}^3 \to \mathbb{R}$ such that $\Phi = (f_1, f_2, f_3) : \mathbb{R}^3 \to \mathbb{R}^3$ has a $\mathcal{C}^1$ inverse function near a.

10. Show that the condition that $dF(a)$ be non-singular is necessary in the Inverse Function Theorem, by showing that if a function F from a neighborhood of a in $\mathbb{R}^p$ to $\mathbb{R}^p$ is differentiable at a and has an inverse function at a which is differentiable at $F(a)$, then $dF(a)$ is non-singular.

11. Let $\gamma : I \to \mathbb{R}^3$ be a smooth parameterized curve, defined on an open interval I, and let t_0 be a point of I with $\gamma'(t_0) \neq 0$, Prove that there are neighborhoods $U \subset I$ of t_0 and V of $\gamma(t_0)$ and a pair f, g of $\mathcal{C}^1$ functions defined in V such that the image of U under γ is the set of solutions in V of the system of equations $f(x, y, z) = 0, g(x, y, z) = 0$. Hint: show that there is a $\mathcal{C}^1$ function F from a neighborhood of $(t_0, 0, 0)$ in $\mathbb{R}^3$ to $\mathbb{R}^3$ with $F(t, 0, 0) = \gamma(t)$ and with $dF(t_0, 0, 0)$ non-singular. Then apply the Inverse Function Theorem to F. The functions f and g are then two of the coordinate functions of F^{-1}.

12. If $F : \mathbb{R}^p \to \mathbb{R}^p$ is a $\mathcal{C}^1$ function, what can you say about F at a point of $\mathbb{R}^p$ where $||F||$ has a local minimum? How about a point where $||F||$ has a local maximum?

9.7   The Implicit Function Theorem

In this section we continue to develop consequences of the Inverse Function Theorem. The most notable of these is the Implicit Function Theorem. First we interpret the Inverse Function Theorem in the context of local systems of coordinates

Local Systems of Coordinates

Let F be a smooth function defined on an open subset U of $\mathbb{R}^p$ which has values in $\mathbb{R}^p$ and which has a smooth local inverse near a point $a \in U$. Then there is a neighborhood V of a and a neighborhood W of $b = F(a)$ such that $F : V \to W$ is one to one and onto and has a smooth inverse function $G = F^{-1} : W \to V$.

We define a change of coordinates for points in V as follows: If

$$F = (f_1, f_2, \cdots, f_p),$$

then we define new coordinates $(u_1, u_2, \cdots, u_p)$ for a point $x = (x_1, x_2, \cdots, x_p)$ in V by setting

$$u_i = f_i(x_1, x_2, \cdots, x_p) \quad \text{for} \quad i = 1, \cdots, p.$$

These new coordinates $u_1, \cdots u_p$ are smooth functions of the old coordinates $x_1, \cdots, x_p$ and, similarly, the old coordinates are smooth functions of the new coordinates since

$$x_j = g_j(u_1, u_2, \cdots, u_p) \quad \text{for} \quad j = 1, \cdots, p,$$

where g_j is the jth coordinate function of the inverse function G.

By subtracting the constant b from F, if necessary, we may assume that $F(a) = 0$ and W is a neighborhood of 0. This just makes the point a the origin in the new coordinate system.

A coordinate hyperplane (intersected with W) in the new coordinates is a set of the form

$$H_i = \{u \in W : u_i = 0\}.$$

In the original coordinates, this is the set

$$\{x \in V : f_i(x) = 0\}.$$

This means that the level set $\{x \in V : f_i(x) = 0\}$ for the function f_i looks like a smoothly deformed hyperplane (intersected with V). Similarly, the subset obtained by setting k of the coordinates $\{u_1, \cdots, u_p\}$ equal to zero is a $p - k$ dimensional subspace of $\mathbb{R}^p$. In the old coordinates this looks like a smoothly deformed $p - k$ subspace intersected with V. If $k = p - 1$ the result is a line through the origin in the new coordinates and a curve through a in the old coordinates.

Parameterizing a Curve

A key question raised in in the last subsection of Section 9.4 is: when does a level set for a smooth function from one Euclidean space to another locally have a smooth parameterization and, hence, a tangent space at each of its points? The following example gives an answer to this question in the case of a level set for a real valued function on $\mathbb{R}^2$. The method used in this example is a model for the proof of the Implicit Function Theorem, which will be proved next.

Example 9.7.1. Show that if $f : \mathbb{R}^2 \to \mathbb{R}^1$ is a smooth function and (a, b) is a point of $\mathbb{R}^2$ such that $f(a, b) = 0$ and $df(a, b) \neq 0$, then there is a neighborhood V of (a, b) in which $S = \{(x, y) : f(x, y) = 0\}$ is the image of a smooth parameterized curve. Find the tangent line to this curve at (a, b).

Solution: Since $df(a, b) \neq 0$, either $\dfrac{\partial f}{\partial x}$ or $\dfrac{\partial f}{\partial y}$ is non-zero at (a, b). Assume $\dfrac{\partial f}{\partial y}(a, b) \neq 0$ (the analysis in the other case is the same, but with the roles of x and y reversed). We define a function $H : \mathbb{R}^2 \to \mathbb{R}^2$ by

$$H(x, y) = (x, f(x, y)).$$

The differential matrix of this function is

$$\begin{pmatrix} 1 & 0 \\ \dfrac{\partial f}{\partial x} & \dfrac{\partial f}{\partial y} \end{pmatrix}.$$

which has determinant $\dfrac{\partial f}{\partial y}$. Since $\dfrac{\partial f}{\partial y}(a, b) \neq 0$, this matrix is non-singular at (a, b). Hence, there is a neighborhood V of (a, b), a neighborhood W of $(a, 0)$, and a smooth inverse function $H^{-1} : W \to V$ for H. We have

$$H^{-1}(x, 0) = (k(x), g(x)),$$

for some smooth real valued functions k, g, defined for all x with $(x, 0) \in W$. Then,

$$(x, 0) = H \circ H^{-1}(x, 0) = (k(x), f(k(x), g(x)) \quad \text{whenever} \quad (x, 0) \in W.$$

It follows that $k(x) = x$ and $f(x, g(x)) = 0$ for all such x. On the other hand, if $(x, y) \in V$ and $f(x, y) = 0$, then $H(x, y) = (x, 0)$ and so

$$(x, y) = H^{-1} \circ H(x, y) = H^{-1}(x, 0) = (x, g(x)).$$

Thus, $y = g(x)$. We conclude that, for $(x, y) \in V$, $f(x, y) = 0$ if and only if $y = g(x)$. Since, $(a, b) \in V$ and $f(a, b) = 0$, this means, in particular, that $g(a) = b$. Thus, we have proved that, near (a, b), S is the graph of the smooth function g and

$$\gamma(x) = (x, g(x))$$

is a smooth parameterization of S near (a, b).

The tangent line to S at (a, b) is given parametrically by

$$\tau(x) = (a, b) + \gamma'(a, b)(x - a)$$
$$= (a, b) + (1, g'(a))(x - a) = (x, b + g'(a)(x - a)),$$

where, since $f(x, g(x)) = 0$, the chain rule tells us that

$$g' = -\left(\frac{\partial f}{\partial y}\right)^{-1} \frac{\partial f}{\partial x}.$$

The tangent line can also be described as the set of all (x, y) such that $(x-a, y-b)$ is orthogonal to the gradient of f at (a, b) – that is, all solutions to the equation

$$\frac{\partial f}{\partial x}(a, b)(x - a) + \frac{\partial f}{\partial y}(a, b)(y - b) = 0.$$

The Implicit Function Theorem

The proof of the Implicit Function Theorem follows exactly the same pattern as the solution to the preceding exercise.

The Implicit Function Theorem provides the answer to a very simple question: When can an equation of the form

$$F(x, y) = 0$$

be solved for y as a function of x? That is, when can we find a function g such that $F(x, g(x)) = 0$? We note several things about this problem:

1. The problem makes perfectly good sense if F is a real valued function of 2 real variables (as in the previous example), but it also makes sense if F is a vector valued function of variables x and y which are also vectors.

2. As was the case with the Inverse Function Theorem, we might expect that there are local solutions to this problem for (x, y) near a point (a, b) where $F(a, b) = 0$, even though global solutions may not be possible.

3. Whether such a local solution is possible near a given point may depend on conditions on the differential matrix of F at the point.

In the statement and the proof of the Implicit Function Theorem, we will need to deal with certain submatrices of the full differential matrix of a function F. In this regard, the following notation will be useful. If $f_1, f_2, \cdots, f_k$ are smooth functions defined on an open set U in some Euclidean space $\mathbb{R}^d$ (these may be some or all of the coordinate functions of a vector function F defined on U) and if $y_1, \cdots, y_m$ are some of the coordinates describing points in $\mathbb{R}^d$), then we set

$$\frac{\partial(f_1, \cdots, f_k)}{\partial(y_1, \cdots, y_m)} = \begin{pmatrix} \dfrac{\partial f_1}{\partial y_1} & \dfrac{\partial f_1}{\partial y_2} & \cdots & \dfrac{\partial f_1}{\partial y_m} \\ \dfrac{\partial f_2}{\partial y_1} & \dfrac{\partial f_2}{\partial y_2} & \cdots & \dfrac{\partial f_2}{\partial y_m} \\ . & . & \cdots & . \\ . & . & \cdots & . \\ . & . & \cdots & . \\ \dfrac{\partial f_k}{\partial y_1} & \dfrac{\partial f_k}{\partial y_2} & \cdots & \dfrac{\partial f_k}{\partial y_m} \end{pmatrix}$$

If $F = (f_1, \cdots, f_q) : U \to \mathbb{R}^q$ is a function on a subset U of $\mathbb{R}^p$ with the coordinates in $\mathbb{R}^p$ labeled $x_1, \cdots, x_p$, then $\dfrac{\partial(f_1, \cdots, f_q)}{\partial(x_1, \cdots, x_p)}$ is just another notation

for dF. However, we will want to use this notation in cases where only some of the coordinate functions and/or some of the variables of F are used.

In the following theorem, $\mathbb{R}^{p+q}$ will be identified with $\mathbb{R}^p \times \mathbb{R}^q$ and points in this space will be expressed in the form $(x, y) = (x_1, \cdots, x_p, y_1, \cdots, y_q)$.

Theorem 9.7.2. *Let $U \subset \mathbb{R}^{p+q}$ be open, let $F = (f_1, \cdots, f_q) : U \to \mathbb{R}^q$ be a smooth function, and let (a, b) be a point of U with $F(a, b) = 0$. Also, suppose the square matrix*

$$\frac{\partial(f_1, \cdots, f_q)}{\partial(y_1, \cdots, y_q)}$$

is non-singular. Then there are neighborhoods $V \subset U$ of (a, b) and A of a and a smooth function $G : A \to \mathbb{R}^q$ such that $(x, G(x)) \in V$ for all $x \in A$, $G(a) = b$, and

$$F(x, y) = 0 \quad for \quad x, y \in V \quad if \ and \ only \ if \quad y = G(x).$$

Furthermore the differential of G on A is given by

$$dG = \frac{\partial(g_1, \cdots, g_q)}{\partial(x_1, \cdots, x_p)} = -\left(\frac{\partial(f_1, \cdots, f_q)}{\partial(y_1, \cdots, y_q)}\right)^{-1} \frac{\partial(f_1, \cdots, f_q)}{\partial(x_1, \cdots, x_p)}. \tag{9.7.1}$$

Proof. We will prove this by applying the Inverse Function Theorem to another function H, constructed from F. We define $H : U \to \mathbb{R}^p \times \mathbb{R}^q$ by

$$H(x, y) = (x, F(x, y)). \quad \text{for} \quad (x, y) \in U.$$

The function H is $\mathcal{C}^1$ on U because F is $\mathcal{C}^1$. The differential of H is

$$dH = \begin{pmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \dfrac{\partial f_1}{\partial x_1} & \cdots & \dfrac{\partial f_1}{\partial x_p} & \dfrac{\partial f_1}{\partial y_1} & \cdots & \dfrac{\partial f_1}{\partial y_q} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \dfrac{\partial f_q}{\partial x_1} & \cdots & \dfrac{\partial f_q}{\partial x_p} & \dfrac{\partial f_q}{\partial y_1} & \cdots & \dfrac{\partial f_q}{\partial y_q} \end{pmatrix}$$

with an identity matrix in the upper left $p \times p$ block and a 0 matrix in the upper right $p \times q$ block. The bottom q rows form the differential matrix dF for F. The determinant of dH is just the determinant of the lower right $q \times q$ block – that is, the determinant of $\dfrac{\partial(f_1, \cdots, f_q)}{\partial(y_1, \cdots, y_q)}$. This determinant is non-zero at (a, b) by hypothesis. Hence, dH also has a non-zero determinant at (a, b) and is, therefore, non-singular at this point.

By the Inverse Function Theorem (Theorem 9.6.5) there are neighborhoods $V \subset U$ of (a, b) and W of $H(a, b)$ such that H has a smooth inverse function $H^{-1} : W \to V$. We have

$$H^{-1}(x, 0) = (K(x), G(x)),$$

for some smooth functions K and G, defined on $A = \{x \in \mathbb{R}^p : (x, 0) \in W\}$ with values in $\mathbb{R}^q$. The set A is open because it is the inverse image of W under the continuous function $x \to (x, 0) : \mathbb{R}^p \to \mathbb{R}^p \times \mathbb{R}^q$. Furthermore,

$$(x, 0) = H \circ H^{-1}(x, 0) = (K(x), F(K(x), G(x))) \quad \text{whenever} \quad x \in A.$$

Thus, $K(x) = x$ and $F(x, G(x)) = 0$ for all $x \in A$. On the other hand, if $(x, y) \in V$ and $F(x, y) = 0$, then $H(x, y) = (x, 0)$ and so

$$(x, y) = H^{-1} \circ H(x, y) = H^{-1}(x, 0) = (x, G(x)).$$

Thus, $y = G(x)$. We conclude that if $(x, y) \in V$, then $F(x, y) = 0$ if and only if $y = G(x)$. This applies, in particular, when $(x, y) = (a, b)$ and so $G(a) = b$.

If we take the differential of both sides of the equation $F(x, G(x)) = 0$ the result is

$$\frac{\partial(f_1, \cdots, f_q)}{\partial(x_1, \cdots, x_p)} + \frac{\partial(f_1, \cdots, f_q)}{\partial(y_1, \cdots, y_q)} \frac{\partial(g_1, \cdots, g_q)}{\partial(x_1, \cdots, x_p)} = 0.$$

On solving this for $\dfrac{\partial(g_1, \cdots, g_q)}{\partial(x_1, \cdots, x_p)}$, we obtain (9.7.1).

$\square$

The Implicit Function Theorem leads to conditions under which a level set of a function has a smooth parameterization and, hence, a tangent space. This is the issue raised at the end of Section 9.4. This is also a key issue in the hypotheses of the theorem concerning the method of Largrange Multipliers (Theorem 9.5.15).

Corollary 9.7.3. *Let $U \subset \mathbb{R}^d$ be an open set and $F : U \to \mathbb{R}^q$ a smooth function. Suppose $c \in U$, $F(c) = 0$, and $dF(c)$ has rank q. Then there is a neighborhood V of c, $V \subset U$, such that the level set $S = \{u \in V : F(u) = 0$ is a smooth p-surface, where $p = d - q$. That is, S has a smooth parameterization of dimension p. Hence, S has a tangent space at each point of S. Furthermore, the tangent space at c is the set of solutions u to the equation*

$$dF(c)(u - c) = 0.$$

Proof. Since $dF(c)$ has rank q, there is a $q \times q$ submatrix of the $q \times d$ matrix $dF(c)$ which is non-singular. By rearranging the variables in F, if necessary, we may assume that the last q columns of dF form a non-singular matrix. With $p = d - q$, we may represent $\mathbb{R}^d$ as $\mathbb{R}^p \times \times \mathbb{R}^q$ and label the variables by $(x, y) = (x_1, \cdots, x_p, y_1, \cdots, y_q)$, as in the preceding theorem. Then the hypothese of that theorem are satisfied, with $c = (a, b)$.

By the Implicit Function Theorem, there are neighborhoods V of $c = (a, b)$ and A of a and a smooth function $G : A \to \mathbb{R}^q$ with $(x, G(x)) \in V$ for all $x \in A$ and such that $F(x, y) = 0$ for $(x, y) \in V$ if and only if $y = G(x)$.

Thus, $S = \{u = (x, y) \in V : F(u) = 0\}$ is the graph of the smooth function G. Then the function $H(x) = (x, G(x))$ is a smooth parameterization of S. $\square$

Example 9.7.4. For the system of equations

$$u^2 + v^2 - x = 0$$
$$u + v + y = 0,$$

find the points on the solution set S at which it may not be possible to solve for u and v as smooth functions of x and y in some neighborhood of the point.

Solution According to the Implicit Function Theorem, there will be smooth solutions in a neighborhood of any point where the following matrix is non-singular:

$$\frac{\partial(f_1, f_2)}{\partial(u, v)} = \begin{pmatrix} 2u & 2v \\ 1 & 1 \end{pmatrix},$$

where $f_1(x, y, u, v) = u^2 + v^2 - x$ and $f_2(x, y, u, v) = u + v + y$. This matrix is singular only when $u = v$. This happens at a point on S if and only if $u = v$ and $y^2 = 2x$.

Recall that the kernel of an affine transformation $L : \mathbb{R}^p \to \mathbb{R}$ of rank 1 is a hyperplane in $\mathbb{R}^p$. The Implicit Function Theorem allows us to draw a similar conclusion for functions which are not affine.

Example 9.7.5. For the equation

$$x^2 + y^2 + z^3 = 0,$$

at which points on its solution set S can we be assured that there is a neighborhood of the point in which S is a smoothly parameterized surface? Find an equation of the tangent space at each such point.

Solution: By the corollary to the Implicit Function Theorem, there will be a smooth parameterization of S in a neighborhood of any point at which df has rank 1, where $f(x, y, z) = x^2 + y^2 + z^3$, Since

$$df(x, y, z) = (2x, 2y, 3z^2),$$

the only point at which such a parameterization may not be possible is the origin.

At any point (a, b, c) which is not the origin, an equation for the tangent space is

$$df(a, b, c)(x - a, y - b, z - c) = 0,$$

or

$$2a(x - a) + 2b(y - b) + 3c^2(z - c) = 0.$$

Exercise Set 9.7

1. Are there any points on the graph of the equation $x^3 + 3xy^2 + 2y^3 = 1$ where it may not be possible to solve for y as a smooth function of x in some neighborhood of the point?

2. Can the equation $xz + yz + \sin(x+y+z) = 0$ be solved, in a neighborhood of $(0,0,0)$ for z as a smooth function $z = g(x,y)$ of (x,y), with $g(0,0) = 0$?

3. Find $\dfrac{\partial(f_1, f_2)}{\partial(u, v)}$ if

$$f_1(x, y, u, v) = u^2 + v^2 + x^2 + y^2$$
$$f_2(x, y, u, v) = xu + yv + x - y.$$

   At which points (x, y, u, v) is this matrix non-singular?

4. Show that the system of equations

$$u^2 + v^2 + 2u - xy + z = 0$$
$$u^3 + \sin v - xu + yv + z^2 = 0$$

   has a solution for (u, v) as a smooth function of (x, y, z), in some neighborhood of $(0,0,0)$, with the property that $(u, v) = (0, 0)$ when $(x, y, z) = (0, 0, 0)$.

5. Show that the system of equations

$$u^3 + x^2v^2 - 2y + w = 0$$
$$v^3 + y^2u^2 - 2x + w = 0$$
$$w^2 + wx - y^2 = 0$$

   has a solution for u, v, w as functions of (x, y) in a neighborhood of the point $(x, y, u, v, w) = (1, 1, 1, 1, 0)$ with $u(1, 1) = 1, v(1, 1) = 1, w(1, 1) = 0$.

6. For the equation $xy + yz + xz = 1$, at which points on the solution set S is there a neighborhood in which S is a smooth 2 surface? At each such point (a, b, c), find an equation of the tangent plane.

7. For the system of equations

$$x^2 + y^2 - z^2 = 0$$
$$x + y + z = 0,$$

   at which points of the solution set S is there a neighborhood in which S is a smooth curve? At each such point, find an equation of the tangent line.

8. For the system of equations

$$x^2 + y^2 + u^2 - 3v = 1$$
$$2x + xy - y + 3u^2 - 9v = 0,$$

   find all points on the solution set S for which there is a neighborhood in which S is a smooth 2 surface.

9. If $F(x, y, u, v) = (x\,e^u + y\,e^u, xv + yu) \in \mathbb{R}^2$, find those points (x, y, u, v) at which the level set of F, containing this point, is a smooth 2-surface in a neighborhood of the point.

10. If $F : \mathbb{R}^p \to \mathbb{R}^q$ is a smooth function and dF has rank q at a certain point $a \in \mathbb{R}^p$, prove that there is a neighborhood of a in which dF has rank q.

Chapter 10

Integration in Several Variables

Integration theory for functions of several variables has much in common with integration for functions of a single variable, Many of the proofs are almost identical. However, there are some fundamental differences.

In one variable, we only have to worry about integrating over an interval. However, in several variables the sets we integrate over can be much more complicated. There are issues concerning the boundary of the set and how large it can be. Such issues don't arise in the theory of integration of a function of one variable. In one variable, the change of variable formula for integration (the substitution formula) is quite simple and has a simple proof – it follows directly from the chain rule for differentiation and the Fundamental Theorem of Calculus. The analogous formula in several variables is much more complicated – it involves the determinant of the differential of the change of variables transformation. Its proof is long and complicated.

We begin with a definition of the integral of a function over a multidimensional rectangle.

10.1   Integration over a Rectangle

An *aligned rectangle* in $\mathbb{R}^d$ is a set of the form

$$R = [a_1, b_1] \times \cdots \times [a_d, b_d] = \{(x_1, \cdots, x_d) \in \mathbb{R}^d : a_k \leq x_k \leq b_k, \ k = 1 \cdots d\}.$$

We call such a rectangle *aligned* because each of its edges is parallel to a coordinate axis. Unless otherwise specified, in this chapter the term *rectangle* will mean *aligned rectangle*.

The *d-volume* of a rectangle is the product of the lengths of its edges – that

is, the d-volume $V(R)$ of the rectangle R above is

$$V(R) = \prod_{k=1}^{d}(b_k - a_k).$$

Thus, the 1-volume of a rectangle (an interval) in $\mathbb{R}$ is its length; the 2-volume of a rectangle in $\mathbb{R}^2$ is its area. The 3-volume of a rectangle in $\mathbb{R}^3$ is its ordinary volume.

Note that it is possible for one of the intervals $[a_k, b_k]$ defining a rectangle in $\mathbb{R}^d$ to be degenerate – that is, it could be that $a_k = b_k$. In this case, the rectangle has d-volume 0. This makes sense, because it is actually a rectangle of dimension $d - 1$ in this case.

As long as the dimension of the ambient space $\mathbb{R}^d$ is understood, we will drop the d and just refer to the d-volume of a rectangle as its volume.

An *aligned partition* P of an aligned rectangle $R = [a_1, b_1] \times \cdots \times [a_d, b_d]$ is a partition

$$\{a_k = x_{0,k} \leq x_{1,k} \leq \cdots \leq x_{d,k} = b_k\}$$

of each of the intervals $[a_k, b_k]$. Such a thing divides R up into subrectangles of the form

$$[x_{j_1-1,1}, x_{j_1,1}] \times \cdots \times [x_{j_d-1,d}, x_{j_d,d}]$$
$$= \{(x_1, \cdots, x_d) \in \mathbb{R}^d : x_{j_k-1,k} \leq x_k \leq x_{j_k,k}, \ k = 1 \cdots d\}.$$

Each of these will be called a subrectangle for the partition P of the rectangle R. If n is the number of subrectangles for P, then we will number these subrectangles in some fashion so that we have a list $\{R_1, R_2, \cdots, R_n\}$ of all the subrectangles for P. We will not attempt to arrange this numbering scheme in a way that has anything to do with the indexing of the points in the corresponding partitions of the individual intervals $[a_k, b_k]$. To do so would lead to an awful mess.

Note that R is the union of the subrectangles determined by a partition of R and any two of these subrectangles are either disjoint or have a lower dimensional rectangle as intersection. The volume of R is the sum of the volumes of the subrectangles determined by the partition.

Unless otherwise specified, in this chapter, the term *partition* will mean *aligned partition*.

Upper and Lower Sums

Let f be a bounded real valued function defined on a rectangle R and let P be a partition of R determining a list of subrectangles $R_1, R_2, \cdots, R_n$,

Definition 10.1.1. If f, R, P, and $\{R_1, R_2, \cdots, R_n\}$ are as above, then we
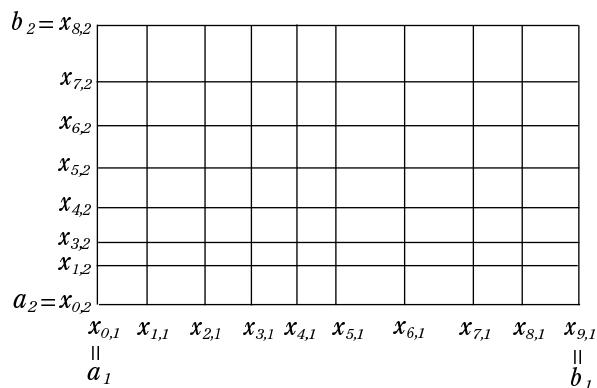
Figure 10.1: Partition of a Rectangle

define the *upper and lower sums* for f and P by

$$U(f,P) = \sum_{j=1}^{n} M_j V(R_j),$$

$$L(f,P) = \sum_{j=1}^{n} m_j V(R_j),$$

(10.1.1)

where

$$M_j = \sup_{R_j} f \quad \text{and} \quad m_j = \inf_{R_j} f.$$

This is exactly the way we defined the upper and lower sums for f and the partition P in Definition 5.1.1, except there we were partitioning intervals into subintervals and here we are partitioning d-dimensional rectangles into subrectangles.

As in Section 5.1, a *Riemann Sum* for f and P on R is a sum of the form

$$\sum_{j=1}^{n} f(u_j)V(R_j)$$

(10.1.2)

where, for each j, u_j is some point in the rectangle R_j. For each j, the term $f(u_j)V(R_j)$ represents the volume (or minus the volume, if $f(\bar{u}_j) < 0$) of a $d+1$-dimensional rectangle with base R_j and with height $|f(u_j)|$. Now, for each j we have

$$m_j \leq f(u_j) \leq M_j,$$

which implies

$$L(f,P) \leq \sum_{j=1}^{n} f(u_j)V(R_j) \leq U(f,P).$$

Thus, as in Section 5.1, every Riemann sum for f and P lies between the lower and upper sums for f and P.

Refinement

If R is a rectangle in $\mathbb{R}^d$, and P and Q partitions of R, then Q is said to be a *refinement* of P if every subrectangle of R determined by Q is a subset of some subrectangle determined by P.

If $R = [a_1, b_1] \times \cdots \times [a_d, b_d]$, then the partition P consists of a partition of each of the intervals $[a_k, b_k]$, as does the partition Q. It is not difficult to see that Q is a refinement of P if and only if, for $k = 1, \cdots, d$, the partition of $[a_k, b_k]$ determined by Q is a refinement of the partition of this same interval determined by P. For this reason, it is also easy to see that any two partitions P, Q of R have a common refinement, since this is true for partitions of intervals.

If Q is a refinement of P, then since R is the union of the subrectangles of itself determined by a given partition, each subrectangle for P is a union of the subrectangles for Q which it contains. This is the key fact needed to prove the following theorem in essentially the same way as the analogous theorem in one variable (Theorem 5.1.4). The details are left to the exercises.

Theorem 10.1.2. *Let f be a bounded function on a rectangle R in $\mathbb{R}^d$. If Q and P are partitions of R and Q is a refinement of P, then*

$$L(f, P) \leq L(f, Q) \leq U(f, Q) \leq U(f, P). \tag{10.1.3}$$

Let P_1 and P_2 be any two partitions of R and let Q be a common refinement of P_1 and P_2, then (10.1.3) holds with P replaced by P_1 and with P replaced by P_2. The resulting inequalities imply the following.

Theorem 10.1.3. *If P_1 and P_2 are partitions of R, then*

$$L(f, P_1) \leq U(f, P_2).$$

Thus, any lower sum for f is less than or equal to any upper sum for f.

Upper and Lower Integrals

Definition 10.1.4. Let R be a rectangle in $\mathbb{R}^d$ and f a bounded real valued function on R. The *upper and lower integrals* of f on R are defined by

$$\overline{\int}_R f(x) dV(x) = \inf\{U(f, P) : \text{P a partition of } \mathbb{R}\}$$
$$\underline{\int}_R f(x) dV(x) = \sup\{L(f, P) : \text{P a partition of } \mathbb{R}\} \tag{10.1.4}$$

The set of all upper sums for f is bounded below by any lower sum and the set of lower sums is bounded above by any upper sum. Thus, the inf (greatest lower bound) of the set of upper sums is greater than or equal to any lower sum and, hence, also greater than or equal to the sup (least upper bound) of the set of all lower sums. Thus,

Theorem 10.1.5. *If f is a bounded real valued function on a rectangle R and if P and Q are arbitrary partitions of R then*

$$L(f,P) \le \underline{\int}_R f(x)dV(x) \le \overline{\int}_R f(x)dV(x) \le U(f,Q)$$

The Integral

A bounded function on R is *integrable* if its upper and lower integrals are the same. That is:

Definition 10.1.6. Let R be a rectangle in $\mathbb{R}^d$ and f a bounded real valued function on R. If $\underline{\int}_R f(x)dV(x) = \overline{\int}_R f(x)dV(x)$, then we will say that f is *integrable* on R. In this case, we will call the common value of these two expressions the *Riemann integral* of f on R and denote it by

$$\int_R f(x)dV(x).$$

The proofs of the following two theorems are exactly the same as the proofs of Theorems 5.1.7 and 5.1.8 and we will not repeat them here.

Theorem 10.1.7. *If f is a bounded function on a rectangle R, then f is Riemann integrable on R if and only if, for each $\epsilon > 0$, there is a partition P of R such that*

$$U(f,P) - L(f,P) < \epsilon. \tag{10.1.5}$$

Theorem 10.1.8. *With f and R as above, f is Riemann integrable on R if and only if there is a sequence $\{P_n\}$ of partitions of R such that*

$$\lim(U(f,P_n) - L(f,P_n)) = 0. \tag{10.1.6}$$

In this case,

$$\int_R f(x)dV(x) = \lim S_n(f)$$

where, for each n, $S_n(f)$ may be chosen to be $U(f,P_n)$, $L(f,P_n)$ or any Riemann sum (10.1.2) for f and the partition P_n.

Remark 10.1.9. The preceding two theorems both involve the difference between the upper and lower Riemann sums for f and P. This can be written as

$$U(f,P) - L(f,P) = \sum_{j=1}^{n}(M_j - m_j)V(R_j). \tag{10.1.7}$$

The factors $M_j - m_j$ that appear in this expression are non-negative numbers, as are the numbers V_j. Hence, any operation that reduces or eliminates some of the terms in this sum will result in a smaller sum.

Properties of the Integral

The next theorem states one of the most important properties of the integral. The proof of this theorem differs in no essential way from the proof of the analogous theorem for functions of one variable (Theorem 5.2.3). In fact, the only difference is that intervals on the line are replaced by aligned rectangles in $\mathbb{R}^d$. We will not repeat the proof here.

Theorem 10.1.10. *If f and g are integrable functions on an aligned rectangle R in $\mathbb{R}^d$ and c is a constant, then*

(a) cf is integrable and $\displaystyle\int_R cf(x)dV(x) = c\int_R f(x)dV(x)$;

(b) $f+g$ is integrable and $\displaystyle\int_R (f+g)(x)dV(x)) = \int_R f(x)dV(x) + \int_R g(x)dV(x)$.

Taken together, the statements of the above theorem mean that the integrable functions on R form a vector space under pointwise addition and scalar multiplication of functions, and the integral is a linear transformation from this vector space to the vector space $\mathbb{R}$.

The order preserving property is another key property of the integral. The version stated in the next theorem is somewhat more general than the analogous result, proved earlier for functions of a single variable (Theorem 5.2.4), and it has a different proof. Hence, we include the proof.

Theorem 10.1.11. *If f and g are functions on an aligned rectangle R in $\mathbb{R}^d$, and $f(x) \le g(x)$ for all $x \in [a,b]$, then*

(a) $\displaystyle\overline{\int}_R f(x)dV(x) \le \overline{\int}_R g(x)dV(x)$ and $\displaystyle\underline{\int}_R f(x)dV(x) \le \underline{\int}_R g(x)dV(x))$;

(b) $\displaystyle\int_R f(x)dV(x) \le \int_R g(x)dV(x)$ if f and g are integrable.

Proof. We will prove this result for the upper integrals. The result for the lower integrals has an analogous proof. The result for the integral in the case of integrable functions then follows because upper integral, lower integral, and integral are all the same for an integrable function.

Given a partition P of R, determining subrectangles $\{R_1, \cdots, R_n\}$ of R, we set

$$M_j(f) = \sup_{R_j} f \quad \text{and} \quad M_j(g) = \sup_{R_j} g.$$

Then $M_j(f) \le M_j(g)$ for all j because $f(x) \le g(x)$ for all $x \in R$. Hence,

$$U(f,P) = \sum_{j=1}^n M_j(f)V(R_j) \le \sum_{j=1}^n M_j(g)V(R_j) = U(g,P).$$

It follows that

$$\overline{\int}_{R} f(x)dV(x) = \inf_{P} U(f,P) \le \inf_{P} U(g,P) = \overline{\int}_{R} g(x)dV(x)$$

This completes the proof. □

A Simple Example

So far we have not computed a single integral or shown that a single function is integrable. We do so now. The function we will integrate is very simple, though not continuous, but the computation of its integral is an important step in our development of integration theory.

Definition 10.1.12. Let E be a subset of $\mathbb{R}^d$. Then the *characteristic function* of E, denoted χ_E is the real valued function on $\mathbb{R}^d$ defined by

$$\chi_E(x) = \begin{cases} 1 & \text{if} \quad x \in E \\ 0 & \text{if} \quad x \notin E. \end{cases}$$

Our example is as follows:

Example 10.1.13. Let R and S be aligned rectangles with $S \subset R$. Show that χ_S is an integrable function on R and

$$\int_{R} \chi_S(x)dV(x) = V(S).$$

Solution: Let

$$R = [a_1, b_1] \times \cdots \times [a_d, b_d] \text{ and}$$
$$S = [s_1, t_1] \times \cdots \times [s_d, t_d],$$

where $a_j \le s_j \le t_j \le b_j$ for each j. Given $\epsilon > 0$, We choose a partition of R as follows: for each j, we partition each interval $[a_j, b_j]$ with the points $\{a_j \le u_j \le s_j \le t_j \le v_j \le b_j\}$, where the points u_j and v_j are chosen so that if A is the rectangle

$$A = [u_1, v_1] \times \cdots \times [u_d, v_d]$$

Then $V(A) < V(S) + \epsilon$ (see Figure 10.2 for a two dimensional version of this setup).

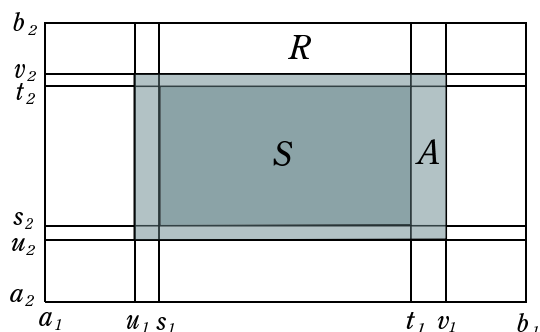The sup of χ_S on a given subrectangle R_j is 1 if $R_j \cap S \ne \emptyset$ and is 0 otherwise. The inf of χ_S on R_j is 1 if $R_j \subset S$ and is 0 otherwise.

There is only one subrectangle for this partition which is contained in S and that is S itself. Thus,

$$L(\chi_S, P) = V(S).$$

The union of the subrectangles R_j that meet S is A. Hence,

$$U(\chi_S, P) = V(A).$$

Figure 10.2: Computing the Integral of χ_S

Since $V(S) < V(A) < V(S) + \epsilon$, we have $V(A) - V(S) < \epsilon$. Hence,

$$U(\chi_S, P) - L(\chi_S, P) < \epsilon.$$

By Theorem 10.1.7, χ_S is integrable on R. Its integral is within ϵ of $L(\chi_S, P) = V(S)$ for every $\epsilon > 0$ and so $\int_R \chi_S(x)dV(x) = V(S)$.

Exercise Set 10.1

1. Let $R = [0,1] \times [0,1]$ be the square with vertices at $(0,0)$, $(1,0)$, $(1,1)$, and $(0,1)$ and let P be the partition of R consisting of the partition $\{0, 1/4, 1/2, 3/4, 1\}$ in both factors of $[0,1] \times [0,1]$. Find $U(f,P)$ and $L(f,P)$ if $f(x,y) = xy$.

2. With R and P as in the previous problem, find $U(\chi_\Delta, P)$ and $L(\chi_\Delta, P)$ if Δ is the closed, solid triangle with vertices at $(0,0)$, $(1,0)$, $(1,1)$.

3. Suppose f and g are functions defined on an aligned rectangle R. Suppose there is a positive constant K such that $|f(x) - f(y)| \leq K|g(x) - g(y)|$ for all $x, y \in R$. Prove that if g is integrable on R, then so is f.

4. Use the result of the preceding exercise to prove that if f is an integrable function on an aligned rectangle R, then $|f|$ is also integrable on R.

5. Prove that if f is integrable on R, then f^2 is also integrable on R.

6. Use the result of the preceding exercise to prove that if f and g are integrable on R, then fg is also integrable on R.

7. Show that each constant function k is integrable and $\int_R k\,dV(x) = kV(R)$.

8. If f is an integrable function defined on the rectangle R and $|f(x)| \leq M$ on R, where M is a positive constant, then prove that

$$\left| \int_R f(x)dV(x) \right| \leq MV(R).$$

9. Prove that if R is an aligned rectangle and f is a continuous function on R, then f is integrable on R.

10. If A and B are subsets of $\mathbb{R}^d$, then

   (a) describe $\chi_{A \cap B}$ in terms of χ_A and χ_B;
   (b) describe $\chi_{A \cup B}$ in terms of χ_A and χ_B;
   (c) describe the meaning of $B \subset A$ in terms of χ_A and χ_B;
   (d) if $B \subset A$, describe $\chi_{A \setminus B}$ in terms of χ_A and χ_B.

10.2 Jordan Regions

The concept of characteristic function of a set (Definition 10.1.12) allows us to define the volume of a set in terms of the integral that we just defined. The volume (or inner or outer volume) of a set E, as defined below, depends very much on the dimension of the ambient space $\mathbb{R}^d$ and so, technically, it should be called the d-volume (or inner or outer d-volume) of the set. However, as with rectangles, we will drop the d when the dimension of the ambient space is understood.

Definition 10.2.1. If E is a bounded subset of $\mathbb{R}^d$, let R be an aligned rectangle containing E. Then we define the outer volume $\overline{V}(E)$, inner volume $\underline{V}(E)$, and volume $V(E)$ (if it exists) for E by

(a) $\overline{V}(E) = \overline{\int}_R \chi_E(x) dV(x)$;

(b) $\underline{V}(E) = \underline{\int}_R \chi_E(x) dV(x)$; and

(c) $V(E) = \int_R \chi_E(x) dV(x)$ if the latter exists – that is if $\underline{\int}_R \chi_E dV(x) = \overline{\int}_R \chi_E(x) dV(x)$.

If $V(E)$ exists, then we call E a *Jordan region*.

Note that E is a Jordan region if and only if $\underline{V}(E) = \overline{V}(E)$ and, in this case, $V(E)$ is their common value.

Note also that, if E is an aligned rectangle, then E is a Jordan region and the above definition of $V(E)$ agrees with our earlier definition. This is demonstrated in Example 10.1.13.

Implicit in the above definition is the fact that the upper and lower integrals of χ_E over R do not depend on the rectangle R, as long as R contains E. We leave a proof of this to the exercises (Exercise 10.2.1).

Example 10.2.2. Show that the closed, solid right triangle Δ in $\mathbb{R}^2$ with vertices at $(0,0)$, $(a,0)$, and $(0,b)$ is a Jordan region and has area (2-volume) $ab/2$.

Solution: We choose R to be the rectangle $[0,a] \times [0,b]$. This contains the triangle Δ. For each n, we choose a partition P_n of R consisting of partitions
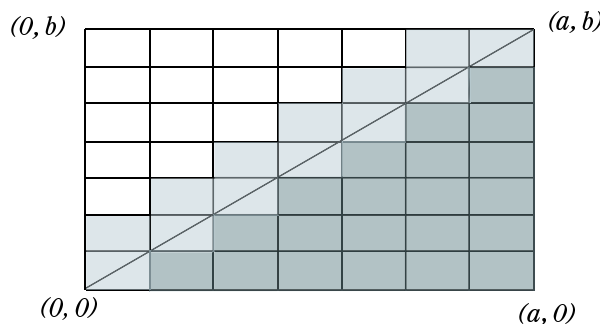
Figure 10.3: Computing the Area of a Triangle

$\{0, a/n, 2a/n, \cdots, na/n = a\}$ of $[0, a]$ and $\{0, b/n, 2b/n, \cdots, nb/n = b\}$ of $[0, b]$. This determines n^2 subrectangles of R, each of volume ab/n^2.

Now for each of these subrectangles R_j, the sup, M_j, and inf, m_j, of χ_Δ on R_j is either 1 or 0. In fact,

$$M_j = 1 \quad \text{if and only if} \quad R_j \cap \Delta \neq \emptyset$$
$$m_j = 1 \quad \text{if and only if} \quad R_j \subset \Delta.$$

Thus, the only subrectangles R_j on which $M_j \neq m_j$ are those which are not contained in Δ but have non-empty intersection with it (the light grey subrectangles in Figure 10.3). There are two kinds of these, those of the form $[(k-1)a/n, ka/n] \times [(k-1)b/n, kb/n]$ which are bisected by the line from $(0, 0)$ to (a, b) and those of the form $[(k-1)a/n, ka/n] \times [kb/n, (k+1)b/n]$ which just have a lower right vertex on this line. There are n of the former and $n-1$ of the latter. The difference $U(\chi_\Delta, P_n) - L(\chi_\Delta, P_n)$ is just the sum of the areas of these $2n-1$ rectangles, which is $(2n-1)ab/n^2$. Hence,

$$\lim_{n \to \infty} (U(\chi_\Delta, P_n) - L(\chi_\Delta, P_n)) = \lim_{n \to \infty} \frac{(2n-1)ab}{n^2} = 0.$$

By Theorem 10.1.8, the Riemann integral $\int_R \chi_\Delta(x) dV(x)$ exists and so the 2-volume (area) of the set Δ exists – that is, Δ is a Jordan region.

Also by Theorem 10.1.8 the integral $\int_R \chi_\Delta(x) dV(x)$ is the limit of the sequence $\{L(\chi_\Delta, P_n)\}$. However, $L(\chi_\Delta, P_n)$ is the sum of the areas of the subrectanges that are contained in Δ (the dark grey subrectangles in Figure 10.3). There are $n(n-1)/2$ of these (half the number remaining after the ones that are bisected by the line from $(0, 0)$ to (a, b) are removed). Hence,

$$V(\Delta) = \int_R \chi_\Delta(x) dV(x) = \lim_{n \to \infty} \frac{n(n-1)ab}{2n^2} = \frac{ab}{2}.$$

Properties of Volume

Many properties of the integral translate directly into properties of volume. For example, Theorem 10.1.11 implies that

Theorem 10.2.3. *If E and F are bounded subsets of $\mathbb{R}^d$ and $E \subset F$, then*

$$\underline{V}(E) \leq \underline{V}(F) \quad and \quad \overline{V}(E) \leq \overline{V}(F).$$

If E and F are Jordan regions, then $V(E) \leq V(F)$.

Theorem10.1.10 and the fact that $\chi_{E \cup F} = \chi_E + \chi_R - \chi_{E \cap F}$ (Exercise 10.1.10) imply

Theorem 10.2.4. *If E, F and $E \cap F$ are Jordan regions and $V(E \cap F) = 0$, then $E \cup F$ is a Jordan region and*

$$V(E \cup F) = V(E) + V(F).$$

In particular, this identity holds if E and F are disjoint Jordan regions.

In particular, if R is an aligned rectangle in R^d and $R_j \neq R_k$ are two of the subrectangles determined by a partition P, then $R_j \cap R_k$ is either empty or is a degenerate aligned rectangle in R – that is, its dimension is lower than that of R. Hence, $V(R_j \cap R_k) = 0$. Thus, by Theorem 10.1.10,

$$V(R_j \cup R_k) = V(R_j) + V(R_k).$$

An induction argument then shows that if F is the union of any number of the subrectangles determined by P, then F is a Jordan region and $V(F)$ is the sum of the volumes of these subrectangles. This is used in the proof of the following theorem.

Theorem 10.2.5. *If E is a bounded subset of $\mathbb{R}^d$, then $\overline{V}(E) = \overline{V}(\overline{E})$ and $\underline{V}(E) = \underline{V}(E^\circ)$.*

Proof. Let R be an aligned rectangle containing E, let P be a partition of R, and let $\{R_j\}$ be the list of subrectangles of R determined by P. Then $U(\chi_E, P)$ is the sum of the volumes of the rectangles R_j in this list that have a non-empty intersection with E (those for which χ_E takes on the value 1 somewhere on R_j). If we set

$$F = \bigcup \{R_j : E \cap R_j \neq \emptyset\},$$

then $U(\chi_E, P) = V(F)$, by the paragraph preceding this theorem.

Now F is a finite union of closed sets and so it is also closed. Since $E \subset F$, we also have $\overline{E} \subset F$. Then

$$\overline{V}(E) \leq \overline{V}(\overline{E}) \leq \overline{V}(F) = V(F) = U(\chi_E, P).$$

Since $\overline{V}(E) = \inf\{U(\chi_E, P) : P$ a partition of $R\}$, we have

$$\overline{V}(E) \leq \overline{V}(\overline{E}) \leq \overline{V}(E).$$

Thus, $\overline{V}(E) = \overline{V}(\overline{E})$.

Similarly, if we set

$$G = \bigcup \{R_j : R_j \subset E\},$$

then, since $G^\circ \subset E^\circ$,

$$\underline{V}(G^\circ) \leq \underline{V}(E^\circ) \leq \underline{V}(E).$$

However, $V(G^\circ) = V(G) = L(\chi_E, P)$, since the boundary of G consists of a finite union of rectangles of dimension lower than d, and these all have volume 0. Since $\sup_P L(\chi_E, P) = \underline{V}(E)$, we conclude that $\underline{V}(E^\circ) = \underline{V}(E)$. This completes the proof.

$\square$

Theorem 10.2.6. *If E is a Jordan region, then so are $\overline{E}$ and E°. Furthermore, $V(E) = V(\overline{E}) = V(E^\circ)$.*

Proof. In view of the previous theorem,

$$\underline{V}(E) \leq \underline{V}(\overline{E}) \leq \overline{V}(\overline{E}) \leq \overline{V}(E).$$

If E is a Jordan region, then $\underline{V}(E) = \overline{V}(E)$ and, hence, each of the above inequalities is an equality. This implies $\overline{E}$ is a Jordan region and $V(\overline{E}) = V(E)$. The proof of the statement for E° is similar.

$\square$

Sets of Volume Zero

We leave the proof of the following theorem to the exercises.

Theorem 10.2.7. *If E is a bounded set with $\overline{V}(E) = 0$, then E is a Jordan region with volume 0. Any subset of a Jordan region of volume 0 is also a Jordan region of volume 0. A finite union of Jordan regions of volume 0 is also a Jordan region of volume 0.*

We will, henceforth, refer to a set E with $\overline{V}(E) = 0$ as simply a *set of volume 0*.

Theorem 10.2.8. *A set E is a set of volume 0 if and only if, for each $\epsilon > 0$, there is a finite set $\{R_1, \cdots, R_n\}$ of aligned rectangles such that*

$$E \subset \bigcup_{j=1}^n R_j \quad and \quad \sum_{j=1}^n V(R_j) < \epsilon.$$

Proof. If $\overline{V}(E) = 0$, then there exist an aligned rectangle R with E in its interior and a partition P of R such that $U(\chi_E, P) < \epsilon$. This just means that those subrectangles determined by P which meet E have volumes which add up to a number less than ϵ. Since E is contained in the union of these rectangles, the proof of the "only if" part of the theorem is complete.

On the other hand, if $E \subset F = \cup_{j=1}^{n} R_j$ for a set of aligned rectangles with volumes adding up to a number less than ϵ, then $\overline{V}(F) < \epsilon$ since

$$\chi_F \leq \sum_{j=1}^{n} \chi_{R_j},$$

which, together with the fact that each χ_{R_j} is integrable, implies

$$\overline{V}(F) = \overline{\int}_R \chi_F(x) dV(x) \leq \overline{\int}_R \sum_{j=1}^{n} \chi_{R_j}(x) dV(x)$$

$$= \sum_{j=1}^{n} \int_R \chi_{R_j}(x) dV(x) = \sum_{j=1}^{n} V(R_j) < \epsilon.$$

This proves the "if" part of the theorem.   □

A Characterization of Jordan Regions

Theorem 10.2.9. *A bounded set E is a Jordan region if and only if its boundary, ∂E, is a set of volume 0.*

Proof. If P is a partition of R determining a list of subrectangles $\{R_j\}$, then $L(\chi_{E^\circ}, P)$ is the sum of the areas of those R_j which are entirely contained in E°, while $U(\chi_{\overline{E}}, P)$ is the sum of the areas of those R_j which have non-empty intersection with $\overline{E}$. It follows that

$$U(\chi_{\overline{E}}, P) - L(\chi_{E^\circ}, P) = U(\chi_{\partial E}, P).$$

Hence, a sequence $\{P_n\}$ of partitions has the property that $\lim U(\chi_{\partial E}, P_n) = 0$ if and only if it has the property that

$$\lim(U(\chi_{\overline{E}}, P_n) - L(\chi_{E^\circ}, P_n)) = 0.$$

Since, for an appropriately chosen sequence of partitions, this limit is

$$\overline{V}(\overline{E}) - \underline{V}(E^\circ) = \overline{V}(E) - \underline{V}(E),$$

by Theorem 10.2.5, we conclude that $\underline{V}(E) = \overline{V}(E)$ if and only if $\overline{V}(\partial E) = 0$ – that is, E is a Jordan region if and only if ∂E is a set of volume 0.   □

Theorem 10.2.10. *If A and B are Jordan regions, then $A \cap B$, $A \cup B$, and $A \setminus (A \cap B)$ are also Jordan regions. Furthermore,*

$$V(A \cup B) = V(A) + V(B) - V(A \cap B), \text{ and}$$
$$V(A \setminus (A \cap B)) = V(A) - V(A \cap B). \tag{10.2.1}$$

Proof. Each of the sets $A \cap B$, $A \cup B$, and $A \setminus (A \cap B)$ has its boundary contained in $\partial A \cup \partial B$. Since A and B are Jordan regions, ∂A and ∂B are sets of volume 0. Then Theorem 10.2.7 implies that $\partial A \cup \partial B$ has volume 0, as does each of its subsets. It follows from the previous theorem that $A \cap B$, $A \cup B$, and $A \setminus (A \cap B)$ are Jordan regions.

The second statement of the theorem follows from the identities

$$\chi_{A \cup B} = \chi_A + \chi_B - \chi_{A \cap B}, \text{ and}$$
$$\chi_{A \setminus (A \cap B)} = \chi_A - \chi_{A \cap B}. \tag{10.2.2}$$

$\square$

Example 10.2.11. Let K be a compact subset of $\mathbb{R}^{d-1}$ and let $f : K \to \mathbb{R}$ be a continuous function. Show that the graph $G(f)$ of f is a set of d-volume 0, where $G(f) = \{(x, f(x)) : x \in K\}$.

Solution: Since K is compact, it is bounded, and so we may choose a rectangle R in $\mathbb{R}^{d-1}$ which contains K. Let W be the $(d-1)$-volume of R.

Since K is compact and f is continuous, f is actually uniformly continuous. Thus, given $\epsilon > 0$ we may choose a $\delta > 0$ such that

$$|f(x) - f(y)| < \epsilon/W \quad \text{whenever} \quad ||x - y|| < \delta.$$

We let P be a partition of R such that the diameter of each subrectangle for the partition is less than δ (diameter in this case means maximal distance between two points in the subrectangle). Let $R_1, R_2, \cdots, R_n$ be a list of those subrectangles for this partition which meet K. If

$$m_j = \min\{f(x) : x \in K \cap R_j\} \quad \text{and} \quad M_j = \max\{f(x) : x \in K \cap R_j\},$$

then

$$G(f) \subset \bigcup_j (R_j \times [m_j, M_j]).$$

The sum of the volumes of the rectangles $R_j \times [m_j, M_j]$ is

$$\sum_j V(R_j)(M_j - m_j) \leq \frac{\epsilon}{W} \sum V(R_j) \leq \frac{\epsilon}{W} W = \epsilon.$$

By Theorem 10.2.8 the graph $G(f)$ of f is a set of volume 0.

Exercise Set 10.2

1. Prove that $\overline{\int}_R \chi_E(x) dV(x)$ and $\underline{\int}_R \chi_E(x) dV(x)$ do not depend on the choice of the aligned rectangle R as long as it contains E.

2. Prove Theorem 10.2.7 – that is, show that if a subset A of $\mathbb{R}^d$ has outer volume zero, then it and each of its subsets is a Jordan region of volume 0.

3. Show that a finite set in R^d has volume 0.

4. If E is the subset of the unit square $[0,1] \times [0,1]$ consisting of points with both coordinates rational numbers, find its inner volume $\underline{V}(E)$ and outer volume $\overline{V}(E)$. Is E a Jordan region?

5. Show that if A and B are sets of volume 0 in $\mathbb{R}^d$, then $A \cup B$ is also a set of volume 0.

6. Let U be an open subset of $\mathbb{R}^2$ and $K \subset U$ a compact set. Suppose $f : U \to R$ is a smooth function and $E = \{(x,y) \in K : f(x,y) = 0\}$. If df is never 0 on E, then show that E is a set of area 0 in $\mathbb{R}^2$.

7. Show that an ellipse in $\mathbb{R}^2$ is a set of area 0 in $\mathbb{R}^2$ and the solid ellipse that it bounds is a Jordan region.

8. Show that a bounded subset of $\mathbb{R}^2$ whose boundary is a finite union of smooth parameterized curves, is a Jordan region.

9. Consider the following three reflection transformations of $\mathbb{R}^2$:

$$T_1(x,y) = (-x,y), \quad T_2(x,y) = (x,-y) \quad \text{and} \quad T_3(x,y) = (y,x).$$

These are reflection through the y-axis, reflection through the x-axis, and reflection through the line $y = x$, respectively. Prove that if E is a Jordan region, then, for $j = 1,2,3$, so is $T_j(E)$ and $V(T_j(E)) = V(E)$. Hint: what do these reflections do to aligned rectangles and their volumes?

10. Using the previous two exercises and theorems from this section, but without using Example 10.2.2, give a proof that the area of a triangle with one side parallel to a coordinate axis is one half its base times its height. Hint: prove this first for right triangles with legs parallel to the axes.

11. Using the result of the preceding exercise, show that a parallelogram in $\mathbb{R}^2$ with one side parallel to a coordinate axis has area equal to its base times its height.

12. Suppose $B \subset \mathbb{R}^d$ is a compact Jordan region and f and g continuous real valued functions on B with $g(x) \leq f(x)$. Show that the set

$$A = \{(x,t) \in \mathbb{R}^{d+1} : x \in B, \text{ and } g(x) \leq t \leq f(x)\}$$

is also a Jordan region.

10.3   The Integral over a Jordan Region

In this section we extend the definition of the integral to cover integration over a Jordan region. We also prove an existence theorem which shows that the class of integrable functions is quite large.

An Existence Theorem

So far we have only proved the existence of the integral for a few functions of the form χ_E. Our next objective is to prove a general existence theorem for the integral over an aligned rectangle. We will then extend this theorem to integrals over Jordan regions.

Theorem 10.3.1. *Let f be a bounded function on an aligned rectangle R. If the set of points of R at which f is not continuous is a set of volume 0, then f is integrable on R.*

Proof. Let E be the set of points of R at which f is not continuous. Since E is a set of volume 0, its outer volume $\overline{V}(E)$ is 0. Hence, given $\epsilon > 0$, there is a partition P of R such that $U(\chi_E, P) < \epsilon/(4M)$, where M is the sup of $|f|$ on R. If A is the union of the subrectangles for P which meet E, then this means that

$$V(A) = U(\chi_E, P) < \frac{\epsilon}{4M}.$$

Let B be the union of the subrectangles for P which do not meet E. Note that $A \cup B = R$ and B is a closed, bounded (hence compact) set on which f is continuous. Hence, f is uniformly continuous on B by Theorem 8.2.12. This implies that we may choose a $\delta > 0$ such that

$$|f(x) - f(y)| < \frac{\epsilon}{2V(R)} \quad \text{whenever} \quad ||x - y|| < \delta.$$

We next choose a refinement Q for the partition P in such a way that the diameter of each subrectangle for Q is at most δ. If $R_1, R_2, \cdots, R_n$ is a list of the subrectangles for Q, then each R_j is either in A or in B. We let S be the set of integers j in $[1, n]$ such that $R_j \subset A$ and T the set of integers j in this interval such that $R_j \subset B$. If M_j and m_j are the sup and inf of f on R_j, then

$$U(f, Q) - L(f, Q) = \sum_{j=1}^{n} (M_j - m_j) V(R_j)$$
$$= \sum_{j \in S} (M_j - m_j) V(R_j) + \sum_{j \in T} (M_j - m_j) V(R_j)$$
$$\leq 2MV(A) + \frac{\epsilon}{2V(R)} V(B) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.$$

In view of Theorem 10.1.7, the proof is complete. $\qquad\qquad\square$

The Integral over a Jordan Region

Definition 10.3.2. Let A be a Jordan region and f a bounded function defined on a set containing A. We define a new function f_A, with domain all of $\mathbb{R}^d$, as follows:

$$f_A(x) = \begin{cases} f(x) & \text{if } x \in A \\ 0 & \text{if } x \in \mathbb{R}^d \setminus A. \end{cases}$$

Thus, f_A is a function defined on all of $\mathbb{R}^d$. It agrees with f on A and is 0 on the complement of A. Note that f may be originally defined on a larger set than A or it may be defined just on A. In the definition of f_A, it doesn't matter.

Example 10.3.3. Let $A = D_1(0,0)$ in $\mathbb{R}^2$. Find f_A and g_A if f is defined on $\mathbb{R}^2$ by $f(x,y) = x^2 + y^2$ and g is defined on A by $g(x,y) = \sqrt{1 - x^2 - y^2}$.
Solution: From the above definition, we have

$$f_A(x,y) = \begin{cases} x^2 + y^2 & \text{if} \quad (x,y) \in D_1(0) \\ 0 & \text{if} \quad (x,y) \notin D_1(0). \end{cases}$$

and

$$g_A(x,y) = \begin{cases} \sqrt{1 - x^2 - y^2} & \text{if} \quad (x,y) \in D_1(0) \\ 0 & \text{if} \quad (x,y) \notin D_1(0). \end{cases}$$

Note that here f is defined originally on all of R^2 while g is defined only on A.

Definition 10.3.4. With A, f and f_A as in the preceding definition, let R be an aligned rectangle containing A. If f_A is integrable on R we say f is integrable on A and we write

$$\int_A f(x)dV(x) = \int_R f_A(x)dV(x).$$

Implicit in the above definition is the assumption that $\int_R f_A(x)dV(x)$ does not depend on which rectangle R is chosen, as long as it contains A. We leave the proof of this to the exercises.

If A happens to be an aligned rectangle, then one choice for R in the above definition is $R = A$. Then $f = f_A$ on the rectangle R and

$$\int_A f(x)dV(x) = \int_R f_A(x)dV(x) = \int_R f(x)dV(x),$$

where, on the right, the integral over R is the one defined in Section 10.1, while the one on the left is our new definition of the integral over a Jordan region. Fortunately, the two agree.

Existence of the Integral over a Jordan Region

Theorem 10.3.5. *Let A be a Jordan region and f a bounded function defined on A. If the set E of points of A at which f is not continuous is a set of volume 0, then f is integrable on A.*

Proof. Since both E and ∂A are sets of volume 0, their union $F = E \cup \partial A$ is also. We choose an aligned rectangle R such that $\overline{A} \subset R$. Then f_A is continuous on $R \setminus F$. It follows from Theorem 10.3.1 that f_A is integrable on R and, by definition, f is integrable on A. □

Properties of the Integral

For integrals over rectangles, the following theorem is Exercise 10.1.6. The extension of this result to integrals over Jordan regions is left to the exercises.

Theorem 10.3.6. *If A is a Jordan region and f and g are integrable functions on A, then fg is also integrable on A.*

Example 10.3.7. Prove that if $B \subset A$ and A and B are Jordan regions, then each function f which is integrable on A is also integrable on B.

    Solution: This follows immediately from the preceding theorem and the observation that $f_B = \chi_B f_A$.

    The next three theorems follow from Theorems 10.1.10, 10.1.11, and 10.3.6 and some observations about the passage from f to f_A. We leave the details to the exercises.

Theorem 10.3.8. *Let A be a Jordan region, f and g integrable functions on A and c a scalar constant. Then $f + g$ and cf are integrable on A, and*

(a) $\int_A 1 \, dV(x) = V(A)$;

(b) $\displaystyle\int_A (f(x) + g(x))dV(x) = \int_A f(x)dV(x) + \int_A g(x)dV(x)$;

(c) $\displaystyle\int_A cf(x)dV(x) = c\int_A f(x)dV(x)$.

Parts (b) and (c) mean that the integral over A is a linear transformation.

Theorem 10.3.9. *Let A and B be Jordan regions with $V(A \cap B) = 0$ and let f be a bounded function on $A \cup B$. Then f is integrable on A and on B if and only if it is integrable on $A \cup B$. In this case,*

$$\int_{A \cup B} f(x)dV(x) = \int_A f(x)dV(x) + \int_B f(x)dV(x).$$

Theorem 10.3.10. *If A is a Jordan region and f and g are integrable functions on A with $f(x) \leq g(x)$ for all $x \in A$, then*

$$\int_A f(x)dV(x) \leq \int_A g(x)dV(x).$$

Integral of a Sequence

Theorem 10.3.11. *Let A be a Jordan region and $\{f_n\}$ a sequence of integrable functions on A. If $\{f_n\}$ converges uniformly on A to a function f, then f is integrable and*

$$\lim_{n \to \infty} \int_A f_n(x)dV(x) = \int_A f(x)dV(x).$$

Proof. We prove this first in the case where A is an aligned rectangle R.

Given $\epsilon > 0$, there is an N such that $|f(x) - f_n(x)| < \epsilon/V(A)$ whenever $x \in R$ and $n \geq N$. this means that, for $n \geq N$,

$$f_n(x) - \frac{\epsilon}{V(R)} < f(x) < f_n(x) + \frac{\epsilon}{V(R)},$$

for all $x \in R$. By Theorem 10.1.11 this implies that

$$\underline{\int}_R (f_n(x) - \epsilon/V(R))dV(x) \leq \underline{\int}_R f(x)dV(x)$$
$$\leq \overline{\int}_R f(x)dV(x) \leq \overline{\int}_R (f_n(x) + \epsilon/V(R))dV(x).$$

Since f_n and the constant $\epsilon/(2V(R))$ are integrable, their upper and lower integrals are the same and are equal to their integrals. Thus,

$$\int_R f_n(x)dV(x) - \epsilon \leq \underline{\int}_R f(x)dV(x) \leq \overline{\int}_R f(x)dV(x) \leq \int_R f_n(x)dV(x) + \epsilon.$$

Since ϵ is an arbitrary positive number, we conclude that

$$\underline{\int}_R f(x)dV(x) = \overline{\int}_R f(x)dV(x)$$

and, hence, that f is integrable on R. These inequalities also show that

$$\left| \int_R f_n(x)dV(x) - \int_R f(x)dV(x) \right| < \epsilon \quad \text{whenever} \quad n \geq N.$$

Thus, $\lim \int_R f_n(x)dV(x) = \int_R f(x)dV(x)$.

Now if A is not an aligned rectangle, we simply choose an aligned rectangle R which contains A and replace f and f_n by f_A and $(f_n)_A$ in the above argument. We note that $\{(f_n)_A\}$ converges uniformly to f_A on R if $\{f_n\}$ converges uniformly to f on A. The conclusion is that f_A is integrable on R and

$$\lim \int_R (f_n)_A(x)dV(x) = \int_R f_A(x)dV(x).$$

This implies that f is integrable on A and

$$\lim \int_A f_n(x)dV(x) = \int_A f(x)dV(x).$$

$\square$

Example 10.3.12. Show that if f is a bounded function on a Jordan region A and if $\{x \in A : f(x) < r\}$ is a Jordan region for each $r \in \mathbb{R}$, then f is integrable on A.

Solution: Since f is bounded, there is an $M > 0$ such that $-M < f(x) < M$ for all $x \in A$. We set

$$g(x) = \frac{f(x) + M}{2M} \quad \text{so that} \quad f(x) = 2Mg(x) - M.$$

The function g also satifies the hypothesis of the theorem, and $0 < g(x) < 1$ for all $x \in A$. We will show that g is integrable. This clearly implies that f is integrable.

We will show that g is integrable by expressing it as a uniform limit of a sequence of integrable functions. This sequence is constructed as follows. For each positive integer n and each positive integer $k \leq n$, we set

$$E(n,k) = \{x \in A : (k-1)/n \leq f(x) < k/n\}$$
$$= \{x \in A : f(x) < k/n\} \setminus \{x \in A : f(x) < (k-1)/n\}.$$

By hypothesis, $E(n,k)$ is a Jordan region and so $\chi_{E(n,k)}$ is integrable. Also, for each n, $A = \cup_{k=1}^{n} E(n,k)$. We define an integrable function g_n on A by

$$g_n(x) = \sum_{k=1}^{n} \frac{k-1}{n} \chi_{E(n,k)}.$$

That is,

$$g_n(x) = \frac{k-1}{n} \quad \text{if} \quad x \in E(n,k).$$

Since g_n is a linear combination of integrable functions, it is integrable. Also

$$0 \geq g(x) - g_n(x) < k/n - (k-1)/n = 1/n \quad \text{if} \quad x \in E(n,k).$$

Since every $x \in A$ is in $E(n,k)$ for some k, we conclude that

$$|g(x) - g_n(x)| < 1/n \quad \text{for all} \quad x \in A.$$

This implies that $\{g_n\}$ converges uniformly to g on A. By the previous theorem, g is integrable on A. Hence, f is integrable on A.

Exercise Set 10.3

1. Prove that the integral $\int_R f_A(x) dV(x)$ that appears in Definition 10.3.4 does not depend on the choice of R as long as R contains A.

2. Prove Theorem 10.3.6. You may use the result of Exercise 10.1.6.

3. Prove Theorem 10.3.8.

4. Prove Theorem 10.3.9.

5. Prove Theorem 10.3.10.

6. Prove that if A and B are Jordan regions with $B \subset A$ and f is a non-negative integrable function on A, then $\int_B f(x)dV(x) \leq \int_A f(x)dV(x)$.

7. Prove that if f is an integrable function on a Jordan region A, then $|f|$ is integrable and

$$\left| \int_A f(x)dV(x) \right| \leq \int_A |f(x)|dV(x).$$

8. Let A be a Jordan region and f an integrable function on A. For each $x \in A$ define $f^+(x)$ and $f^-(x)$ by

$$f^+(x) = \max\{f(x), 0\} \quad \text{and} \quad f^-(x) = \max\{-f(x), 0\} = (-f(x))^+.$$

Prove that f^+ and f^- are non-negative functions on A with $f = f^+ - f^-$ and $|f| = f^+ + f^-$. Then prove that f^+ and f^- are integrable.

9. Prove that if f is a bounded function on a set A of volume 0, then f is integrable on A and $\int_A f(x)dV(x) = 0$.

10. Let U be an open Jordan region and $\{K_n\}$ an increasing sequence of compact Jordan subsets of U such that $U = \cup_n K_n^\circ$. Prove that, for each integrable function f on U,

$$\int_U f(x)\, dV(x) = \lim_n \int_{K_n} f(x)\, dx.$$

11. Prove that if U is an open Jordan region, then there always exists a sequence $\{K_n\}$ like the one in the previous exercise.

12. Let A be a Jordan region and f an integrable function on A. The average value of f on A is defined to be the number

$$\operatorname{avg}(f, A) = \frac{1}{V(A)} \int_A f(x)dV(x).$$

If A is compact and connected and f is continuous on A, prove that there is a point $x_0 \in A$ at which $f(x_0) = \operatorname{avg}(f, A)$.

13. Suppose A is a Jordan region in $\mathbb{R}^d$ and g_k is an integrable function on A for $k = 1, 2, \cdots$. Prove that if

$$g(x) = \sum_{k=1}^{\infty} g_k(x),$$

where this series converges uniformly on A, then g is integrable and

$$\int_A g(x)dV(x) = \sum_{k=1}^{\infty} \int_A g_k(x)dV(x).$$

14. Prove that the function g on $\mathbb{R}^2$, defined by

$$g(x,y) = \sum_{k=1}^{\infty} \frac{1}{k^2} \sin(kx) \sin(ky),$$

is integrable on any Jordan region in $\mathbb{R}^2$.

10.4   Iterated Integrals

Integrals of functions of a single variable may be calculated exactly in a wide range of situations. The theorem that makes this possible is the Fundamental Theorem of Calculus. We calculate an integral by finding (if we can) an antiderivative for the integrand, then evaluating at the endpoints and subtracting. Fortunately, there is a theorem which often makes it possible to use this same procedure to compute integrals in several variables. This theorem is Fubini's Theorem, and it tells us that, in many situations, we may calculate an integral in several variables by integrating with respect to one variable at a time.

An Additivity Lemma

We begin our discussion of Fubini's Theorem with a lemma that will play an important role in the proof.

Theorem 10.3.9 says that if A and B are Jordan regions with $V(A \cap B) = 0$, then the integral of an integrable function over $A \cup B$ is the sum of the integrals of the function over A and over B. If f is not integrable, only bounded, the analogous result holds for the upper integral of f and for the lower integral of f. We will only need the following special case of this result.

Lemma 10.4.1. *Suppose $R = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_d, b_d]$ is an aligned rectangle in $\mathbb{R}^d$ and f is a bounded function on R. Suppose that $R = R_1 \cup R_2$, where R_1 and R_2 are obtained from R by partitioning one of the intervals $[a_j, b_j]$ into two adjacent subintervals $[a_j, c]$, $[c, b_j]$ and leaving the others alone. Then*

$$\underline{\int}_R f(x)dV(x) = \underline{\int}_{R_1} f(x)dV(x) + \underline{\int}_{R_2} f(x)dV(x),$$

and

$$\overline{\int}_R f(x)dV(x) = \overline{\int}_{R_1} f(x)dV(x) + \overline{\int}_{R_2} f(x)dV(x).$$

Proof. The proof of this is exactly the same as the proof of the interval additivity theorem for the single variable integral (Theorem 5.2.7). The key to the proof is that a partition P_1 of R_1 and a partition P_2 of R_2, together form a partition P of R, and this partition has the property that

$$L(f, P) = L(f, P_1) + L(f, P_2) \quad \text{and} \quad U(f, P) = U(f, P_1) + U(f, P_2).$$

Furthermore, each partition of R has a refinement which is of this form. ☐

Fubini's Theorem

Let S be an aligned rectangle in $\mathbb{R}^p$ and T an aligned rectangle in R^q. Let f be a bounded function on the aligned rectangle $R = S \times T$ in $\mathbb{R}^{p+q}$. We will denote the typical point of $S \times T$ by (x, y) where $x \in S$ and $y \in T$.

If we hold $x \in S$ fixed and consider $f(x, y)$ as a function of $y \in T$, then this function may or may not be integrable on T. In general, it will be integrable for some values of x and not for others. However, the upper and lower integrals of this function of y exist for all x and yield new functions of x on S which also have upper and lower integrals. The key step in the proof of Fubini's Theorem is the following theorem which relates these to the upper and lower integrals of f over $S \times T$.

Theorem 10.4.2. *With S, T, and f as above,*

$$\underline{\int}_{S \times T} f(x,y) dV(x,y) \leq \underline{\int}_S \underline{\int}_T f(x,y) dV(y) dV(x)$$

$$\leq \overline{\int}_S \overline{\int}_T f(x,y) dV(y) dV(x) \leq \overline{\int}_{S \times T} f(x,y) dV(x,y). \tag{10.4.1}$$

Proof. The typical partition of $S \times T$ has the form $P \times Q$, where P is a partition of S and Q is a partition of T. Recall that a partition of S consists of a partition of each of the intervals whose cartesian product is S, while a partition of T consists of a partition of each of the intervals whose cartesian product is T. Taken together, these partitions yield partitions of each of the intervals whose product is $S \times T$. It is this partition of $S \times T$ that we denote by $P \times Q$.

Let $\{S_i\}_{i=1}^n$ be a list of the subrectangles of S determined by the partition P and $\{T_j\}_{j=1}^m$ be a list of the subrectangles of T determined by the partition Q. Then $\{S_i \times T_j\}_{i,j=1}^{n,m}$ is a list of the subrectangles for the partition $P \times Q$. Let

$$M_{ij} = \sup_{S_i \times T_j} f \quad \text{and} \quad m_{ij} = \inf_{S_i \times T_j} f.$$

Then, for $x \in S_i$, Theorem 10.1.11 implies

$$m_{ij} V(T_j) \leq \underline{\int}_{T_j} f(x,y) dV(y) \leq \overline{\int}_{T_j} f(x,y) dV(y) \leq M_{ij} V(T_j).$$

Applying Theorem 10.1.11 again, in the variable x, implies

$$m_{ij} V(S_i) V(T_j) \leq \underline{\int}_{S_i} \underline{\int}_{T_j} f(x,y) dV(y) dV(x)$$

$$\leq \overline{\int}_{S_i} \overline{\int}_{T_j} f(x,y) dV(y) dV(x) \leq M_{ij} V(S_i) V(T_j).$$

If we sum this inequality over i and j, note that $V(S_i) V(T_j) = V(S_i \times T_j)$, and

make repeated use of the preceding lemma, the result is

$$L(f, P \times Q) \leq \underline{\int}_S \underline{\int}_T f(x,y) dV(y) dV(x)$$

$$\leq \overline{\int}_S \overline{\int}_T f(x,y) dV(y) dV(x) \leq U(f, P \times Q)$$

Since the two expressions in the middle of this inequality give an upper bound for $\{L(f, P \times Q)\}$ and a lower bound for $\{U(f, P \times Q)\}$, and since the least upper bound for $\{L(f, P \times Q)\}$ is $\underline{\int}_{S \times T} f(x,y) dV(x,y)$ and the greatest lower bound for $\{U(f, P \times Q)\}$ is $\overline{\int}_{S \times T} f(x,y) dV(x,y)$, we conclude that (10.4.1) holds. $\quad\square$

In the case where f is integrable on $S \times T$, this yields Fubini's Theorem:

Theorem 10.4.3. *Let S and T be aligned rectangles in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and let f be an integrable function on $S \times T$, then*

$$\int_{S \times T} f(x,y) dV(x,y)$$

$$= \underline{\int}_S \underline{\int}_T f(x,y) dV(y) dV(x) = \overline{\int}_S \overline{\int}_T f(x,y) dV(y) dV(x). \tag{10.4.2}$$

Furthermore, if $f(x,y)$ is an integrable function of y on T for each fixed $x \in S$, then $\int_T f(x,y) dV(y)$ is an integrable function of x on S, and

$$\int_{S \times T} f(x,y) dV(x,y) = \int_S \int_T f(x,y) dV(y) dV(x). \tag{10.4.3}$$

Proof. If f is integrable on $S \times T$, then the first and last expressions in the string of inequalities (10.4.1) are equal. Hence, each of the inequalities in (10.4.1) is actually an equality in this case. This proves (10.4.2).

If $f(x,y)$ is an integrable function of y on T for each $x \in S$, then

$$\underline{\int}_T f(x,y) dV(y) = \overline{\int}_T f(x,y) dV(y) = \int_T f(x,y) dV(y)$$

for each $x \in S$. Then (10.4.2) implies that

$$\underline{\int}_S \int_T f(x,y) dV(y) dV(x) = \overline{\int}_S \int_T f(x,y) dV(y) dV(x),$$

which means that $\int_T f(x,y) dV(y)$ is an integrable function of x. Then (10.4.2) implies (10.4.3). $\quad\square$

Remark 10.4.4. In (10.4.2) there is nothing special about the order in which the iterated integrals are taken. The theorem is equally valid if we integrate first with respect to x and then with respect to y. Of course, for the analogue of (10.4.3) to be valid with the order of integration reversed, we must assume that $f(x,y)$ is an integrable function of x for each fixed y.

This leads to the following consequence of Fubini's Theorem.

Theorem 10.4.5. *Let S and T be aligned rectangles in $\mathbb{R}^p$ and $\mathbb{R}^q$, respectively, and let $f(x, y)$ be an integrable function on $S \times T$ which is also integrable as a function of x for each fixed y and integrable as a function of y for each fixed x. Then $\int_S f(x, y)dV(x)$ is an integrable function of y on T and $\int_T f(x, y)dV(y)$ is an integrable function of x on S, and*

$$\int_{S \times T} f(x, y)dV(x, y)$$
$$= \int_S \int_T f(x, y)dV(y)dV(x) = \int_T \int_S f(x, y)dV(x)dV(y). \tag{10.4.4}$$

Note that the integrability conditions in this theorem will all be satisfied if f is a continuous function on the rectangle $S \times T$.

The ability to reverse the order of integration in an iterated integral is a real advantage, as the following example shows.

Example 10.4.6. Find $\displaystyle\int_0^1 \int_0^{\sqrt{\pi}} y^3 \sin(xy^2) \, dydx$.

Solution: Computing the inside integral looks difficult. However, if we reverse the order of integration, the inside integral is just $\int_0^1 y^3 \sin(xy^2)dx = y - y\cos(y^2)$ and the iterated integral becomes

$$\int_0^{\sqrt{\pi}} \int_0^1 y^3 \sin(xy^2) \, dxdy = \int_0^{\sqrt{\pi}} (y - y\cos(y^2))dy = \pi/2.$$

Iterated Integrals over Non-rectangular Regions

A great advantage of integrals in one real variable is that we can often use the Fundamental Theorem of Calculus to calculate them. In order to take advantage of this, we would like to interpret an integral over a Jordan region A in R^d as the result of repeated applications of integration in one variable. Fubini's Theorem is the tool which allows us to do this.

The issue is complicated by the fact that we wish to integrate over a Jordan region, rather than over a rectangle. To do this, we replace the function f to be integrated with f_A, where f is an integrable function on A (then f_A is an integrable function on any aligned rectangle containing A). We then attempt to apply Fubini's Theorem repeatedly to express the integral of f_A over a rectangle containing A as the result of a succession of single variable integrations. In order for this to work, A must have a special form.

We begin with a result which is a direct application of Fubini's Theorem. It will form the basis for the induction argument in the proof of our main theorem. It concerns the case of an integral over a compact Jordan region $A \subset \mathbb{R}^{k+1}$, which is constructed as follows: Suppose there is a compact Jordan region $B \subset R^k$ such that A has the form

$$A = \{(x, t) : x \in B, \text{ and } \psi(x) \leq t \leq \phi(x)\},$$

where ψ and ϕ are continuous functions on B. In this case, $f_A(x,t) = 0$ if $x \notin B$ or if $t \notin [\psi(x), \phi(x)]$. Then (10.4.3) implies

Theorem 10.4.7. *With A, B, ψ, and ϕ as above and f an integrable function on A,*

$$\int_A f(x,t)\, dV(x,t) = \int_B \int_{\psi(x)}^{\phi(x)} f(x,t)\, dt dV(x).$$

provided $f(x,t)$ is an integrable function of t on $[\psi(x), \phi(x)]$ for each $x \in B$.

If we write

$$g(x) = \int_{\psi(x)}^{\phi(x)} f(x,t)\, dt,$$

then the above theorem reduces the problem of computing $\int_A f(x,t) dV(x,t)$ to the problem of computing the lower dimensional integral $\int_B g(x) dV(x)$. This is the basis for the induction argument in the proof of Theorem 10.4.9. Before we state and prove that theorem, we need the following technical result.

Theorem 10.4.8. *Let A, B, ψ, ϕ, and f be as in the previous theorem. If f is continuous on A, then the function*

$$g(x) = \int_{\psi(x)}^{\phi(x)} f(x,t)\, dt$$

is continuous on B.

Proof. Since A is compact and f continuous on A, $|f|$ has a maximum on A. Let M_1 be a positive number greater than or equal to this maximum.

Since ψ and ϕ are continuous on B and $\psi(x) \leq \phi(x)$, the non-negative function $\phi - \psi$ is also continuous and, hence, has a maximum. Let M_2 be a positive number greater than or equal to this maximum.

Let x_0 be a point of B. We will prove that g is continuous at x_0. We need to consider two cases: (1) $\phi(x_0) - \psi(x_0) = 0$, and (2) $\phi(x_0) - \psi(x_0) > 0$.

In case (1), $g(x_0) = 0$. Furthermore, the continuity of $\phi - \psi$ implies that, given $\epsilon > 0$, there is a $\delta > 0$ such that

$$\phi(x) - \psi(x) < \frac{\epsilon}{M_1} \quad \text{whenever} \quad ||x - x_0|| < \delta.$$

Then,

$$|g(x) - g(x_0)| = |g(x)| = \left| \int_{\psi(x)}^{\phi(x)} f(x,t)\, dt \right| \leq M_1(\phi(x) - \psi(x)) < \epsilon.$$

This completes the proof in case (1).

In case (2), we have $\phi(x_0) - \psi(x_0) > 0$. Given $\epsilon > 0$, we may choose a positive number ρ such that

$$\rho < \frac{1}{2}(\phi(x_0) - \psi(x_0)) \quad \text{and} \quad \rho < \frac{\epsilon}{12M_1}.$$

We then set $a = \psi(x_0) + \rho$ and $b = \phi(x_0) - \rho$. Since ψ and ϕ are continuous at x_0, there is a $\delta > 0$ such that

$$|\psi(x) - \psi(x_0)| < \rho \quad \text{and} \quad |\phi(x) - \phi(x_0)| < \rho,$$

whenever $x \in B$ and $||x - x_0|| < \delta$. For each such x, we have

$$\psi(x) < a < b < \phi(x).$$

Also, each of the intervals $[\psi(x), a]$ and $[b, \phi(x)]$ has length less than 2ρ, and so the sum of their lengths is less than 4ρ.

Since f is continuous on the compact set A, it is uniformly continuous on A. Hence, we may choose δ small enough that it is also true that

$$|f(x_1, t_1) - f(x_2, t_2)| < \frac{\epsilon}{3M_2},$$

whenever (x_1, t_1) and (x_2, t_2) are in A and $||(x_1, t_1) - (x_2, t_2)|| < \delta$. In particular,

$$|f(x, t) - f(x_0, t)| < \frac{\epsilon}{3M_2} \quad \text{whenever} \quad ||x - x_0|| < \delta,$$

provided that (x, t) and (x_0, t) are both in A. Then,

$$
\begin{aligned}
|g(x) - g(x_0)| &= \left| \int_{\psi(x)}^{\phi(x)} f(x, t)\, dt - \int_{\psi(x_0)}^{\phi(x_0)} f(x_0, t)\, dt \right| \\
&\leq \left| \int_{\psi(x)}^{\phi(x)} f(x, t)\, dt - \int_{a}^{b} f(x, t)\, dt \right| + \left| \int_{a}^{b} (f(x, t) - f(x_0, t))\, dt \right| \\
&\quad + \left| \int_{a}^{b} f(x_0, t)\, dt - \int_{\psi(x_0)}^{\phi(x_0)} f(x_0, t)\, dt \right| \\
&\leq 4\rho M_1 + \frac{\epsilon}{3M_2} M_2 + 4\rho M_1 = \epsilon.
\end{aligned}
$$

This completes the proof in case (2).   □

We can now state and prove the form of Fubini's Theorem which represents an integral over a Jordan region as the result of repeated single variable integrations.

Theorem 10.4.9. *Suppose f is an integrable function on the closed Jordan region A. Suppose also that A is the set of $x = (x_1, \cdots, x_d) \in R^d$ which satisfy the inequalities*

$$
\begin{aligned}
\psi_1 &\leq x_1 \leq \phi_1, \\
\psi_2(x_1) &\leq x_2 \leq \phi_2(x_1)
\end{aligned}
$$

$$\vdots$$

$$\psi_d(x_1, \cdots, x_{d-1}) \leq x_d \leq \phi_d(x_1, \cdots, x_{d-1}),$$

where ψ_1 and ϕ_1 are numbers and $\psi_j(x_1, \cdots, x_{j-1})$ and $\phi_j(x_1, \cdots, x_{j-1})$ are continuous functions on the set of $(x_1, \cdots, x_{j-1})$ which satisfy the inequalities in this list that precede the jth one. Then

$$\int_A f(x)dV(x)$$
$$= \int_{\psi_1}^{\phi_1} \int_{\psi_2(x_1)}^{\phi_2(x_1)} \cdots \int_{\psi_d(x_1, \cdots, x_{d-1})}^{\phi_d(x_1, \cdots, x_{d-1})} f(x_1, \cdots x_d) \, dx_d \cdots dx_1. \tag{10.4.5}$$

provided that each of the successive iterated integrals exists. This condition is satisfied if f is continuous on A.

Proof. We prove this by induction on d. If $d = 1$, then there is nothing to prove, since the two sides of (10.4.5) are the same integral over an interval in this case.

Now suppose the theorem is true in dimension $d - 1$. To complete the proof we need to prove that it is then true in dimension d. Let A be a Jordan region defined by d inequalities as in the hypothesis of the theorem and let f be an integrable function on A. Let B be the set defined by the first $d - 1$ of these inequalities. Then A, B, and f satisfy the conditions of Theorem 10.4.7. Hence, if $x = (\tilde{x}, x_d)$ where $\tilde{x} = (x_1, \cdots, x_{d-1})$, and $f(\tilde{x}, x_d)$ is an integrable function of x_d on $[\psi_d(\tilde{x}), \phi_d(\tilde{x})]$ for each $\tilde{x} \in B$, then this theorem implies that $g(\tilde{x}) = \int_{\psi(\tilde{x})}^{\phi(\tilde{x})} f(\tilde{x}, x_d) \, dx_d$ is integrable on B and

$$\int_A f(x) \, dV(x) = \int_B \int_{\phi_d(\tilde{x})}^{\psi_d(\tilde{x})} f(\tilde{x}, x_d) \, dx_d dV(\tilde{x}). \tag{10.4.6}$$

Now the set B and the function g satisfy the conditions of our theorem in dimension $d - 1$. Since we are assuming the theorem is true in dimension $d - 1$, we have

$$\int_B g(\tilde{x}) dV(\tilde{x}) = \int_{\psi_1}^{\phi_1} \int_{\psi_2(x_1)}^{\phi_2(x_1)} \cdots \int_{\psi_d(x_1, \cdots, x_{d-2})}^{\phi_d(x_1, \cdots, x_{d-2})} g(x_1, \cdots x_{d-1}) \, dx_{d-1} \cdots dx_1.$$

If we combine this with (10.4.6), the result is (10.4.5).

It remains to prove that each of the successive iterated integrals exists if f is continuous on A. However, this also follows from induction on d. It is clearly true if $d = 1$ since a continuous function on an interval is integrable. Assuming it is true in dimension $d - 1$, then if f is continuous on an A of the form describe in the theorem in dimension d, we conclude that f is continuous, hence, integrable in its last variable and the function g, defined by integrating in this last variable is continuous on the corresponding set B by Theorem 10.4.8. Since we are assuming the result to be true in dimension $d - 1$, we conclude that each of the successive iterated integrals of g exists. Hence, the same thing is true of f. $\square$

Example 10.4.10. Find $\int_A xyz \, dV(x, y, z)$ if A is the Jordan region in $\mathbb{R}^3$ defined by the inequalities $0 \le x \le 1$, $0 \le y \le x$, $0 \le z \le 1 - x^2$.

Solution: According to the previous theorem,

$$
\begin{aligned}
\int_A xyz\, dV(x,y,z) &= \int_0^1 \int_0^x \int_0^{1-x^2} xyz\, dz dy dx \\
&= \int_0^1 \int_0^x \frac{1}{2} xy(1-x^2)^2\, dy dx \\
&= \int_0^1 \frac{1}{4} x^3 (1-x^2)^2\, dx = \frac{1}{4} \int_0^1 (x^3 - 2x^5 + x^7)\, dx \\
&= \frac{1}{4} \left(\frac{1}{4} - \frac{1}{3} + \frac{1}{8} \right) = \frac{1}{96}.
\end{aligned}
$$

Exercise Set 10.4

1. Find the integral of the function g of Exercise 10.3.14 over the square $[-\pi, \pi] \times [-\pi, \pi]$.

2. Evaluate $\displaystyle \int_0^1 \int_0^1 \frac{y^3 x}{(1+y^2 x^2)^2}\, dy dx$.

3. Find the area of the triangle Δ with vertices at $(0,0), (a,0), (a,b)$ by calculating $\int_\Delta 1\, dV(x,y)$ (use Theorem 10.4.9).

4. Calculate the area of a disc of radius one by representing it as the integral of 1 over the disc, expressing this integral as an iterated integral, and then evaluating this iterated integral.

5. Interpret the iterated integral $\displaystyle \int_0^1 \int_{x^2}^x (x^2 + y^2)\, dy dx$ as an integral of $x^2 + y^2$ over a certain Jordan region in $\mathbb{R}^2$. This, in turn, is equal to a certain iterated integral, first with respect to x and then with respect to y. Describe this integral and then evaluate it.

6. Write down an integral in $\mathbb{R}^3$ which represents the volume of a sphere of radius 1. Then express this as a triple iterated integral. You do not need to evaluate this integral.

7. Find $\int_A x dV(x,y,z)$ if A is defined by the inequalities

$$
0 \leq x \leq 1,\ 0 \leq y \leq x^2,\ 0 \leq z \leq x + y.
$$

8. Show that if f and g are continuous real valued functions on a Jordan region $B \subset \mathbb{R}^d$ and $g(x) \leq f(x)$ for all $x \in B$, then the Jordan region $A = \{(x,t) \in \mathbb{R}^{d+1} : x \in B \text{ and } g(x) \leq t \leq f(x)\}$ of Exercise 10.2.12 has volume

$$
V(A) = \int_B (f(x) - g(x))\, dV(x).
$$

9. Prove that if A is any bounded subset of R^p and B is a subset of R^q of volume 0, then $A \times B$ is a subset of $\mathbb{R}^{p+q}$ of volume 0. Use this to prove that the Cartesian product $A \times B$ of two Jordan regions is a Jordan region.

10. Use Fubini's Theorem and the previous exercise to prove that if $A \subset \mathbb{R}^p$ and $B \subset R^q$ are Jordan regions, then $V(A \times B) = V(A)V(B)$.

11. Suppose A is a compact Jordan region in $\mathbb{R}^p$, B is a compact subset of $\mathbb{R}^q$, and f is a continuous function on $B \times A$. Prove that $\int_A f(x,y)dV(y)$ is a continuous function of x on B. Hint: this is similar to but not exactly the same as Theorem 10.4.8.

12. Prove that if $f(t,x)$ is a continuous function on $I \times A$, where I is an open interval in $\mathbb{R}$ and A is a compact Jordan region in $\mathbb{R}^d$, and if $\dfrac{\partial f}{\partial t}(t,x)$ exists and is continuous on $I \times A$, then

$$\frac{d}{dt} \int_A f(t,x)\, dV(x) = \int_A \frac{\partial f}{\partial t}(t,x)\, dV(x).$$

Hint: fix t and consider the function

$$g(h,x) = \begin{cases} \dfrac{f(t+h,x) - f(t,x)}{h} & \text{if } h \neq 0 \\[2mm] \dfrac{\partial f}{\partial t}(t,x) & \text{if } h = 0. \end{cases}$$

Show that this is a continuous function of (h,x) on $J \times A$ for some interval J containing 0 (the Mean Value Theorem is useful in proving this). Then apply the preceding exercise.

10.5 The Change of Variables Formula

Recall the substitution formula (Theorem 5.3.6) from Chapter 5:

$$\int_a^b f(g(t))g'(t)\, dt = \int_{g(a)}^{g(b)} f(u)\, du.$$

Here, if $I = [a, b]$ and $J = g(I)$, then f is assumed continuous on J and g is assumed differentiable with an integrable derivative on I.

This can be thought of as a change of variables formula, where $u = g(t)$ is the transformation from the variable t to the variable u, and the integral formula relates the integral of f as a function of u to an integral involving the composite function $f \circ g$ as a function of t. The formula requires an extra factor $g'(t)$ in the integrand of the latter integral. This is related to how the transformation g changes lengths.

In this section we will derive a similar formula for integrals in several variables. In this case, the extra factor that is needed measures how the transformation changes volume.

Factorization of Matrices

We begin by studying how a linear transformation effects the volume of a Jordan region. The simple way to do this is to factor a given linear transformation as a product of elementary linear transformations whose effect on volume is easy to determine. Such a factorization is given by the process of Gauss elimination (row reduction). The elementary linear transformations in this factorization correspond to the elementary matrices as described below.

The elementary $d \times d$ matrices are of three types:

1. The interchange matrices E_{ij}. For $i \neq j$, the interchange matrix E_{ij} is obtained from the identity matrix by interchanging its ith and jth rows.

2. The shear matrices S_{ij}. For $i \neq j$ the shear matrix S_{ij} is obtained from the identity matrix by adding its jth row to its ith row – that is, by adding a 1 to the ij position in the identity matrix.

3. The scale matrices $T_i(a)$. For $i = 1, \cdots, d$ and $a \neq 0$, $T_i(a)$ is obtained from the identity matrix by multiplying its ith row by the scalar a. – that is, it is the matrix that is a in the ith position on the main diagonal, 1 in the other positions on the main diagonal and 0 in all other positions.

Note that if A is any $d \times d$ matrix, then $E_{ij}A$ is the result of interchanging the ith and jth rows in A and leaving the other rows unchanged, $S_{ij}A$ is the result of adding the jth row of A to its ith row and leaving all but the ith row unchanged, while $T_i(a)A$ is the result of multiplying the ith row of A by a and leaving the other rows unchanged.

The process of Gauss elimination is that of successively multiplying a matrix A on the left by elementary matrices until what is left is a matrix of reduced row echelon form. In the case of a non-singular matrix A its reduced row echelon form is just the identity matrix. Thus, for each non-singular $d \times d$ matrix A there is a matrix B which is a product of elementary matrices and satisfies $BA = I$. Then

$$A = B^{-1}.$$

Note that the inverse of an elementary matrix is an elementary matrix or a product of elementary matrices (Exercise 10.5.1) and so B^{-1} is also a product of elementary matrices. Thus, we have proved

Theorem 10.5.1. *Each non-singular $d \times d$ matrix A is a product of matrices of the form $E_{ij}, S_{ij}, T_i(a)$.*

The determinants of the elementary matrices are easily calculated.

Theorem 10.5.2. *For each i and each $j \neq i$ we have $\det E_{ij} = 1$, $\det S_{ij} = 1$, and $\det T_i(a) = a$.*

Since the determinant is multiplicative ($\det AB = \det A \det B$ for all pairs A, B of $d \times d$ matrices), it follows that the determinant of a given non-singular matrix A is just the product of the scale factors a that appear in its factorization as a product of elementary matrices.

Linear Transformations and Volume

We wish to understand how the volume of a Jordan region is effected by a linear transformation. Some linear transformations clearly have no effect on volume. A transformation that takes each aligned rectangle to an aligned rectangle of the same volume has no effect on the volume of a Jordan region. The elementary interchanges E_{ij} have this property. The shear matrices S_{ij} also preserve volumes of Jordan regions, but the proof of this fact is a little more complicated.

Theorem 10.5.3. *A shear transformation S_{ij} takes a Jordan region to a Jordan region of the same volume.*

Proof. The shear matrix S_{12} on $\mathbb{R}^2$ is the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

It takes the aligned rectangle $[a, b] \times [c, d]$, which has vertices $(a, c), (b, c), (b, d)$, and (a, d) to the parallelogram with vertices $(a + c, c), (b + c, c), b + d, d)$, and $(a + d, d)$. This parallelogram has base of length $(b + c) - (a + c) = b - a$ and height $d - c$. Thus, its area is $(b - a)(d - c)$ (Exercise 10.2.11), which is the same as the volume of the original rectangle.

In general, an aligned rectangle R in $\mathbb{R}^d$ for $d > 2$ has the form $S \times T$ where S is an aligned rectangle in $\mathbb{R}^2$ and T is an aligned rectangle in $\mathbb{R}^{d-2}$. The shear transformation S_{12} on $\mathbb{R}^d$ sends this to $P \times T$ where, by the above discussion, P is a parallelogram with the same area as S. It follows from this and Exercise 10.4.10 that S_{12} sends R to a Jordan region with the same volume as R. Since, for any $i \neq j$, S_{ij} is just S_{12} composed with some elementary interchanges, it follows that it also takes an aligned rectangle to a Jordan region with the same volume.

Let A be a Jordan region, R an aligned rectangle containing A, and P a partition of R. Let $R_1, R_2, \cdots, R_n$ be a list of the subrectangles of R determined by the partition P. Set

$$E = \bigcup \{R_k : R_k \subset A\}$$
$$F = \bigcup \{R_k : R_k \cap A \neq \emptyset\}.$$

Then $U(\chi_A, P) = V(F)$ and $L(\chi_A, P) = V(E)$. Since A is a Jordan region, given $\epsilon > 0$, there is a partition P such that $V(F) - V(E) < \epsilon$. Of course, regardless of how the partition is chosen

$$V(E) \leq V(A) \leq V(F). \tag{10.5.1}$$

Note $S_{ij}F$ is the union of those $S_{ij}R_k$ such that $R_k \cap A \neq \emptyset$, and any two of these sets meet (if at all) in a set of volume 0. Since $V(S_{ij}R_k) = V(R_k)$, we conclude that
$$V(S_{ij}F) = V(F).$$

A similar argument shows that

$$V(S_{ij}E) = V(E).$$

Hence,

$$V(E) = V(S_{ij}E) \le \underline{V}(S_{ij}A) \le \overline{V}(S_{ij}A) \le V(S_{ij}F) = V(F). \qquad (10.5.2)$$

Since, $V(F) - V(E) < \epsilon$, we conclude that

$$\overline{V}(S_{ij}A) - \underline{V}(S_{ij}A) < \epsilon.$$

Since ϵ was arbitrary, this difference is actually 0. This proves that $S_{ij}A$ is a Jordan region. That it has the same volume as A follows from (10.5.1) and (10.5.2). $\qquad \square$

Theorem 10.5.4. *If $L : \mathbb{R}^d \to \mathbb{R}^d$ is a linear transformation and E is a Jordan region, then $L(E)$ is also a Jordan region and $V(L(E)) = |\det L| V(E)$, where $\det L$ denotes the determinant of the matrix corresponding to L.*

Proof. We first note that if this theorem is true for linear transformations L_1 and L_2, then it is also true for the composition $L_1 \circ L_2$, by the following computation:

$$\begin{aligned} V(L_1 \circ L_2(E)) &= |\det L_1| V(L_2(E)) \\ &= |\det L_1| |\det L_2| V(E) = |\det L_1 L_2| V(E), \end{aligned}$$

since determinant and absolute value are both multiplicative functions.

The elementary interchanges E_{ij} and shear transformations S_{ij} do not effect volume and they are matrices of determinant ± 1. Thus, the theorem is true for these linear transformations.

The scale matrix $T_i(a)$ takes each aligned rectangle to an aligned rectangle with edges of the same length as the original except for the ith edge, which has its length multiplied by $|a|$. Hence, each aligned rectangle is sent to an aligned rectangle of volume $|a|$ times the volume of the original. It follows that $T_i(a)$ takes a Jordan region to another Jordan region with volume $|a|$ times the volume of the original. Since $a = \det T_i(a)$, the theorem is true for the transformations $T_i(a)$.

Since every non-singular $d \times d$ matrix is a product of interchanges, shear transformations, and scale transformations, the theorem is true for all non-singular linear functions from $\mathbb{R}^d$ to $\mathbb{R}^d$.

If L is singular, then its determinant is 0. Thus, to finish the proof, we need to show that if L is a singular linear transformation, then $L(E) = 0$ for every Jordan region E. We leave this as an exercise. $\qquad \square$

Example 10.5.5. If $L : \mathbb{R}^2 \to \mathbb{R}^2$ is the linear transformation with matrix

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

what is the area of the image of the unit disc $D_1(0,0)$ under the transformation L?

Solution: The unit disc has area π. By the previous theorem, its image under L has area $|\det L|\pi = 2\pi$.

Example 10.5.6. What is the area of an ellipse, with two vertices at distance 3 from $(0,0)$ along the line $y = x$ and two vertices at distance 2 from $(0,0)$ along the line $y = -x$?

Solution: This ellipse may be obtained from the unit disc by first applying the transformation with matrix

$$\begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$$

and then applying the linear transformation which is rotation through the angle $\pi/4$. The first transformation has determinant 6, while the second has determinant 1. Hence the area of the indicated ellipse is 6π.

Smooth Image of a Rectangle

We will prove that, under appropriate conditions, the image of an aligned rectangle under a smooth map is a Jordan region. We first prove the the image of a degenerate rectangle under such a map is a set of volume 0.

Theorem 10.5.7. *Let ϕ be a one-to-one smooth transformation from an open set $U \subset \mathbb{R}^p$ to R^p and suppose $d\phi(x)$ is non-singular at each point of U. If R is a degenerate aligned rectangle contained in U, then $\phi(R)$ is a set of volume 0 in $\mathbb{R}^p$.*

Proof. Since R is degenerate, it is a rectangle of dimension at most $p - 1$. We may as well assume that it is contained in $\mathbb{R}^{p-1} = \{x = (x_1, \cdots, x_p) : x_p = 0\}$. Let a be a point of R. We will show first that there is a neighborhood of $b = \phi(a)$ whose intersection with $\phi(R)$ has volume 0. If we can do this for each $a \in R$, then, since $\phi(R)$ is compact, we may cover $\phi(R)$ with finitely many open sets whose intersections with $\phi(R)$ have volume 0. It follows from this that $\phi(R)$ itself has volume 0.

Since translations do not effect volume, we may as well assume that a and $b = \phi(a)$ are both equal to 0. Also, since applying a non-singular linear transformation does not effect whether or not a set has volume 0, we may replace ϕ by $(d\phi(0))^{-1}\phi$. In other words, we may as well assume that $d\phi(0) = I$ – the identity transformation.

If $\phi = (\phi_1, \cdots, \phi_p)$, and points of $\mathbb{R}^p$ are denoted (x, y) with $x \in \mathbb{R}^{p-1}$ and $y \in \mathbb{R}$, then we define $g : U \cap \mathbb{R}^{p-1} \to \mathbb{R}^{p-1}$ by

$$g(x) = (\phi_1(x, 0), \cdots, \phi_{p-1}(x, 0)).$$

Then $dg(0)$ is the upper left $(p-1) \times (p-1)$ subdeterminant of $d\phi(0)$ and so it too is the identity transformation. The Inverse Function Theorem then implies

that there are neighborhoods V and W of 0 in $\mathbb{R}^{p-1}$ such that g maps V onto W and has a smooth inverse $g^{-1} : W \to V$. Then

$$\phi(g^{-1}(x), 0) = (x, \phi_p \circ g^{-1}(x))$$

for $x \in W$. That is, the part of $\phi(R)$ consisting of points with first coordinate in W is the graph of the smooth function $\phi_p \circ g^{-1}$. It therefore has volume 0 by Example 10.2.11. This completes the proof. □

Theorem 10.5.8. *Let $\phi : U \to \mathbb{R}^p$ satisfy the conditions of the previous theorem. If R is a rectangle in U, then $\phi(R)$ is a Jordan region.*

Proof. If R is a rectangle in U, then its boundary is a union of finitely many rectangles of dimension $p-1$ – that is, it is the union of finitely many degenerate rectangles. The image of each of these under ϕ has volume 0 by the previous theorem. Hence, $\phi(\partial R)$ has volume zero. The proof will be complete if we can show that $\partial \phi(R) = \phi(\partial R)$.

The image of ϕ is an open set V by Exercise 9.6.8, and $\phi : U \to V$ is one-to-one and onto. Thus, ϕ has an inverse transformation $\phi^{-1} : V \to U$ which is a smooth transformation, by the Inverse Function Theorem. It is, in particular, continuous. Since both ϕ and ϕ^{-1} are continuous, a subset $A \subset U$ is open if and only if its image $\phi(A) \subset V$ is open. It follows that ϕ takes the interior of R to the interior of $\phi(R)$ and, hence, the boundary of R to the boundary of $\phi(R)$. □

Integral over the Smooth Image of a Rectangle

Our next objective is to prove the change of variables formula for integration over a rectangle. We will need the following lemma, which says that the relative error in approximating the volume of the image of a rectangle under a smooth map by the volume of its image under the differential of the map can be made arbitrarily small. In the lemma, it is crucial that we don't allow rectangles R to become too skinny. By this, we mean that we don't want the ratio of the length of the shortest edge of R to the diameter of R (greatest distance between two points of R) to be too small. We will call this ratio the *aspect ratio* of the rectangle.

Lemma 10.5.9. *Let λ and K be positive constants. Let U be an open subset of $\mathbb{R}^p$ and $\phi : U \to R^p$ a smooth one-to-one transformation. Suppose $d\phi(a)$ is non-singular and $|\det d\phi(a)| \leq K$ for all $a \in U$. Then, given $\epsilon > 0$, there is a $\delta > 0$ such that if R is a rectangle in U with diameter less than δ and aspect ratio at least λ, then $|V(\phi(R)) - V(d\phi(a)R)| < \epsilon V(R)$, where a is the center of the rectangle R.*

Proof. Let R be a rectangle in U with diameter less than a positive number δ to be detemined below and aspect ratio at least λ. Note that $\phi(R)$ is a Jordan region, by the previous theorem and, hence, it has volume.

Since translation does not effect volume, we may assume that the center of the rectangle R is 0 and $\phi(0) = 0$. By hypothesis

$$|\det d\phi(0)| \leq K. \tag{10.5.3}$$

If $0 < \rho < 1$, then $(1 + \rho)R$ is the rectangle created from R by expanding each edge in a symmetric way about its center by the factor $(1 + \rho)$. Similarly, $(1 - \rho)R$ is the rectangle created from R by shrinking each edge in a symmetric way about its center by the factor $1 - \rho$. Also,

$$(1 - \rho)R \subset R \subset (1 + \rho)R,$$

and, since $d\phi(0)$ is linear,

$$(1 - \rho)d\phi(0)R \subset d\phi(0)R \subset (1 + \rho)d\phi(0)R.$$

Comparing volumes and using (10.5.3) yields,

$$
\begin{aligned}
V((1 + \rho)d\phi(0)R) &- V((1 - \rho)d\phi(0)R) \\
&= ((1 + \rho)^d - (1 - \rho)^d)V(d\phi(0)R) \\
&= ((1 + \rho)^d - (1 - \rho)^d)|\det d\phi(0)|V(R) \qquad (10.5.4) \\
&\leq 2\rho d(1 + \rho)^{d-1}|\det d\phi(0)|V(R) \\
&\leq 2^d \rho dK V(R).
\end{aligned}
$$

If we choose

$$\rho = \frac{\epsilon}{2^d dK},$$

then it follows from (10.5.4) that

$$V((1 + \rho)d\phi(0)R) - V((1 - \rho)d\phi(0)R) \leq \epsilon V(R).$$

The proof will be complete if we can show that, for small enough δ, any rectangle R containing 0, of diameter less than δ, satisfies

$$(1 - \rho)d\phi(0)R \subset \phi(R) \subset (1 + \rho)d\phi(0)R, \tag{10.5.5}$$

since these containments are also satisfied with $\phi(R)$ replaced by $d\phi(0)R$.

If x is any non-zero vector in $\mathbb{R}^d$, then

$$||x|| = ||(d\phi(0))^{-1}d\phi(0)x|| \leq ||(d\phi(0))^{-1}||\,||d\phi(0)x||.$$

Thus, $||d\phi(0)x|| \geq ||(d\phi(0))^{-1}||^{-1}||x||$. In other words, if L is any line segment in $\mathbb{R}^d$, then the length of the line segment $d\phi(0)L$ is at least the factor

$$A = ||(d\phi(0))^{-1}||^{-1}$$

times the length of L. It follows that the distance from $d\phi(0)R$ to the complement of $(1 + \rho)d\phi(0)R$ is at least $A\rho r$, where r is one half the length of the

shortest edge of R. By the definition of the differential $d\phi(0)$, we may choose δ such that $||x|| < \delta$ and $x \in R$ implies

$$||\phi(x) - d\phi(0)x|| < A\rho\lambda||x|| < A\rho r.$$

This implies that $\phi(x) \in (1+\rho)d\phi(0)R$. A similar argument shows that, with, δ chosen as above, $x \in R$ implies that $(1-\rho)d\phi(0)x \in \phi(R)$. Hence, (10.5.5) holds if R has diameter less than δ. This completes the proof. □

Theorem 10.5.10. *Let U be an open subset of $\mathbb{R}^p$ and $\phi : U \to R^p$ a smooth one-to -one transformation with $d\phi$ non-singular at each point of U. Let R be an aligned rectangle in U and f a continuous function on $\phi(R)$. Then*

$$\int_{\phi(R)} f(u)\,dV(u) = \int_R f(\phi(x))\,|\det d\phi(x)|\,dV(x).$$

Proof. For each subrectangle S of R we set

$$\Delta(S) = \int_{\phi(S)} f(u)\,dV(u) - \int_S f(\phi(x))\,|\det d\phi(x)|dV(x),$$

$$Q(S) = \frac{\Delta(S)}{V(S)}.$$

To prove the theorem, we need to show that $\Delta(R) = 0$. This is equivalent to showing that $Q(R) = 0$.

Let h be the diameter of R. We will choose inductively a downwardly nested sequence $\{R_i\}_{i=0}^\infty$ of subrectangles of R in such a way that R_i has diameter $h/2^i$ and $|Q(R_i)| \geq |Q(R)|$. We begin by setting $R_0 = R$.

Suppose $R_0, \cdots, R_m$ have been chosen in such a way that the conditions of the previous paragraph are met. If $R_m = [a_1, b_1] \times \cdots \times [a_p, b_p]$, we partition R_m by partitioning each interval $[a_k, b_k]$ into two subintervals of equal length. There are 2^p subrectangles of R_m for this partition and each of them has diameter $h/2^{m+1}$ since R_m has diameter $h/2^m$. If $\{S_1, \cdots, S_n\}$ is a list of these subrectangles of R_m, then $R_m = \cup_j S_j$ and

$$\Delta(R_m) = \sum_{j=1}^n \Delta(S_j) = \sum_{j=1}^n Q(S_j)V(S_j).$$

For at least one of the rectangles S_j, we must have $|Q(S_j)| \geq |Q(R_m)|$, for if $|Q(S_j)| < |Q(R_m)|$ for all j, then

$$\Delta(R_m) = \sum_{j=1}^n Q(S_j)V(S_j) < \sum_{j=1}^n Q(R_m)V(S_j) = Q(R_m)V(R_m) = \Delta(R_m),$$

which is impossible. Thus, for some j, we have $|Q(S_j)| \geq |Q(R_m)|$. We choose R_{m+1} to be an S_j which satisfies this inequality. This proves by induction that a sequence $\{R_i\}$ with the required properties can be chosen.

Since the sequence $\{R_i\}$ is a downwardly nested sequence of compact sets, it has a non-empty intersection. Let a be a point in this intersection.

Since ϕ is smooth, we may choose a neighborhood V of a in which $|\det d\phi(x)|$ is bounded above by a positive constant K. If λ is the aspect ratio of R, then each of the rectangles R_j has the same aspect ratio. By the previous lemma, there is a $\delta > 0$ such that each rectangle R in V with aspect ratio at least λ and with diameter less than δ satisfies

$$|V(\phi(R)) - V(d\phi(b)R)| < \epsilon V(R),$$

where b is the center of the rectangle R. These conditions will be met for all R_j with $R_j \subset B_\delta(a)$. We will denote the center of R_j by a_j. If we also choose δ small enough that

$$|f(\phi(x)) - f(\phi(y))| < \epsilon, \quad \text{and} \quad |f(\phi(x))|\det d\phi(x)| - f(\phi(y))|\det d\phi(y)|| < \epsilon$$

for all $x, y \in B_\delta(a)$, then

$$|\Delta(R_j)| = \left| \int_{\phi(R_j)} f(u)\, dV(u) - \int_{R_j} f(\phi(x))\,|\det d\phi(x)| dV(x) \right|$$

$$\leq \left| \int_{\phi(R_j)} f(\phi(a_j))\, dV(u) - \int_{R_j} f(\phi(a_j))\,|\det d\phi(a_j)| dV(x) \right|$$

$$+ \int_{\phi(R_j)} |f(u) - f(\phi(a_j))|\, dV(u)$$

$$+ \int_{R_j} |f(\phi(x))\,|\det d\phi(x)| - f(\phi(a_j))\,|\det d\phi(a_j)|| dV(x)|$$

$$\leq |f(\phi(a_j))|\,|V(\phi(R_j)) - V(d\phi(a_j)R_j)| + \epsilon V(\phi(R_j)) + \epsilon V(R_j).$$

Since $|V(\phi(R_j)) - V(d\phi(a_j)R_j)| < \epsilon V(R_j)$ and $V(d\phi(R_j)) = |\det \phi(a_)|V(R_j)$, it follows that

$$|\Delta(R_j)| \leq \epsilon V(R_j)(|f(\phi(a_j))| + |\det d\phi(a_j)| + \epsilon + 1).$$

Since ϵ was arbitrary and $\phi(a_j) \to \phi(a)$ and $d\phi(a_j) \to d\phi(a)$ as $j \to \infty$, this implies that $Q(R_j) = \Delta(R_j)/V(R_j)$ can be made smaller than any positive number by choosing j large enough. Since $Q(R) \leq Q(R_j)$ for all j, this implies that $Q(R) = 0$, as required. $\qquad\square$

This has the following corollary, the proof of which is left to the exercises.

Corollary 10.5.11. *Let U be an open subset of $\mathbb{R}^d$ and $\phi : U \to \mathbb{R}^d$ a smooth one-to-one transformation with non-singular differential on U. If R is an aligned rectangle in U, then*

$$V(\phi(R)) = \int_R |\det d\phi(x)|\, dV(x).$$

Furthermore, if $M = \sup_R |\det d\phi|$ and $m = \inf_R |\det d\phi|$, then

$$mV(R) \leq V(\phi(R)) \leq MV(R).$$

Integral over the Smooth Image of a Jordan Region

We can now prove the general change of variables formula. The proof uses the following lemma, which follows easily from the previous corollary. The proof is left to the exercises.

Lemma 10.5.12. *If $\phi : U \to \mathbb{R}^d$ is a smooth one-to-one function with $d\phi$ non-singular on U and if $K \subset U$ is a compact set of volume 0, then $\phi(K)$ is also a set of volume 0.*

Theorem 10.5.13. *Let A be a compact Jordan region contained in an open set $U \subset \mathbb{R}^d$. Let $\phi : U \to \mathbb{R}^d$ be a smooth one-to-one function with a differential which is non-singular on A, and let f be a function which is bounded on $\phi(A)$ and continuous except on a subset E of $\phi(A)$ of volume 0. Then, $\phi(A)$ is a Jordan region, f is integrable on $\phi(A)$, $f \circ \phi$ is integrable on A and*

$$\int_{\phi(A)} f(u)\, dV(u) = \int_A f(\phi(x)) |\det d\phi(x)|\, dV(x).$$

Proof. Let $V = \phi(U)$. By the Inverse Function Theorem, V is an open set and $\phi^{-1} : V \to U$ is a smooth function with non-singular differential.

The boundary of A is a set of volume 0 since A is a Jordan region. Since ϕ and ϕ^{-1} are both continuous, $\partial \phi(A) = \phi(\partial A)$. It follows from the previous lemma that $\partial \phi(A)$ is also a set of volume 0 and, hence, that $\phi(A)$ is a Jordan region. Hence, we may extend f to be 0 on the complement of $\phi(A)$ in V and it will still be a function which is continuous except on a set of volume 0. It follows from Theorem 10.3.5 that f is integrable on $\phi(A)$.

Let K be the closure of $\partial \phi(A) \cup E$. Then f, extended to be 0 on the complement of $\phi(A)$, is continuous on the complement of K. The set K has volume 0. Hence, by the previous lemma, $\phi^{-1}(K)$ is a set of volume 0. Since $f \circ \phi$ is continuous on U except at points of $\phi^{-1}(K)$, it follows that $f \circ \phi$ is integrable on A.

Let ϵ be any positive number. Let R be a rectangle containing A and P a partition of R. We choose P so that $R_1, R_2, \cdots, R_n$ is a list of those rectangles for this partition which are contained in U. If the partition is fine enough, then it will be true that $A \subset \cup_j R_j$. Also, the partition may be chosen fine enough that, if S is the set of j for which $R_j \cap K \neq \emptyset$, then

$$\sum_{j \in S} V(R_j) < \epsilon.$$

If $K \cap R_j = \emptyset$, then either $A \cap R_j = \emptyset$ or R_j is a rectangle contained in the interior of A and f is continuous on $\phi(R_j)$. If the latter is true, then

$$\int_{\phi(R_j)} f(u)\, dV(u) = \int_{R_j} f(\phi(x)) |\det d\phi(x)|\, dV(x).$$

Since f is 0 on the complement of $\phi(A)$, we have

$$\left| \int_{\phi(A)} f(u)\, dV(u) - \int_A f(\phi(x)) |\det \phi(x)|\, dV(x) \right|$$

$$= \left| \sum_j \left(\int_{\phi(R_j)} f(u)\, dV(u) - \int_{R_j} f(\phi(x)) |\det \phi(x)|\, dV(x) \right) \right|$$

$$= \left| \sum_{j \in S} \left(\int_{\phi(R_j)} f(u)\, dV(u) - \int_{R_j} f(\phi(x)) |\det \phi(x)|\, dV(x) \right) \right|$$

$$\leq \sum_{j \in S} \left(\int_{\phi(R_j)} M\, dV(u) + \int_{R_j} MK\, dV(x) \right)$$

$$= \sum_{j \in S} (MV(\phi(R_j) + MKV(R_j)) \leq 2MK\epsilon.$$

where $M = \sup_A |f(\phi(x)))|$ and $K = \sup_A |\det d\phi(x)|$. Since, ϵ is arbitrary, this implies the equality of the theorem. □

With some additional hypotheses, the above theorem can be strengthened so as to apply to integrals over the full open sets U and $\phi(U)$ rather that just to integrals over compact subsets. The next theorem is such a result.

Theorem 10.5.14. *Let U be an open Jordan region in $\mathbb{R}^d$ and let $\phi : A \to \mathbb{R}^d$ be a one to one smooth fuction on U with image $\phi(U)$ which is also a Jordan region. Suppose $d\phi$ is non-singular on U and f is bounded on $\phi(U)$ and continuous except on a subset of volume 0. Then f is integrable on $\phi(U)$. If, in addition, $f \circ \phi |\det \phi|$ is bounded on U, then it too is integrable on U and*

$$\int_{\phi(U)} f(u)\, dV(u) = \int_U f(\phi(x)) |\det d\phi(x)|\, dV(x).$$

Proof. Since $d\phi$ is non-singular on U, Theorem 9.6.5 implies that $\phi : U \to \mathbb{R}^d$ is a one-to-one open map onto an open set V.

Since f is bounded on $\phi(U)$ and is continuous except on a set of volume 0, it is integrable on $\phi(U)$. The function $g(x) = f(\phi(x)) |\det d\phi(x)|$ is continuous and bounded and, hence, is an integrable function on U.

Let K_n be a sequence of compact Jordan subsets of U such that $\cup_n K_n^\circ = U$. Such a sequence exists by Exercise 10.3.11. Then, by Exercise 10.3.10,

$$\int_U g(x)\, dV(x) = \lim_n \int_{K_n} g(x)\, dV(x). \qquad (10.5.6)$$

Also, since $\{\phi(K_n)\}$ is a sequence of compact subsets of $V = \phi(U)$ with the union of the interiors of the sets in the sequence equal to V, we conclude

$$\int_{\phi(U)} f(u)\, dV(u) = \lim_n \int_{\phi(K_n)} f(u)\, dV(u). \qquad (10.5.7)$$

The previous theorem implies that

$$\int_{\phi(K_n)} f(u)\, dV(u) = \int_{K_n} g(x)\, dV(x),$$

for each n. This, together with (10.5.6) and (10.5.7), completes the proof.   □

The change of variables theorem has the following corollary, the proof of which is left to the exercises.

Corollary 10.5.15. *Let U be an open Jordan region in $\mathbb{R}^d$ and $\phi : U \to \mathbb{R}^d$ a function satisfying the conditions of the previous theorem. Then*

$$V(\phi(U)) = \int_U |\det d\phi(x)|\, dV(x).$$

Note that, in the change of variables formulas in the above theorem and its corollary, the sets U and $\phi(U)$ may be replaced by their closures, even though the transformation ϕ may not be defined on the closure of U. This is due to the fact that the boundaries of U and $\phi(U)$ have volume 0.

Example 10.5.16. Use the preceding corollary to find the area enclosed by an ellipse with major and minor axes of lengths $2a$ and $2b$ without assuming knowledge of the area of a circle.
Solution: Such an ellipse has equation $x^2/a^2 + y^2/b^2 = 1$. The region it encloses is the image of the square $A = \{(r,\theta) : 0 \le r \le 1, 0 \le \theta \le 2\pi\}$, under the transformation $\phi(r,\theta) = (ar\cos\theta, br\sin\theta)$. The differential of this map is

$$d\phi(r,\theta) = \begin{pmatrix} a\cos\theta & -ar\sin\theta \\ b\sin\theta & br\cos\theta \end{pmatrix}$$

The determinant of this matrix is abr, which is non-zero except at $r = 0$. Thus, the function ϕ is one to one and smooth with non-singular differential on the interior of the square A. The interior of A is taken by ϕ to the interior of the ellipse with the line joining $(0,0)$ to $(1,0)$ removed. This set differs from the ellipse itself by a set of volume 0. Thus, the area we seek is, by the previous corollary and Fubini's Theorem,

$$\int_0^{2\pi} \int_0^1 abr\, dr\, d\theta = \pi\, ab.$$

Example 10.5.17. Find $\displaystyle\int_0^1 \int_0^{\sqrt{1-x^2}} \cos(x^2 + y^2)\, dydx$.
Solution: By Fubini's Theorem, this integral is

$$\int_D \cos(x^2 + y^2)\, dV(x,y),$$

where $D = B_1(0,0)$. If we change to polar coordinates using the transformation

$$\phi(r,\theta) = (r\cos\theta, r\sin\theta),$$

then $\det d\phi(r, \theta) = r$ and $D = \phi(R)$, where R is the rectangle $[0, 1] \times [0, 2\pi]$. On R, ϕ is smooth with non-singular differential except when $r = 0$, and so Theorem 10.5.14 applies with $U = R^\circ$. Hence,

$$\int_{\phi(R)} \cos(x^2 + y^2) \, dV(x, y) = \int_R \cos(r^2) \, rdrd\theta.$$

Applying Fubini's Theorem again yields

$$\int_0^1 \int_0^{\sqrt{1-x^2}} \cos(x^2 + y^2) \, dydx = \int_0^{2\pi} \int_0^1 \cos(r^2) \, rdrd\theta = \pi \sin 1.$$

Exercise Set 10.5

1. Compute the inverse of each elementary matrix E_{ij}, S_{ij}, and $T_i(a)$. Show that each inverse is itself an elementary matrix or a product of elementary matrices.

2. Show that if E is a Jordan region and L is a linear transformation whose matrix is singular, then $L(E)$ has volume 0.

3. Let u and v be two vectors in the plane and define $\phi : \mathbb{R}^2 \to \mathbb{R}^2$ by $\phi(s, t) = su + tv$. Let A be the parallelogram which is the image of $[0, 1] \times [0, 1]$ under ϕ. If f is a continuous function on A, express $\int_A f(x, y) \, dV(x, y)$ as an integral over $[0, 1] \times [0, 1]$.

4. Use the result of the previous exercise to find a formula for the area of the parallelogram determined by two vectors u and v.

5. An orthogonal transformation is a linear transformation A that preserves inner products – that is, $Au \cdot Av = u \cdot v$ for each pair of vectors u, v. Note that a rotation is an orthogonal transformation. Prove that a $d \times d$ orthogonal transformation preserves volume in $\mathbb{R}^d$.

6. Compute $\int_0^a \int_0^{\sqrt{a^2-x^2}} e^{x^2+y^2} \, dydx$.

7. Let $A = \{(x, y) \in \mathbb{R}^2 : x \geq 0, y \geq 0, x^2 + y^2 \leq 4, x^2 - y^2 \geq 1\}$. Compute

$$\int_A \frac{xy}{x^2 + y^2} \, dV(x, y)$$

   by making a change of variables $u = x^2 + y^2, v = x^2 - y^2$ for $x \geq 0$, $y \geq 0$.

8. Compute the volume of a sphere S of radius r by computing the integral

$$\int_S 1 \, dV(x).$$

   Compute this integral by first converting to spherical coordinates.

9. Compute the volume of a right circular cone with height h and radius a. Hint: such a cone can be described in cylindrical coordinates as the set of points

$$\{(r, \theta, z) : 0 \leq r \leq \frac{a}{h}z, 0 \leq \theta \leq 2\pi\}.$$

Here $x = r\cos\theta, y = r\sin\theta, z = z$ describes the transformation from cylindrical to rectangular coordinates.

10. Show by example that the conclusion of Theorem 10.5.13 does not hold if the function ϕ is not one-to-one on A.

11. Prove Corollary 10.5.11

12. Prove Lemma 10.5.12.

Chapter 11

Vector Calculus

Previous chapters have dealt with integration over intervals on the line and over Jordan domains in $\mathbb{R}^d$. In this chapter, we study integration over curves and surfaces in $\mathbb{R}^d$. Here, the surfaces involved could be ordinary two dimensional surfaces in $\mathbb{R}^3$, but they might be p-surfaces in $\mathbb{R}^d$ for any $p \leq d$. In this study, the objects to be integrated are no longer functions, but closely related objects called *differential forms*. Differential forms, like surfaces, have a dimension. Thus, there is a notion of a p-form for each non-negative integer p. When a differential form is integrated over a surface, the dimensions must match. Thus, we integrate p-forms over p-surfaces.

The culmination of this study is a series of very powerful theorems – Green's Theorem, Gauss's Theorem, Stokes's Theorem – which are really all special cases of one very general theorem, which is also usually called Stokes's Theorem.

11.1   1-Forms and Path Integrals

We begin with the one dimensional case: curves and integration of 1-forms over curves.

Smooth Curves

Recall from Section 9.4 that a curve in $\mathbb{R}^d$ is a continuous function $\gamma : I \to \mathbb{R}^d$ which has an interval I on the line as its domain. The interval I is called the parameter interval for the curve. We will be focusing on curves which have a derivative γ' on the interior of I. The derivative is defined in the usual way:

$$\gamma'(t) = \lim_{s \to t} \frac{\gamma(s) - \gamma(t)}{s - t}.$$

Note that if the curve $\gamma(t)$ is expressed in terms of its coordinate functions, $\gamma(t) = (\gamma_1(t), \gamma_2(t), \cdots, \gamma_d(t))$, then $\gamma'(t) = (\gamma_1'(t), \gamma_2'(t), \cdots, \gamma_d'(t))$.
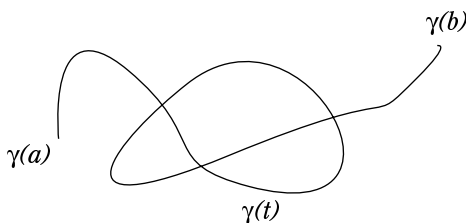
Figure 11.1: A Smooth Curve in $\mathbb{R}^2$

Definition 11.1.1. A smooth curve γ is a curve with a bounded, continuous derivative γ' on the interior of its parameter interval I.

The *trace* of a curve γ with parameter interval I is its image $\gamma(I)$ in $\mathbb{R}^d$. A curve is said to lie in the subset E of $\mathbb{R}^d$ if its trace is contained in E.

Example 11.1.2. Find a smooth curve which traces a straight line from u to v in $\mathbb{R}^d$. What is the derivative of this curve?

Solution: The curve γ, defined by $\gamma(t) = u + t(v - u)$, $t \in [0, 1]$, begins at $u = \gamma(0)$, moves in the direction of the vector $v - u$ as t increases, and ends at $v = \gamma(1)$. The derivative of γ is the constant vector $\gamma'(t) = v - u$.

Piecewise Smooth Curves – Paths

We will also need to consider curves which are only *piecewise smooth* – that is curves which have a bounded, continuous derivative except at finitely many points of the parameter interval I. The precise definition is as follows:

Definition 11.1.3. Let $\gamma : I \to \mathbb{R}^d$ be a curve. We will say that γ is *piecewise smooth* if there is a partition $a = t_0 < t_1 < t_2 < \cdots < t_n = b$ of I such that , for each j, the restriction of γ to the subinterval $[t_{j-1}, t_j]$ is a smooth curve. A piecewise smooth curve will also be called a *path*.

If γ is a path as described above, then γ' exists and is continuous and bounded on $I \setminus \{t_0, \cdots, t_n\}$.

One may think of a path as finitely many smooth curves which join together to form a single curve which is smooth everywhere except at points where two of the original curves join. At such points the curve may abruptly change direction.

Example 11.1.4. Find a path that traces once around the square with vertices $(0, 0), (1, 0), (1, 1), (0, 1)$ in the counterclockwise direction. Find $\gamma'(t)$ on the subintervals where γ is smooth.

Solution: We choose $[0, 1]$ as the parameter interval and define a path γ as follows:

$$\gamma(t) = \begin{cases} (4t, 0) & 0 \le t \le 1/4 \\ (1, 4t - 1) & 1/4 \le t \le 1/2 \\ (3 - 4t, 1) & 1/2 \le t \le 3/4 \\ (0, 4 - 4t) & 3/4 \le t \le 1 \end{cases}$$
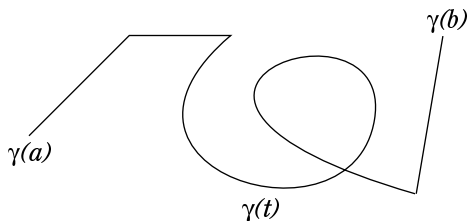
Figure 11.2: A Path in $\mathbb{R}^2$

This is continuous on $[0, 1]$ and smooth on each subinterval in the partition $0 < 1/4 < 1/2 < 3/4 < 1$. It traces each side of the square in succession, moving in the counterclockwise direction. On the first interval, γ' is the constant vector $(4, 0)$, on the second it is $(0, 4)$, on the third it is $(-4, 0)$, and on the fourth it is $(0, -4)$.

Closed Paths

The preceding example is an example of a *closed path* – that is, a path γ which begins and ends at the same point. This means that $\gamma(a) = \gamma(b)$, where $[a, b]$ is the parameter interval. The following is another example of a closed path:

Example 11.1.5. Find a path which traces once around the circle of radius r in $\mathbb{R}^2$, centered at u_0.

Solution: The circle of radius r centered at u_0 consists of all points in $\mathbb{R}^2$ of the form $u_0 + rv$ where $||v|| = 1$. A parameterized curve which traverses this set once in the counter-clockwise direction is given by $\gamma(t) = u_0 + (r \cos t, r \sin t)$ for $0 \le t \le 2\pi$.

Length of a Path

Definition 11.1.6. The *length* of a path $\gamma : [a, b] \to \mathbb{R}^d$ is the number $\ell(\gamma)$ defined by

$$\ell(\gamma) = \int_a^b ||\gamma'(t)|| \, dt.$$

Note that the integral in this definition exists because $\gamma'(t)$ is bounded and is continuous except at finitely many points on $[a, b]$. It follows that $||\gamma'(t)||$ has these same properties and is, therefore, integrable.

Example 11.1.7. Find the length of the path in $\mathbb{R}^2$ given by $\gamma(t) = (2t^3, 3t^2)$ for $t \in [0, 1]$.

Solution: Since $\gamma'(t) = (6t^2, 6t)$ and $||\gamma'(t)|| = \sqrt{36t^4 + 36t^2} = 6t\sqrt{t^2 + 1}$, we conclude

$$\ell(\gamma) = 6 \int_0^1 t\sqrt{t^2 + 1} \, dt = 3 \int_1^2 \sqrt{u} \, du = 2u^{3/2}\big|_1^2 = 2(2\sqrt{2} - 1)$$

where we have made the substitution $u = t^2 + 1$, $du = 2t\,dt$.

Differential 1-Forms

Recall from Chapter 9 that if F is a differentiable function from an open subset of $\mathbb{R}^p$ to $\mathbb{R}^q$, then its differential $dF(x)$ at a point x is a linear transformation from $\mathbb{R}^p$ to $\mathbb{R}^q$ and, as such, may be represented by a $q \times p$ matrix (the matrix of partial derivatives of the coordinate functions). In particular, an $\mathbb{R}$-valued function f on an open subset of $\mathbb{R}^d$ has differential $df(x)$ at a point x in its domain which is a linear function from $\mathbb{R}^d$ to $\mathbb{R}$ – represented by a $1 \times d$ matrix (such a thing is just a d-vector, but we wish to think of it as a linear transformation from $\mathbb{R}^d$ to $\mathbb{R}$). A notation for df that was introduced in Section 9.4 is

$$df = \frac{\partial f}{\partial x_1} dx_1 + \frac{\partial f}{\partial x_2} dx_2 + \cdots + \frac{\partial f}{\partial x_d} dx_d.$$

Here, dx_j may be thought of as the differential of the jth coordinate function x_j. When represented as a $1 \times d$ matrix, dx_j is 1 in the jth entry and 0 in all other entries. This determines the linear transformation which sends a vector of dimension d to its jth component. Similarly, df may be represented as the $1 \times d$ matrix which is $\dfrac{\partial f}{\partial x_j}$ in the jth entry for each j.

A differential 1-form ϕ on a set E in $\mathbb{R}^d$ is just a continuous function which assigns to each point x of E a linear function $\phi(x) : \mathbb{R}^d \to \mathbb{R}$. Since the dx_j form a basis for the vector space of such functions, each differential form ϕ may be written in the form

$$\phi = \phi_1 dx_1 + \phi_2 dx_2 + \cdots + \phi_d dx_d,$$

where the functions ϕ_j are continuous $\mathbb{R}$-valued functions on E. For example, if E is a subset of $\mathbb{R}^2$, then a 1-form on E is an expression of the form $f\,dx + g\,dy$, where f and g are continuous functions on U.

Note that the gradient df of a differentiable function is a special kind of differential 1-form – one in which the functions ϕ_j are the partial derivatives $\dfrac{\partial f}{\partial x_j}$ of f.

Integration Along a Path

Let $\gamma : [a, b] \to \mathbb{R}^d$ be a path. Since γ is a function from a subset of $\mathbb{R}$ to a subset of $\mathbb{R}^d$, its differential $d\gamma$ is a function which assigns to a point $t \in [a, b]$ a linear function from $\mathbb{R}$ to $\mathbb{R}^d$ – that is, a $d \times 1$ matrix. In fact this matrix is just the vector $\gamma'(t)$ regarded as a column vector. In this chapter, we will write

$$d\gamma(t) = \gamma'(t)\,dt$$

where dt is to be thought of as the differential of the identity function (the function that sends t to itself) and $\gamma'(t)$ is to be thought of as a $d \times 1$ matrix.

This formalism may seem unnecessarily complicated, but it is very useful in the coming discussions of transformation laws for paths, differential forms, and integrals under changes of variables.

If ϕ is a differential 1-form defined on a set containing the trace of γ, then $\phi(\gamma(t))$ acts on $d\gamma(t)$ through matrix multiplication to produce a real number $\phi(t)d\gamma(t)$. The resulting real valued function is a bounded function on $[a, b]$ which is continuous except at finitely many points. We may integrate this function.

The resulting integral has a very important property – it is independent of the parameterization of the path. We will prove this in the next section. An integral of this type is called a *line integral* or *path integral*. The formal definition is as follows:

Definition 11.1.8. If $\phi = \phi_1 dx_1 + \phi_2 dx_2 + \cdots + \phi_d dx_d$ is a continuous 1-form defined on a set A in $\mathbb{R}^d$ and $\gamma = (\gamma_1, \gamma_2, \cdots, \gamma_d)$ is a path in A with parameter interval $[a, b]$, then the integral of ϕ over γ is defined to be

$$\int_\gamma \phi = \int_a^b \phi(\gamma(t))d\gamma(t) = \int_a^b \phi(\gamma(t))\gamma'(t)\, dt = \int_a^b \sum_{j=1}^d \phi_j(\gamma(t))\gamma_j'(t)\, dt.$$

A useful device for remembering and applying this definition is suggested by the use of differentials in the change of variable formalism for the Riemann integral: The jth coordinate x_j of a point on the curve γ and its formal differential dx_j are given by

$$\begin{aligned} x_j &= \gamma_j(t), \\ dx_j &= \gamma_j'(t)\, dt. \end{aligned} \tag{11.1.1}$$

The formula for the integral given in Definition 11.1.8 is

$$\int_\gamma (\phi_1(x)\, dx_1 + \cdots + \phi_d(x)\, dx_d) = \int_a^b (\phi_1(\gamma(t))\gamma_1'(t) + \cdots \phi_d(\gamma(t))\gamma_d'(t))\, dt.$$

We may think of the right side of this equation as being obtained from the left side by making the substitutions (11.1.1).

Example 11.1.9. Find $\int_\gamma (y\, dx + x\, dy)$ and $\int_\lambda (y\, dx + x\, dy)$, if

$$\begin{aligned} \gamma(t) &= (1 + 2t, 1 + 3t) \quad \text{for} \quad 0 \le t \le 1 \\ \lambda(t) &= (1 + 2t^2, 1 + 3t^2) \quad \text{for} \quad 0 \le t \le 1. \end{aligned}$$

Solution: On the curve γ, we have $x = 1 + 2t$, $dx = 2\, dt$, $y = 1 + 3t$, and $dy = 3\, dt$. Thus,

$$\int_\gamma (y\, dx + x\, dy) = \int_0^1 ((1 + 3t)2 + (1 + 2t)3)\, dt = \int_0^1 (5 + 12t)\, dt = 11.$$

On the curve λ, we have $x = 1 + 2t^2$, $dx = 4t\,dt$, $y = 1 + 3t^2$, and $dy = 6t\,dt$. Thus,

$$\int_\lambda (y\,dx + x\,dy) = \int_0^1 ((1 + 3t^2)4t + (1 + 2t^2))6t)\,dt \int_0^1 (24t^3 + 10t)\,dt = 11.$$

Thus, the two integrals yield the same result. Note that γ and λ are just different parameterizations of the straight line joining $(1, 1)$ to $(3, 4)$.

The Fundamental Theorem of Calculus

A simple consequence of the Fundamental Theorem of Calculus in the context of differential forms and paths is the following.

Theorem 11.1.10. *Let γ be a path in $\mathbb{R}^d$ with parameter interval $[a, b]$ and let f be a differentiable function on a set containing $\gamma(I)$. Then*

$$\int_\gamma df = f(\gamma(b)) - f(\gamma(a)).$$

Proof. First assume the path γ is a smooth curve. If $\gamma = (\gamma_1, \cdots, \gamma_d)$, then

$$\int_\gamma df = \int_a^b df(\gamma(t))d\gamma(t) = \int_a^b d(f \circ \gamma)(t)$$

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

by the chain rule and the Fundamental Theorem of Calculus.

The proof in the case where γ is not smooth is left to the exercises (Exercise 11.1.9). $\quad\square$

Simple Paths and Smooth Simple Paths

A path γ with parameter interval I is said to be *simple* if it satisfies the following two conditions:

1. if s and t are distinct points of I which are not both endpoints of I, then $\gamma(s) \neq \gamma(t)$;

2. γ' not only exists, but is non-vanishing at all but finitely many points of the interior of I.

The first condition says that γ is one-to-one, except that we allow the endpoints of I to be sent to the same point in the case of a closed path. The second condition says that $\gamma : I \to \mathbb{R}^d$ has a well defined tangent line at all but finitely many interior points of I. Intuitively, a simple path is one which does not cross itself or retrace portions of itself and has a tangent line at all but finitely many points. A *simple closed path* is a closed path which is simple – for example, a circle traversed once.

A smooth simple curve γ is a simple curve which is smooth and which has $\gamma'(t) \neq 0$ at each interior point of I. This means that the tangent vector $T(t) = \gamma'(t)/||\gamma'(t)||$ is defined at each such point. Note that, since a smooth curve may not be simple (it may cross itself), there may be more than one tangent vector at a given point of the trace $\gamma(I)$ of I; however, these will correspond to different parameter values. A smooth simple curve has a well defined tangent vector at each point of $\gamma(I)$ except possibly at $\gamma(a)$ or $\gamma(b)$.

Exercise Set 11.1

1. Find a smooth curve in $\mathbb{R}^2$ which traces the straight line from $(1, 2)$ to $(3, 0)$.

2. Graph the spiral curve in $\mathbb{R}^2$ defined by $\gamma(t) = (t \cos t, t \sin t)$, $0 \leq t \leq 4\pi$, and then find its length.

3. Find the length of the curve $\gamma(t) = (t, t^{3/2})$, $0 \leq t \leq 1$.

4. If ϕ is the 1-form $\phi(x, y) = x\, dx + y\, dy$ and γ is the curve $\gamma(t) = (t^2, t^3)$, $0 \leq t \leq 1$, then find $\int_\gamma \phi$.

5. If ϕ is the 1-form $\phi(x, y) = xy\, dx - x^2\, dy$ and γ is the curve $\gamma(t) = (\cos t, \sin t)$, $0 \leq t \leq \pi/2$, then find $\int_\gamma \phi$.

6. In $\mathbb{R}^3$ let ϕ be the 1-form $\phi(x, y, z) = x^2\, dx + y^2\, dy + dz$. Find $\int_\gamma \phi$ if $\gamma(t) = (\cos(2\pi t), \sin(2\pi t), t - t^2)$, $0 \leq t \leq 1$.

7. In $\mathbb{R}^3$, let ϕ be the 1-form $\phi = \sin z\, dx + \cos z\, dy + y^2\, dz$ and let γ be the smooth curve $\gamma(t) = (\cos t, \sin t, t)$, $0 \leq t \leq 2\pi$. Describe $\gamma(I)$ and find $\int_\gamma \phi$.

8. If $\gamma : [0, 1] \to \mathbb{R}^d$ is a path, set $^-\gamma(t) = \gamma(1 - t)$ – that is, $^-\gamma$ is γ traversed backwards. Show that

$$\int_{-\gamma} \phi = -\int_\gamma \phi$$

for any 1-form ϕ defined on the trace of γ.

9. Theorem 11.1.10 was proved in the case where γ is smooth. Use this to prove that the theorem also holds in the case where γ is not smooth – that is, the case where it is made up of several smooth curves joined together.

10. Prove that if γ is a closed path and f is a smooth function defined on an open set containing the trace of γ, then $\int_\gamma df = 0$.

11.2 Change of Variables

There are some arbitrary choices made in our descriptions of paths and 1-forms in the previous section. A path γ comes with a choice of parameterization. Does the integral along this path depend on the choice of parameterization or is it only the trace $\gamma(I)$ that is important and we are free to parameterize it any way we wish? Also, the descriptions of paths and 1-forms in $\mathbb{R}^d$ involve a choice of a coordinate system for $\mathbb{R}^d$. If this is changed, the expression for a path will change in accordance with this change of coordinates, How should the expression for a 1-form change in order that the integral remains the same? These are crucial questions. Their resolution is the key ingredient in the proofs of the main theorems of this chapter.

Parameter Independence

The equality of the integrals in the Example 11.1.9 is not an accident. The integral of a 1-form over a path is essentially independent of how the path is parameterized. The precise statement of this independence is the next theorem. First a definition:

Definition 11.2.1. Suppose γ and λ are smooth curves in $\mathbb{R}^d$ with parameter intervals $[a,b]$ and $[c,d]$, respectively. Let α be a continuous function from $[c,d]$ onto $[a,b]$ which is smooth with non-vanishing derivative on (c,d). If $\lambda = \gamma \circ \alpha$, then we will say that α determines a *smooth parameter change* from γ to λ. If, in addition, $\alpha' > 0$ on (c,d), then we will say that α is *orientation preserving*. On the other hand, if $\alpha' < 0$ on (c,d) we will say that α is *orientation reversing*.

Note that since $\alpha'(t) \neq 0$ for all $t \in (c,d)$, then α' is either everywhere positive or everywhere negative on (c,d) by the Intermediate Value Theorem (Theorem 3.2.3) applied to α'. This, in turn, implies that α is either increasing on $[c,d]$ or decreasing on $[c,d]$ (recall that such a function is said to be *strictly monotone* on $[c,d]$).

Intuitively, a smooth parameter change replaces γ with a new path λ which traverses the same trace, moving consistently in either the same direction or the reverse direction of the original path γ.

Theorem 11.2.2. *Suppose γ and λ are smooth curves in $\mathbb{R}^d$ with parameter intervals $[a,b]$ and $[c,d]$, respectively, and suppose α determines a smooth parameter change from γ to λ. Then*

$$\int_\lambda \phi = \pm \int_\gamma \phi$$

for each 1-form $\phi = \phi_1 dx_1 + \cdots + \phi_d x_d$ defined on a set containing the common trace of γ and λ. The factor of ± 1 that appears on the right in this equaility will be positive if α is orientation preserving and negative if it is orientation reversing.

Proof. This is a simple application of the chain rule and the change of variable formula for integrals on the line. Suppose first that α is orientation preserving. By the chain rule, we have $d\lambda(t) = d\gamma(\alpha(t))d\alpha(t) = \gamma'(\alpha(t))\alpha'(t)\,dt$, and so

$$
\int_\lambda \phi = \int_c^d \phi(\lambda(t))d\lambda(t) = \int_c^d \phi(\gamma(\alpha(t)))\gamma'(\alpha(t))\alpha'(t)dt
$$
$$
= \int_a^b \phi(\gamma(s))\gamma'(s)ds = \int_\gamma \phi,
$$

where we have made the substitution $s = \alpha(t)$, $ds = d\alpha(t) = \alpha'(t)\,dt$.

If α is orientation reversing, then a and b will be reversed in the fourth integral above and to undo this reversal introduces a factor of -1. ☐

Definition 11.2.3. If γ and λ are two paths which have the same trace and if

$$
\int_\gamma \phi = \int_\lambda \phi
$$

for every 1-form ϕ defined on the common trace of γ and λ, then we will say that γ and λ are *equivalent* paths.

Theorem 11.2.2 says that if there is an orientation preserving smooth parameter change from γ to λ, then the paths γ and λ are equivalent.

Remark 11.2.4. If γ and λ are two path and if there is a smooth parameter change α from γ to λ, then α has an inverse function $\alpha^{-1} : [a,b] \to [c,d]$ and it is a smooth parameter change from λ to γ (see Exercise 11.2.6).

Example 11.2.5. In Example 11.1.9 the two curves γ and λ are shown to be equivalent. Is there a smooth orientation preserving parameter change from γ to λ? Is there a smooth orientation preserving parameter change from λ to γ?

Solution: The function $\alpha(t) = t^2$ is increasing and has the property that $\lambda = \gamma \circ \alpha$. Also, it has positive, continuous derivative on $(0, 1)$ and so it is smooth parameter change. Note that α' is bounded on $(0, 1)$ in this case. The smooth parameter change going the other direction (from λ to γ) is $\alpha^{-1}(s) = \sqrt{s}$. This function does not have a bounded derivative on $(0, 1)$ but that is not a requirement for a smooth parameter change.

Example 11.2.6. Consider the paths in $\mathbb{R}^2$ given by $\gamma(t) = (\cos t, \sin t)$ and $\lambda(t) = (\cos t, -\sin t)$ for $0 \le t \le 2\pi$. Is there a smooth parameter change from γ to λ? Are γ and λ equivalent?

Solution: These paths each traverse the circle of radius 1 centered at $(0, 0)$ in $\mathbb{R}^2$ once, but in opposite directions. The function $\alpha(t) = 2\pi - t$ is a smooth parameter change from γ to λ, since $\cos(2\pi - t) = \cos t$ and $\sin(2\pi - t) = -\sin t$. However, α is orientation reversing, and so Theorem 11.2.2 tells us that γ and λ are not equivalent. We can confirm this by direct calculation if we choose the

1-form $\phi(x,y) = -y dx + x dy$. On γ we have $x = \cos t$, $dx = -\sin t \, dt$, $y = \sin t$, $dy = \cos t \, dt$. Thus,

$$\int_\gamma \phi = \int_0^{2\pi} (\sin^2 t + \cos^2 t) \, dt = \int_0^{2\pi} 1 \, dt = 2\pi,$$

On λ, x and dx are the same, but $y = -\sin t$, $dy = -\cos t \, dt$. Thus,

$$\int_\lambda \phi = \int_0^{2\pi} (-\sin^2 t - \cos^2 t) \, dt = \int_0^{2\pi} (-1) \, dt = -2\pi.$$

Theorem 11.2.2 leads to a strategy which, for many paths γ and λ with the same trace, yields a proof that they are equivalent paths. Suppose that the parameter intervals for the two paths can each be partitioned into n subintervals in such a way that for $j = 1, \cdots, n$, γ on its jth subinterval and λ on its jth subinterval are related by a smooth orientation preserving parameter change α_j, as in Theorem 11.2.2. If this can be done, then it clearly follows that $\int_\gamma \phi = \int_\lambda \phi$ for any 1-form ϕ which is defined on a set containing the common trace of γ and λ. Hence, the two paths are equivalent in this situation.

The question of parameter independence is particularly simple for smooth, simple curves.

Theorem 11.2.7. *If γ and λ are two smooth, simple non-closed curves in $\mathbb{R}^d$ which begin at the same point, end at the same point, and have the same trace, then there is an orientation preserving smooth parameter change from γ to λ. Hence, γ and λ are equivalent in this case.*

Proof. Let the parameter intervals for γ and λ be $[a,b]$ and $[c,d]$. For each $t \in [c,d]$ there is an $s \in [a,b]$ such that $\lambda(t) = \gamma(s)$. This is because both γ and λ have the same trace. Furthermore, since γ is one-to-one, there is only one such s for each t. We denote this s by $\alpha(t)$. This defines a function $\alpha : [c,d] \to [a,b]$ such that $\lambda(t) = \gamma(\alpha(t))$. We will show that α has a continuous positive derivative on (c,d). This follows from the Implicit Function Theorem, as we shall show below.

We set $F(s,t) = \lambda(t) - \gamma(s)$. Then F is a smooth function from $[a,b] \times [c,d]$ to $\mathbb{R}^d$. If t_0 is a point of (c,d) we wish to show that $\alpha'(t)$ exists in a neighborhood of t_0 and is continuous at t_0.

Let $s_0 = \alpha(t_0)$. Since $\gamma'(s) \neq 0$ for each s, it follows that $\frac{\partial f_j}{\partial s}(s_0, t_0) \neq 0$ for at least one of the coordinate functions f_j of F. By the Implicit Function Theorem, there is a smooth function β defined in a neighborhood of t_0 such that $\beta(t_0) = s_0$ and $f_j(s,t) = 0$ for (s,t) in a neighborhood of (s_0, t_0) if and only if $s = \beta(t)$. Since we have $F(\alpha(t), t) = 0$ for all $t \in [c,d]$ by the choice of α, we also have $f_j(\alpha(t), t) = 0$. It follows that $\beta(t) = \alpha(t)$ in some neighborhood of t_0. Thus, α is smooth in a neighborhood of t_0.

The fact that $\alpha'(t)$ is non-vanishing follows from the chain rule. Since $\lambda(t) = \gamma(\alpha(t))$, the chain rule implies that

$$\lambda'(t) = \gamma'(\alpha(t))\alpha'(t).$$

Here, $\alpha'(t)$ is a scalar multiplying the vector $\gamma'(\alpha(dt))$. If there were a point t where $\alpha'(t) = 0$, then we would have $\lambda'(t) = 0$ also, and this not possible, since λ' is non-vanishing. Thus, α is a smooth parameter change from γ to λ

Since α' is non-vanishing on (a, b), it is either strictly positive or strictly negative by the Intermediate Value Theorem. Hence, α is either increasing or decreasing on $[c, d]$. It must be increasing, since it takes c to a and d to b. Thus, α is orientation preserving. $\square$

What if we do not assume the two curves in the preceding theorem are non-closed? What if they are closed? Does the theorem still hold? If not, is there a way to modify the theorem so that it does hold in this case. These questions are dealt with in the exercises.

Arc Length Parameterization

Suppose γ is a smooth curve with parameter interval $[a, b]$. We define a change of variables from t to a new variable s by setting

$$s(t) = \int_a^t ||\gamma'(u)|| \, du$$

for each $t \in [a, b]$. That is, $s(t)$ is the length of that part of the curve γ for which the parameter u lies in the interval $[a, t]$. Furthermore, by the Fundamental Theorem of Calculus,

$$ds = ||\gamma'(t)||dt.$$

Since $||\gamma'(t)||$ is a positive continuous function of t and is the derivative of s, it follows that s, as a function of t, is a continuous, increasing function from $[a, b]$ to $[0, \ell(\gamma)]$ which is smooth on (a, b). Hence, its inverse function defines t as a continuous, increasing function of s for $s \in [0, \ell(\gamma)]$ with image $[a, b]$. Furthermore, it is smooth on $(0, \ell)$. This defines a smooth parameter change from γ to the curve $\lambda(s) = \gamma(t(s))$.

The length of a curve remains the same after a smooth parameter change (Exercise 11.2.7). Thus, given $s \in [0, \ell(\gamma)]$, the length of that part of λ for which the parameter lies between 0 and s is the same as the length of that part of γ for which the parameter lies between a and t. This is exactly

$$\int_0^t ||\gamma'u)|| \, du = s. \tag{11.2.1}$$

That is, s is the length of that part of λ for which the parameter lies in $[0, s]$. A smooth curve or a path with this property is said to be *parameterized by arc length*.

Since each path is made up of a number of smooth curves joined together, we have proved:

Theorem 11.2.8. *Each path in $\mathbb{R}^d$ may be reparameterized so as to be a path parameterized by arc length.*

Equation 11.2.1, when applied to the curve λ parameterized by arc length, yields

$$s = \int_0^s ||\lambda'(t)|| \, dt.$$

where λ_s denotes λ restricted to $[0, s]$. On differentiating and using the Fundamental Theorem of Calculus, we conclude that that $||\lambda'(s)|| = 1$ for each s. That is, $\lambda'(s)$ is a unit vector. This unit vector is often denoted by T and is called the *unit tangent vector* to γ. A simple calculation shows that, in terms of γ, $T = \gamma'/||\gamma'||$.

Classical Form for Path Integrals

Let $\phi = f_1 \, dx_1 + \cdots + f_p \, dx_p$ be a 1-form on a subset A of $\mathbb{R}^p$ and γ a simple path in A with trace C. If $F = (f_1, \cdots, f_p)$ is the vector valued function determined by the components of ϕ, then the path integral of ϕ over γ is classically written as

$$\int_\gamma \phi \, d\gamma = \int_C F \cdot T \, ds, \qquad (11.2.2)$$

where $T = \gamma'(t)/||\gamma'(t)||$ is the unit tangent vector to γ and $ds = ||\gamma'(t)|| \, dt$ is the differential of arc length along γ, as above. Here the integral on the right is just another way of denoting

$$\int_a^b F(\gamma(t)) \cdot T(\gamma(t)) ||\gamma'(t)|| \, dt = \int_a^b F(t) \cdot \gamma'(t) \, dt.$$

Integrals of this type arise in may contexts in Physics. For example, If F is a force field acting on an object, then the above path integral represents the work done by the force field as the object moves along the path γ.

The classical notation represents the integral of a 1-form along a path as the integral of an ordinary function $F \cdot T$ with respect to arc length along the path. Such an integral can be defined for any continuous function along the path. This leads to the definition of an integral along a path for ordinary continuous functions as opposed to 1-forms:

Definition 11.2.9. If f is a continuous real valued function, defined on the trace C of a path γ with parameter interval $[a, b]$, then we define

$$\int_C f \, ds = \int_a^b f(\gamma(t)) ||\gamma'(t)|| \, dt.$$

This is called the integral of f over C with respect to arc length.

Change of Variables for 1-Forms

A smooth parameter change is one kind of change of variables. It is a change in the independent variable of a path. It is equally important to understand how

to deal with a change of variables in the dependent variable space. By this, we mean a smooth one-to-one function from one open set in $\mathbb{R}^d$ to another which has a non-singular differential.

More generally, let U be an open subset of $\mathbb{R}^p$ and $H : U \to R^q$ any smooth function. The function H could be a smooth change of variables or possibly a function which parameterizes a piece of a p-surface in $\mathbb{R}^q$. It is important to understand how functions, paths, and differential forms are transformed by H. Such an understanding will allow us to solve problems concerning functions, paths, and forms on complicated sets by reducing the problem to an analogous problem on a simpler set such as a square or a cube. We have already done this type of thing. This is exactly what is involved when we parameterize a path in order to express the integral of a 1-form over the path as an integral of a function over an interval I on the line.

With $U \subset \mathbb{R}^p$ and $H : U \to \mathbb{R}^q$ as above, if $\gamma : I \to U$ is a path in U, then $H \circ \gamma : I \to \mathbb{R}^q$ is a path in $\mathbb{R}^q$. On the other hand, if f is a function defined on a set containing $H(U)$, then $f \circ H$ is a function defined on U. We will often call this function $H^*(f)$. Note that, while $\gamma \to H \circ \gamma$ takes paths in $\mathbb{R}^p$ to paths in R^q, the operation $f \to H^*(f)$ goes the other way – that is, it takes functions on a subset of $\mathbb{R}^q$ to functions on a subset of $\mathbb{R}^p$. Note that there is the following relationship between the two operations: If we evaluate the function $H^*(f)$ along the curve γ, the result is the real valued function $H^*(f) \circ \gamma$ on I. On the other hand

$$H^*(f) \circ \gamma = (f \circ H) \circ \gamma = f \circ (H \circ \gamma),$$

which is the result of evaluating f along the curve $H \circ \gamma$.

How do 1-forms transform under a function H, as above? This is best understood by seeing how a 1-form of the form df should transform.

Let f be a smooth function defined on U and let df be its differential (considered as a vector valued function on U). Under H, f transforms to $H^*(f) = f \circ H$. The differential of this function, by the chain rule, is the vector-matrix product $(df \circ H)dH$. This suggests that we should regard $(df \circ H)dH$ as the appropriate transform of df under the function H. This, in turn, suggests that the function H should transform every differential 1-form on U in the same manner. That is H should take a differential 1-form ϕ to $(\phi \circ H)dH$, where $\phi \circ H$ is a vector valued function and dH a matrix valued function on V, and $(\phi \circ H)dH$ is the vector-matrix product of $\phi \circ H$ with dH. This leads to the following definition.

Definition 11.2.10. If U is an open subset of $\mathbb{R}^p$ and $H : U \to \mathbb{R}^q$ a smooth function, then for each function (0-form) f on $H(U)$ and each 1-form ϕ on $H(U)$, we define a function $H^*(f)$ and 1-form $H^*(\phi)$ on U by

$$H^*(f) = f \circ H \quad \text{and} \quad H^*(\phi) = (\phi \circ H)dH$$

Example 11.2.11. Let $H : U \to \mathbb{R}^3$ be a smooth function, as above, with U an open subset of $\mathbb{R}^2$. If we regard the coordinates (x, y, z) of points in the image of H to be functions on U of the variables (u, v) through the equation

$(x, y, z) = H(u, v)$ and if $\phi(x, y, z) = f(x, y, z)\, dx + g(x, y, z)\, dy + h(x, y, z)\, dz$ is a 1-form on $H(U)$, then write out $H^*(\phi)$ in the (u, v) coordinates.

Solution: In vector notation, the new 1-form is

$$
H^*(\phi) = (\phi \circ H)dH = \begin{pmatrix} f \circ H, & g \circ H, & h \circ H \end{pmatrix} \begin{pmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \\ \dfrac{\partial z}{\partial u} & \dfrac{\partial z}{\partial v} \end{pmatrix},
$$

$$
= \left(f \circ H \frac{\partial x}{\partial u} + g \circ H \frac{\partial y}{\partial u} + h \circ H \frac{\partial z}{\partial u}, f \circ H \frac{\partial x}{\partial v} + g \circ H \frac{\partial y}{\partial v} + h \circ H \frac{\partial z}{\partial v} \right),
$$

where all functions are evaluated at (u, v). If we write this in terms of the basis vectors du and dv, it becomes

$$
\left(f \circ H \frac{\partial x}{\partial u} + g \circ H \frac{\partial y}{\partial u} + h \circ H \frac{\partial z}{\partial u} \right) du
$$

$$
+ \left(f \circ H \frac{\partial x}{\partial v} + g \circ H \frac{\partial y}{\partial v} + h \circ H \frac{\partial z}{\partial v} \right) dv.
$$

Remark 11.2.12. An easy way to remember how a 1-form $\phi = f\,dx + g\,dy + h\,dz$ in $\mathbb{R}^3$ transforms under a function $H : U \to \mathbb{R}^3$ with $U \subset \mathbb{R}^2$ is to think of making the replacements

$$
(x, y, z) = H(u, v)
$$

for (x, y, z) in $f(x, y, z)$, $g(x, y, z)$ and $h(x, y, z)$ and the replacements

$$
dx = \frac{\partial x}{\partial u} du + \frac{\partial x}{\partial v} dv
$$

$$
dy = \frac{\partial y}{\partial u} du + \frac{\partial y}{\partial v} dv
$$

$$
dz = \frac{\partial z}{\partial u} du + \frac{\partial z}{\partial v} dv
$$

This leads to the same expression for the transformed 1-form as is obtained in the preceding example. The same formalism works for transforming 1-forms on $\mathbb{R}^q$ to 1-forms on $\mathbb{R}^p$ under any smooth function H from an open subset of $\mathbb{R}^p$ to $\mathbb{R}^q$. Note that, when $p = q = 1$, this formalism is just the procedure for replacing $f(x)\, dx$ by the appropriate expression when doing a substitution $x = H(u)$ in an integral on the line.

Example 11.2.13. Consider the function $H(r, \theta) = (r \cos \theta, r \sin \theta)$ for $r > 0$ and $-\pi < \theta < \pi$. This is the change of variables $x = r \cos \theta, y = r \sin \theta$ between rectangular and polar coordinates. For the 1-form $\phi(x, y) = x\, dx + y\, dy$, what is $H^*(\phi)$?

Solution: We make the replacements

$$
x = r \cos \theta, \quad y = r \sin \theta,
$$

$$
dx = \cos \theta\, dr - r \sin \theta\, d\theta,
$$

$$
dy = \sin \theta\, dr + r \cos \theta\, d\theta.
$$

Then $\phi(x, y) = x\, dx + y\, dy$ is transformed to

$$H^*(\phi) = r\cos^2\theta\, dr - r^2\sin\theta\cos\theta\, d\theta + r\sin^2\theta\, dr + r^2\sin\theta\cos\theta\, d\theta$$

$$= r(\cos^2\theta + \sin^2\theta)\, dr = r\, dr$$

Change of Variables in Path Integrals

The transformation law for 1-forms under a smooth transformation is the correct one if we want path integrals to be preserved.

Theorem 11.2.14. *If U is an open subset of $\mathbb{R}^p$, $H : U \to \mathbb{R}^q$ is a smooth transformation, ϕ is a 1-form on $H(U)$, and $\gamma : I \to U$ is a path in U, then*

$$\int_\gamma H^*(\phi) = \int_{H\circ\gamma} \phi.$$

Proof. Ultimately, this reduces to the chain rule and the definition of the integral of a 1-form over a path. That is, if $I = [a, b]$,

$$\int_\gamma H^*(\phi) = \int_\gamma \phi \circ H\, dH = \int_a^b \phi(H(\gamma(t))dH(\gamma(t))\gamma'(t)\, dt$$

$$= \int_a^b \phi(H \circ \gamma(t))(H \circ \gamma)'(t)\, dt = \int_{H\circ\gamma} \phi.$$

$\square$

Example 11.2.15. Find $\int_\lambda (xdx + ydy)$ for the path $\lambda(t) = (\cos t, \sin t)$ with $-\pi \le t \le \pi$, by first changing to polar coordinates (as in Example 11.2.13) and then integrating the resulting 1-form over the path given by

$$(r, \theta) = \gamma(t) = (1, t) \quad \text{for} \quad -\pi \le t \le \pi.$$

Solution: By Example 11.2.13, the form $x\, dx + y\, dy$ transforms to $r\, dr$ under the transform H to polar coordinates. Also, $\lambda = H \circ \gamma$. Hence, by the previous theorem.

$$\int_\lambda (x\, dx + y\, dy) = \int_\gamma r\, dr = 0,$$

since $r = 1$ and $dr = 0$ on γ.

Exercise Set 11.2

1. Are the curves $\gamma(t) = (t^3, t^2)$, $0 \le t \le 1$ and $\lambda(t) = (\sin^3 t, 1 - \cos^2 t)$, $0 \le t \le \pi/2$, equivalent curves? Justify your answer.

2. Are the curves $\gamma(t) = (\cos t, \sin t)$, $0 \le t \le 2\pi$, and $\lambda(t) = (\cos t, \sin t)$, $0 \le t \le 4\pi$, equivalent curves? Justify your answer.

3. Are the curves $\gamma(s) = (s, \sqrt{1 - s^2})$, $-1 \le s \le 1$, and $\lambda(t) = (\cos t, \sin t)$, $0 \le t \le \pi$, equivalent curves? Does the answer change if the parameter interval for λ is changed to $-\pi \le t \le 0$? Justify your answer.

4. If γ is a path with parameter interval $[a, b]$, show how to define a smooth parameter change from γ to an equivalent path λ which has $[0, 1]$ as parameter interval. Hint: you simply need to find a smooth increasing function $\alpha : [0, 1] \to [a, b]$ and then set $\lambda = \gamma \circ \alpha$. There are many such functions, but there is one which is particularly simple.

5. Give an example to show that the conclusion of Theorem 11.2.7 does not hold if we do not assume the paths are non-closed. Tell how to restate the theorem so that it does hold for closed curves as well as non-closed curves.

6. Prove that if γ and λ are smooth paths, $\alpha : [c, d] \to [a, b]$ is a smooth parameter change from γ to λ, then α has a smooth inverse function $\alpha^{-1} : [a, b] \to [c, d]$ which is a smooth parameter change from λ to γ. Furthermore, α is orientation preserving if and only if α^{-1} is orientation preserving.

7. Show that a smooth parameter change does not change the length of a smooth curve.

8. If $\gamma(t) = (\cos 2\pi t, \sin 2\pi t)$ for $0 \le t \le 1$, describe a curve equivalent to γ which is parameterized by arc length.

9. Express the differential form $y\,dx - x\,dy$ in polar coordinates (see Example 11.2.13).

10. Calculate $\int_\lambda (y\,dx - x\,dy)$, where $\lambda(t) = (\cos t, \sin t)$ for $-\pi \le t \le \pi$, by first expressing this integral in polar coordinates, as in Example 11.2.15.

11. Give a different solution to the problem in Example 11.2.13 by noticing that $x\,dx + y\,dy$ is df for the function $f(x, y) = (x^2 + y^2)/2$. What does f transform into under the change to polar coordinates? How does this lead immediately to the solution in Example 11.2.15.

12. What does the differential form $x\,dx + y\,dy + z\,dz$ on $\mathbb{R}^3$ transform to under the change to spherical coordinates?

13. What does the differential form $y\,dx - x\,dy + dz$ transform to under the change of coordinates $x = u + 2v$, $y = 3u - v$, $z = u + v + w$.

14. If $H : (-\pi, \pi) \times (-\pi, \pi) \to \mathbb{R}^3$ is defined by

$$H(u, v) = (\cos u \cos v, \sin u \cos v, \sin v),$$

what does the differential form $x\,dx + y\,dy + z\,dz$ transform to under H?

11.3   Differential Forms of Higher Order

The statements and proofs of our main integration theorems (Green's Theorem, Gauss's Theorem, and Stokes' Theorem) all involve the algebra of differential forms. We have already seen how differential 1-forms enter into the definition of path integrals. Second order differential forms are involved in the definitions of surface integrals and third order forms are related to integrals over solid regions in $\mathbb{R}^3$. In this section we introduce higher order differential forms, the operations we shall perform on them, and the transformation rules that govern them.

2-Forms

If coordinate functions $x_1, \cdots x_d$ are chosen for $\mathbb{R}^d$, then we begin by constructing a vector space over $\mathbb{R}$ that has certain symbols $dx_i \wedge dx_j$ as basis elements. Here, we declare that

$$dx_j \wedge dx_i = -dx_i \wedge dx_j, \quad \text{and} \quad dx_i \wedge dx_i = 0, \tag{11.3.1}$$

for all i and j. Our basis vectors will then be the expressions $dx_i \wedge dx_j$ for which $i < j$. Whenever a symbol $x_j \wedge x_i$ with $j > i$ occurs in a calculation, we simply replace it by $-dx_i \wedge dx_j$. Of course, if $dx_i \wedge dx_i$ occurs, it is replaced by 0.

Given a subset E of $\mathbb{R}^d$, a *differential 2-form* is a continuous function on E with values in the vector space described above. Thus, a differential 2-form, when written out in terms of the basis described above yields an expression of the form

$$\phi(x) = \sum_{i<j}^{d} f_{ij}(x) dx_i \wedge dx_j,$$

where each f_{ij} is a continuous function on E.

We may construct 2-forms from 1-forms in two ways.

First, there is a product operation, called *exterior* or *wedge* product, which assigns to each pair ϕ, ψ of 1-forms a 2-form $\phi \wedge \psi$. If $\phi = \sum_{i=1}^{d} f_i \, dx_i$ and $\psi = \sum_{i=1}^{d} g_i \, dx_i$, then

$$\phi \wedge \psi = \sum_{i,j=1}^{d} f_i g_j \, dx_i \wedge dx_j = \sum_{i<j} (f_i g_j - f_j g_i) \, dx_i \wedge dx_j.$$

Here, in going from the first to the second sum, we have used the relations (11.3.1) to express the sum in terms of the basis vectors $dx_i \wedge dx_j$ for $i < j$.

Second, we may take the differential of a 1-form: if $\phi = \sum_{j=1}^{d} f_j \, dx_j$ is a 1-form defined on an open set $U \subset \mathbb{R}^d$, then we define a 2-form $d\phi$, called the *differential* of ϕ, by

$$d\phi = \sum_{j=1}^{d} df_j \wedge dx_j = \sum_{i,j=1}^{d} \frac{\partial f_j}{\partial x_i} dx_i \wedge dx_j$$

$$= \sum_{i<j} \left(\frac{\partial f_j}{\partial x_i} - \frac{\partial f_i}{\partial x_j} \right) dx_i \wedge dx_j.$$

Note that we previously defined the differential of a function f (a 0-form) to be a certain 1-form df. Now we have defined the differential of a 1-form to be a certain 2-form. In general, the differential of a p-form will be a p+1-form.

The following theorem follows directly from the definitions. The proof is left to the exercises.

Theorem 11.3.1. *Let ϕ, θ and ψ be differentiable 1-forms and f a differentiable function defined on an open set U. Then*

(a) $\phi \wedge \psi = -\psi \wedge \phi$;

(b) $\phi \wedge (\theta + \psi) = \phi \wedge \theta + \phi \wedge \psi$;

(c) $f(\phi \wedge \psi) = (f\phi) \wedge \psi = \phi \wedge (f\psi)$;

(d) $d(\phi + \psi) = d\phi + d\psi$;

(e) $d(f\phi) = df \wedge \phi + f d\phi$.

On $\mathbb{R}^2$, 2-forms are particularly simple. If x and y are the coordinate functions, then $dx \wedge dy$ is the only basis vector for 2-forms and so all 2-forms can be expressed as $f\, dx \wedge dy$ for some continuous function f.

Example 11.3.2. Given 1-forms $\phi = f\, dx + g\, dy$, and $\psi = h\, dx + k\, dy$ find

$$(a) \quad \phi \wedge \psi; \quad \text{and} \quad (b) \quad d\phi.$$

Solution:

$$(a) \quad \phi \wedge \psi = fh\, dx \wedge dx + fk\, dx \wedge dy + gh\, dy \wedge dx + gk\, dy \wedge dy$$
$$= (fk - gh) dx \wedge dy;$$

$$(b) \qquad d\phi = df \wedge dx + dg \wedge dy$$
$$= \left(\frac{\partial f}{\partial x}\, dx + \frac{\partial f}{\partial y}\, dy \right) \wedge dx + \left(\frac{\partial g}{\partial x}\, dx + \frac{\partial g}{\partial y}\, dy \right) \wedge dy$$
$$= \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx \wedge dy.$$

On $\mathbb{R}^3$, the basis vectors $dx \wedge dy$, $dy \wedge dz$, and $dx \wedge dz$ are independent and generate a 3-dimensional vector space. Thus, a typical differential 2-form on an open subset U of $\mathbb{R}^3$ has the form

$$f_1\, dy \wedge dz + f_2\, dx \wedge dz + f_3\, dx \wedge dy$$

where f_1, f_2, and f_3 are continuous functions on U.

In some contexts, a function $F = (f_1, f_2, f_3)$ from $U \subset \mathbb{R}^3$ to $\mathbb{R}^3$ is called a *vector field* on U. Thus, a 2-form $\phi = f_1\, dy \wedge dz + f_2\, dx \wedge dz + f_3\, dx \wedge dy$ in $\mathbb{R}^3$ determines a vector field $F = (f_1, f_2, f_3)$. We will call this the component vector field of ϕ. Of course, a 1-form $g_1\, dx + g_2\, dy + g_3\, dz$ in $\mathbb{R}^3$ also determines a component vector field $G = (g_1, g_2, g_3)$.

Example 11.3.3. If $\phi = f_1 dx + f_2 dy + f_3 dz$ and $\psi = g_1 dx + g_2 dy + g_3 dz$ are 1-forms on $U \subset \mathbb{R}^3$, then find

$$(a) \quad \phi \wedge \psi; \quad \text{and} \quad (b) \quad d\phi.$$

Solution: Using (11.3.1) and collecting terms involving $dx \wedge dy$, $dy \wedge dz$, and $dx \wedge dz$ we obtain:

$$\phi \wedge \psi = (f_2 g_3 - f_3 g_2) \, dy \wedge dz + (f_3 g_1 - f_1 g_3) \, dz \wedge dx + (f_1 g_2 - f_2 g_1) \, dx \wedge dy,$$

and

$$d\phi = df_1 \wedge dx + df_2 \wedge dy + df_3 \wedge dz$$
$$= \left(\frac{\partial f_3}{\partial y} - \frac{\partial f_2}{\partial z} \right) dy \wedge dz + \left(\frac{\partial f_1}{\partial z} - \frac{\partial f_3}{\partial x} \right) dz \wedge dx + \left(\frac{\partial f_2}{\partial x} - \frac{\partial f_1}{\partial y} \right) dx \wedge dy.$$

Remark 11.3.4. Note that, if F is the component vector field of the 1-form ϕ and G is the component vector field of the 1-form ψ, then the formulas of the preceding example say that

1. the component vector field of $\phi \wedge \psi$ is $F \times G$, and

2. the component vector field of $d\phi$ is $\operatorname{curl} F$,

in terms of the classical cross product "$\times$" and "curl" operations.

3-Forms

A differential 3-form on an open subset U of $\mathbb{R}^d$ is a sum of expressions of the form

$$f \, dx_i \wedge dx_j \wedge dx_k,$$

where f is a continuous function on U. As in (11.3.1), interchanging any two adjacent terms dx_i, dx_j, dx_k in this expression changes the sign of the expression. If two of i, j, k are equal, then the expression is understood to be equal to 0. It follows from this that every 3-form on U may be expressed as a sum of forms as above with $i < j < k$.

In the obvious way, the wedge product of three 1-forms is a 3-form and the wedge product of a 1-form with a 2-form is a 3-form. We define the exterior differential $d\phi$ of a 2-form

$$\phi = \sum_{i<j} f_{ij} dx_i \wedge dx_j$$

to be the 3-form

$$d\phi = \sum_{i<j} df_{ij} \wedge dx_i \wedge dx_j = \sum_{i<j} \sum_k \frac{\partial f_{ij}}{\partial x_k} \, dx_k \wedge dx_i \wedge dx_j$$

Example 11.3.5. If $\phi = f_1\, dy \wedge dz + f_2\, dz \wedge dx + f_3\, dx \wedge dy$ is a 2-form on an open subset of $\mathbb{R}^3$, find $d\phi$.

Solution: By definition,

$$d\phi = df_1 \wedge dy \wedge dz + df_2 \wedge dz \wedge dx + df_3 \wedge dx \wedge dy.$$

Since $df_1 = \dfrac{\partial f_1}{\partial x}\, dx + \dfrac{\partial f_1}{\partial y}\, dy + \dfrac{\partial f_1}{\partial z}\, dz$ and since $dy \wedge dy \wedge dz = 0$ and $dz \wedge dy \wedge dz = 0$, the only non-zero term in $df_1 \wedge dy \wedge dz$ will be the term involving $dx \wedge dy \wedge dz$. Similar statements hold for $df_2 \wedge dx \wedge dz$ and $df_3 \wedge dx \wedge dy$. It follows that

$$d\phi = \frac{\partial f_1}{\partial x}\, dx \wedge dy \wedge dz + \frac{\partial f_2}{\partial y}\, dy \wedge dz \wedge dx + \frac{\partial f_3}{\partial z}\, dz \wedge dx \wedge dy$$

$$= \left(\frac{\partial f_1}{\partial x} + \frac{\partial f_2}{\partial y} + \frac{\partial f_3}{\partial z} \right) dx \wedge dy \wedge dz = \mathrm{div} F\, dx \wedge dy \wedge dz,$$

if F is the component vector field of ϕ. Here, div is the classical divergence operation on vector fields in $\mathbb{R}^3$.

Theorem 11.3.6. *Let f be a function which is $\mathcal{C}^2$ on an open set $U \subset \mathbb{R}^p$ and ϕ a 1-form with coefficients which are $\mathcal{C}^2$ on U. Then*

(a) $d(df) = 0$; and

(b) $d(d\phi) = 0$.

Proof. We will prove part (a) and leave part (b) for the exercises.

We have

$$df = \sum_{j=1}^{p} \frac{\partial f}{\partial x_j}\, dx_j$$

and

$$d(df) = \sum_{j=1}^{p} \sum_{k=1}^{p} \frac{\partial^2 f}{\partial x_k \partial x_j}\, dx_k \wedge dx_j. \qquad (11.3.2)$$

Now for each pair of indices (j, k) that occurs in this sum, the opposite pair (k, j) also occurs. Furthermore

$$\frac{\partial^2 f}{\partial x_j \partial x_k} = \frac{\partial^2 f}{\partial x_k \partial x_j}, \quad \text{and} \quad dx_j \wedge dx_k = -dx_k \wedge dx_j$$

by Theorem 9.1.6 (since f is $\mathcal{C}^2$) and by Theorem 11.3.1 (a). It follows that the jk^{th} term and the kj^{th} term in (11.3.2) cancel each other and the sum is 0. This proves part (a) of the theorem. $\qquad \square$

Although we won't do it here, one can, of course, define differential forms of any non-negative degree p and define the differential of such a form. What the above theorem says for 1-forms and 2-forms is true for any $\mathcal{C}^2$ p-form ϕ – that is $d^2\phi = d(d\phi) = 0$. A differential form ϕ is said to be *closed* if $d\phi = 0$ and *exact*

if $\phi = d\psi$ for some form ψ. Thus, exact $\mathcal{C}^2$ forms are always closed. How about the converse? It turns out the converse is not true in general, but it is true if the domain U of the form satisfies certain topological conditions. In particular, it is true if U is convex. We explicitly state this here for 1-forms. The proof is left to the exercises.

Theorem 11.3.7. *If U is a convex set and ϕ is a closed 1-form on U ($d\phi = 0$), then ϕ is exact ($\phi = df$ for some $\mathcal{C}^2$ f on U).*

Remark 11.3.8. We may summarize the relationship between the exterior differential operation d and its classical counterparts for vector functions on $\mathbb{R}^3$ as follows: If f is a function, ϕ a 1-form with component vector field F and ω a 2-form with component vector field G, all defined on an open subset of $\mathbb{R}^3$, then

1. the component vector field of df is $\operatorname{grad} f$

2. the component vector field of $d\phi$ is $\operatorname{curl} F$, and

3. the coefficient function of $d\omega$ is $\operatorname{div} G$.

Transformation Laws for 2-Forms and 3-Forms

If $H : U \to \mathbb{R}^m$ is a function defined on an open subset U of $\mathbb{R}^d$, then we would like 2-forms and 3-forms to transform under H in a way which is consistent with our earlier rules for transforming functions and 1-forms, and in a way that preserves wedge products. This leads to,

Definition 11.3.9. With H as above, if $\phi = \sum_{i<j} f_{ij} dx_i \wedge dx_j$ is a 2-form and $\omega = \sum_{i<j<k} f_{ijk} dx_i \wedge dx_j \wedge d_k$ is a 3-form defined on a set containing $H(U)$, then we define $H^*(\phi)$ and $H^*(\omega)$ as follows:

$$H^*(\phi) = \sum_{i<j} H^*(f_{ij}) H^*(dx_i) \wedge H^*(dx_j),$$

$$H^*(\omega) = \sum_{i<j<k} H^*(f_{ijk}) H^*(dx_i) \wedge H^*(dx_j) \wedge H^*(dx_k).$$

Of course, we may define p-forms on U for any non-negative integer p, not just for $p = 0, 1, 2, 3$. The appropriate transformation law for such a p-form under $H : U \to \mathbb{R}^m$ is the obvious extension of the laws already described, as above, for $p \leq 3$. Note that if p is greater than the dimension of the underlying space, then 0 is the only p-form.

In the following theorem, parts (a) and (b) follows immediately from the definitions and part (c) has a simple proof which is left to the exercises.

Theorem 11.3.10. *Let ϕ and ψ be two differential forms on an open set V in $\mathbb{R}^q$ and let f be a function on V. If U is an open subset of $\mathbb{R}^p$ and $H : U \to B$ a smooth function, then*

(a) $H^*(f\phi) = H^*(f)H^*(\phi);$

(b) $H^*(\phi \wedge \psi) = H^*(\phi) \wedge H^*(\psi);$ *and*

(c) $H^*(d\phi) = dH^*(\phi).$

Example 11.3.11. If U is an open subset of $\mathbb{R}^2$, $H : U \to \mathbb{R}^2$ is a smooth transformation, and $\phi(x, y) = f(x, y)\, dx \wedge dy$ is a 2-form defined on $H(U)$, then find an explicit expression for $H^*(\phi)$.

   Solution: As in Remark 11.2.12, we may think of H as a change of variables

$$x = h_1(u, v), \quad y = h_2(u, v)$$

and simply replace x and y by $h_1(u, v)$ and $h_2(u, v)$ in $f(x, y)$ and in dx and dy. This leads to

$$dx = \frac{\partial h_1}{\partial u}\, du + \frac{\partial h_1}{\partial v}\, dv, \quad dy = \frac{\partial h_2}{\partial u}\, du + \frac{\partial h_2}{\partial v}\, dv, \quad \text{and}$$

$$dx \wedge dy = \left(\frac{\partial h_1}{\partial u} \frac{\partial h_2}{\partial v} - \frac{\partial h_1}{\partial v} \frac{\partial h_2}{\partial u} \right) du \wedge dv$$
$$= \det(dH)\, du \wedge dv.$$

More precisely , $dx \wedge dy$, when expressed in the u, v coordinates, becomes

$$H^*(dx \wedge dy) = \det(dH)\, du \wedge dv.$$

Since $H * (f) = f \circ H$, we conclude that

$$H^*(\phi) = H^*(f)H^*(dx \wedge dy) = f \circ H \det(dH) du \wedge dv.$$

Example 11.3.12. If U is an open subset of $\mathbb{R}^2$, $H : U \to \mathbb{R}^3$ is a smooth transformation, and

$$\phi(x, y, z) = f_1(x, y, z)\, dy \wedge dz + f_2(x, y, z)\, dz \wedge dx + f_3(x, y, z)\, dx \wedge dy$$

is a 2-form defined on $H(U)$, then find an explicit expression for $H^*(\phi)$.

   Solution: If $H(u, v) = (h_1(u, v), h_2(u, v), h_3(u, v))$, then we may think of H as defining a change of variables

$$x = h_1(u, v), \quad y = h_2(u, v), \quad z = h_3(u, v).$$

Then

$$dx = \frac{\partial h_1}{\partial u}\, du + \frac{\partial h_1}{\partial v}\, dv, \quad dy = \frac{\partial h_2}{\partial u}\, du + \frac{\partial h_2}{\partial v}\, dv, \quad dz = \frac{\partial h_3}{\partial u}\, du + \frac{\partial h_3}{\partial v}\, dv.$$

If we set

$$\frac{\partial(h_i, h_j)}{\partial(u, v)} = \det \begin{pmatrix} \dfrac{\partial h_i}{\partial u} & \dfrac{\partial h_j}{\partial u} \\ \dfrac{\partial h_i}{\partial v} & \dfrac{\partial h_j}{\partial v} \end{pmatrix},$$

we conclude that

$$H^*(\phi) = \left[(f_1 \circ H) \frac{\partial(h_2, h_3)}{\partial(u, v)} + (f_2 \circ H) \frac{\partial(h_3, h_1)}{\partial(u, v)} + (f_3 \circ H) \frac{\partial(h_1, h_2)}{\partial(u, v)} \right] du \wedge dv.$$

This can also be written as

$$H^*(\phi) = (F \circ H) \cdot \left[\frac{\partial H}{\partial u} \times \frac{\partial H}{\partial v} \right] du \wedge dv, \qquad (11.3.3)$$

if F denotes the vector function $F = (f_1, f_2, f_3)$, and $\dfrac{\partial H}{\partial u}$ and $\dfrac{\partial H}{\partial v}$ denote the vector functions obtained by taking the partial derivatives of the component functions of H.

Example 11.3.13. If $\phi = x\, dy \wedge dz + y\, dz \wedge dx + z\, dx \wedge dy$ and $H : \mathbb{R}^2 \to \mathbb{R}^3$ is the transformation $H(u, v) = (u, v, u^2 + v^2)$, then find $H^*(\phi)$.

Solution: We express the transformation H as a change of variables

$$x = u, \quad y = v, \quad z = u^2 + v^2.$$

Then $dx = du$, $dy = dv$, and $dz = 2u\,du + 2v\,dv$. Thus

$$dy \wedge dz = -2u\, du \wedge dv, \quad dz \wedge dx = -2v\, du \wedge dv, \quad dx \wedge dy = du \wedge dv.$$

Thus, $H^*(\phi) = (u, v, u^2 + v^2) \cdot (-2u, -2v, 1) du \wedge dv = -(u^2 + v^2)\, du \wedge dv$.

Finally, there is a composition law for transformations of forms:

Theorem 11.3.14. *if $H_1 : U \to V$ and $H_2 : V \to W$ are smooth functions between open sets, and if ϕ is any differential form defined on W, then*

$$(H_2 \circ H_1)^*(\phi) = H_1^* \circ H_2^*(\phi).$$

Proof. It follows from the previous theorem that it is enough to check this in the case when ϕ is a function f or the differential of a function (such as the differential dx_j of one of the coordinate functions x_j). In the case of a function f, we have

$$(H_2 \circ H_1)^*(f) = f \circ (H_2 \circ H_1) = (f \circ H_2) \circ H_1 = H_1^*(f \circ H_2)$$
$$= H_1^*(H_2^*(f)) = H_1^* \circ H_2^*(f).$$

In the case when ϕ is the differential df of a function, we have

$$(H_2 \circ H_1)^*(df) = d(f \circ (H_2 \circ H_1)) = d((f \circ H_2) \circ H_1))$$
$$= H_1^*(d(f \circ H_2)) = H_1^*(H_2^*(df)) = (H_1^* \circ H_2^*)(df).$$

This completes the proof. $\square$

Exercise Set 11.3

1. If $\phi = x^2\, dx + xy\, dy$ and $\psi = y\, dx + x^3\, dy$, then find $d\phi$ and $\phi \wedge \psi$.

2. $\phi = \cos z\, dx + \sin z\, dy + xy\, dz$ and $\psi = z\, dx + x\, dy + y\, dz$, then find $d\phi$ and $\phi \wedge \psi$.

3. If $\phi = yz\, dx + xz\, dy + xy\, dz$ and $\omega = z\, dx \wedge dy + x\, dy \wedge dz + dx \wedge dz$, then find $d\phi$ and $\phi \wedge \omega$.

4. Prove Theorem 11.3.1 parts (a), (b), and (c).

5. Prove Theorem 11.3.1 parts (d) and (e).

6. Prove Part (b) of Theorem 11.3.6.

7. Prove Theorem 11.3.7. Hint: fix a point $a \in U$ and then define $f(x)$ to be the integral of ϕ along the line $[a, x]$; show $\phi = df$ by using the condition $d\phi = 0$ and integration by parts.

8. Show that 11.3.7 does not hold if we don't put some restriction on the domain U. In fact show that if

    $$\phi = \frac{-y}{x^2 + y^2}dx + \frac{x}{x^2 + y^2}dy \quad \text{on} \quad U = \{(x,y) \in \mathbb{R}^2 : 1/2 < \|(x,y)\| < 2\},$$

    then ϕ is closed but not exact on U. Hint: Use the result of Exercise 11.1.10.

9. Prove Theorem 11.3.10. Part (a) in the case where ϕ is a 2-form or a 3-form in $\mathbb{R}^3$.

10. Prove Theorem 11.3.10. Part (b) in the case where ϕ is a 1-form and ψ is a 2-form in $\mathbb{R}^3$.

11. Prove Theorem 11.3.10. Part (c) in the case where ϕ is a 2-form in $\mathbb{R}^3$.

12. Prove that the vector $\dfrac{\partial H}{\partial u} \times \dfrac{\partial H}{\partial v}$ that appears in (11.3.3) is perpendicular to the surface $H(U)$ at each point $H(u, v)$ of this surface.

11.4   Green's Theorem

Green's Theorem relates certain double integrals over a region in the plane to path integrals over the boundary of the region. It has a wide variety of applications and it generalizes nicely to higher dimensions. In this section, we prove Green's Theorem for fairly general regions. We begin with the case where the region is a rectangle.

Green's Theorem on a Rectangle

In its simplest form, Green's Theorem follows from two applications of the Fundamental Theorem of Calculus – one in the x-direction and one in the y-direction.

Theorem 11.4.1. *Let* $\phi = f\,dx + g\,dy$ *be a 1-form on the rectangle* $R = [a, b] \times [c, d]$ *and suppose* $d\phi$ *exists and is continuous and bounded on the interior of* R. *Then*

$$\int_R \left(\frac{\partial g}{\partial x}(x, y) - \frac{\partial f}{\partial y}(x, y) \right) dV(x, y) = \int_{\partial R} \phi,$$

where ∂R *is a path which traces the boundary of* R *once in the counter-clockwise direction.*

Proof. We begin by breaking up the double integral on the left and expressing each of the resulting terms as an iterated integral using Fubini's Theorem:

$$\int_R \left(\frac{\partial g}{\partial x}(x, y) - \frac{\partial f}{\partial y}(x, y) \right) dV(x, y)$$
$$= \int_c^d \int_a^b \frac{\partial g}{\partial x}(x, y)\,dxdy - \int_a^b \int_c^d \frac{\partial f}{\partial y}(x, y)\,dydx.$$

The hypotheses on ϕ ensure that the Fundamental Theorem of Calculus applies to the inner integral in each of the latter iterated integrals. This yields

$$\int_c^d (g(b, y) - g(a, y))\,dy - \int_a^b (f(x, d) - f(x, c))\,dx$$
$$= \int_c^d g(b, y)\,dy + \int_b^a f(x, d)\,dx + \int_d^c g(a, y)\,dy + \int_a^b f(x, c)\,dx.$$
$$= \int_{\partial R} \phi,$$

where ∂R is the path obtained by joining together the four straight line paths along the edges of R in such a way that the resulting path traverses the boundary of R once in the counter-clockwise direction. $\square$

In the following example, we use Green's Theorem to avoid parameterizing four different sides of a rectangle R in order to compute a line integral around ∂R.

Example 11.4.2. Find $\int_{\partial R}(y^2\,dx + y \ln x\,dy)$ if $R = [1, 2] \times [0, 1]$
 Solution: By Theorem 11.4.1

$$\int_{\partial R} (y^2\,dx + y \ln x\,dy) = \int_R (y/x - 2y)\,dV(x, y).$$

By Fubini's Theorem, the latter integral is equal to the iterated integral

$$\int_0^1 \int_1^2 (y/x - 2y)\,dxdy = \int_0^1 (y \ln 2 - 2y)\,dy = \frac{\ln 2}{2} - 1.$$

Integration of 2-Forms in the Plane

It will be helpful to interpret the integral of a function over a region in the plane as an integral of a certain 2-form.

If A is a compact Jordan region in the plane, every 2-form on A is of the form $f\,dx \wedge dy$ where f is a continuous function on A. We define the integral of such a 2-form over A to be

$$\int_A f(x,y)\,dx \wedge dy = \int_A f(x,y)dV(x,y). \qquad (11.4.1)$$

That is, it is the ordinary Riemann integral in two variables of the function f over the set A. The advantages of using the 2-form notation in the integral will become apparent below.

In Example 11.3.2 we showed that if $\phi = f\,dx + g\,dy$ is a differentiable 1-form, then

$$d\phi = \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx \wedge dy.$$

This, together with the above 2-form notation for integrals in $\mathbb{R}^2$ allows us to rewrite the left side of the equality in Theorem 11.4.1 as $\int_R d\phi$. Then Green's Theorem on a rectangle becomes

Theorem 11.4.3. *If ϕ is a 1-form defined on a bounded rectangle R and if $d\phi$ is continuous and bounded on the interior of A, then*

$$\int_R d\phi = \int_{\partial R} \phi.$$

Proof. By (11.4.1) and the previous theorem, we have

$$\int_R d\phi = \int \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dx \wedge dy = \int_R \left(\frac{\partial g}{\partial x} - \frac{\partial f}{\partial y} \right) dV = \int_{\partial R} \phi$$

$\square$

Change of Variables for Integrals of 2-forms

Using the 2-form notation for integrals in $\mathbb{R}^2$ also turns the change of variables formula for such integrals into a natural formula involving the transformation law for 2-forms, as discussed in the previous section.

Theorem 11.4.4. *Let $H = (h_1, h_2)$ be a continuous transformation from the open Jordan region U in $\mathbb{R}^2$ to another Jordan region in $\mathbb{R}^2$ and suppose H is one-to-one and smooth with non-singular differential on U. If ϕ is a 2-form on $H(U)$ with ϕ bounded on $H(U)$ and $H^*(\phi)$ bounded on U, then*

$$\int_{H(U)} \phi = \int_U H^*(\phi)$$

provided $\det(dH) > 0$ everywhere on U. If $\det(dH) < 0$ on U, equality holds if the right side of the equation is replaced by its negative.

Proof. Let $\phi(x, y) = f(x, y)dx \wedge dy$. By Example 11.3.11,

$$H^*(\phi) = f \circ H \, \det(dH) \, du \wedge dv. \tag{11.4.2}$$

Recall that the differential dH of the transformation H is the linear transformation with matrix

$$\begin{pmatrix} \dfrac{\partial h_1}{\partial u} & \dfrac{\partial h_1}{\partial v} \\ \dfrac{\partial h_2}{\partial u} & \dfrac{\partial h_2}{\partial v} \end{pmatrix}.$$

The hypotheses of the theorem ensure that the change of variables formula (Theorem 10.5.14)applies. If the determinant $\det(dH)$ is everywhere non-negative on U, then it implies

$$\int_{H(U)} f(x, y) \, dx \wedge dy = \int_U f \circ H(u, v) |\det(dH)(u, v)| \, du \wedge dv,$$
$$= \int_U f \circ H(u, v) \det(dH)(u, v) \, du \wedge dv, = \int_U H^*(\phi). \tag{11.4.3}$$

That is,

$$\int_{H(U)} \phi = \int_U H^*(\phi).$$

If $\det(dH)$ is everywhere non-positive, then $|\det(dH)| = -\det(dH)$ and the right side of the above equation is replaced by its negative. $\qquad\square$

2-Cells

In order to extend Green's Theorem to a much larger class of integrals, we need to change our point of view regarding integrals of 2-forms. We have discussed in previous sections the integration of 1-forms over paths. A path is not a set, but rather a function from an interval into R^d, although we sometimes ignore the distinction between the path and the set which is its trace in R^d. There is a similar and highly useful formulation for integration of 2-forms. We define the integral of a 2-form over an object which is not a set, but rather a 2-dimensional analogue of a smooth path. A 2-cell, as defined below, is such an object.

In what follows, I^2 will denote the square $[0, 1] \times [0, 1]$ in $\mathbb{R}^2$. The boundary path ∂I^2 is the path consisting of the straight line paths along the edges of I^2 joined together so as to traverse the topological boundary of I^2 in the counterclockwise direction.

We will say the function $E : I^2 \to \mathbb{R}^d$ is smooth on I^2 if each of its first order partial derivatives exists and is continuous on I^2. It is clear what this means on the interior of I^2. On each edge and corner of I^2 one or both of the partial derivatives must be interpreted as a one-sided derivative. Thus, at each point of I^2 we require that the appropriate one or two sided derivative exists and we require that the resulting functions on I^2 are continuous. With this understanding, we make the following definition.
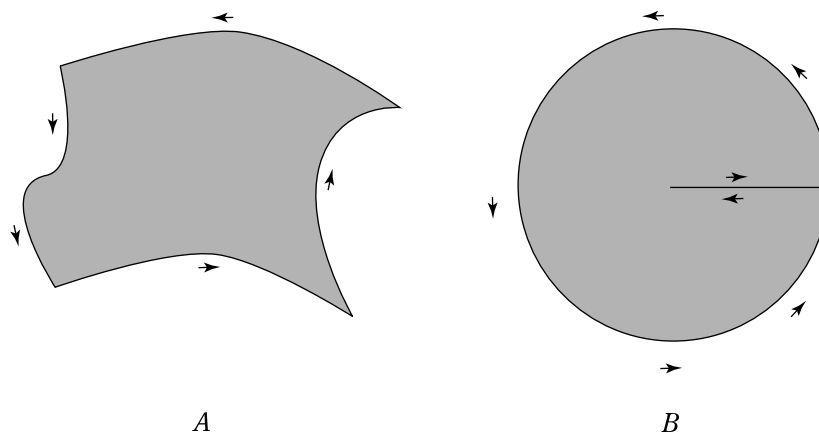
Figure 11.3: Simple, Positively Oriented Cells in $\mathbb{R}^2$

Definition 11.4.5. A 2-cell in $\mathbb{R}^d$ is a smooth function E from I^2 into $\mathbb{R}^d$.

We will say that a 2-cell E is *simple* if, on the interior of I^2, E is one to one and $\det(dE)$ is non-vanishing. If, in addition, $\det(dE) > 0$ on the interior of I^2, we will say that E is *positively oriented*. We will say that E is *negatively oriented* if $\det(dE) < 0$ on the interior of I^2.

In this section we will only be concerned with 2-cells in $\mathbb{R}^2$. In the next section, 2-cells in higher dimensional spaces will become important.

Note also that the conditions on a cell E ensure that the restriction of E to each of the four edges of ∂I^2 is a smooth curve and, hence, that $\partial E = E \circ \partial I^2$ is piecewise smooth – that is, it is a path.

The image $E(I^2)$ of a 2-cell E is called the *trace* of E. As was the case with curves and paths, a 2-cell consists of not only the set $E(I^2)$, but also a parameterization E of that set, with the parameters being the coordinates of points in I^2.

In general, a path may cross itself, retrace portions of itself, or even be constant over portions of its parameter interval. However, a simple path can do none of these things. A simple path is one-to-one and has non-vanishing derivative on the interior of its parameter interval. Similarly, a simple 2-cell is one-to-one with non-singular differential on the interior of I^2.

Note that if E is a 2-cell, then ∂E is a path, not a set, and so it is not the same thing as the topological boundary of $E(I^2)$ even though we use the same notation to denote it. Which is meant should be obvious from the context. Sometimes the trace of ∂E is the same as the topological boundary of the trace of E, but not always (see Figure 11.3).

Orientation for Paths

A path which traverses the boundary of a set such as a square or circle in the counterclockwise direction has a property which can be generalized in a useful way.

An ordered basis in $\mathbb{R}^2$ is a linearly independent ordered pair $\{u, v\}$ of vectors in $\mathbb{R}^2$. An ordered basis is said to be *positively oriented* if the angle θ between the two vectors, measured from u to v satisfies $0 < \theta < \pi$ – that is, if $\sin \theta > 0$. Think of this as meaning that v points to the left of u. This happens if and only if the determinant of the matrix with u as first column and v as second column is positive (Exercise 11.4.8).

At each smooth point of ∂I^2 (at points which are not corners), the tangent vector T to the path is defined. Furthermore, if v is any vector for which (T, v) is a positively oriented ordered basis, then tv belongs to I^2 for all sufficiently small positive t. In other words, the set I^2 lies on the left as we traverse ∂I^2. It turns out that this property is preserved by a positively oriented 2-cell, due to the fact that dE takes a positively oriented basis to a positively oriented basis. That is, at each smooth point a of the path ∂E, if the tangent vector T to the path at a and a vector v form a positively oriented pair (T, v), then each sufficiently small positive multiple of v lies in $E(I^2)$ (we won't prove this here). Intuitively, this means that as we traverse ∂E, the set $E(I^2)$ lies on the left (see Figure 11.3). If the cell is negatively oriented, the set $E(I^2)$ lies on the right as we traverse the path ∂E – that is, the orientation of the boundary path is reversed by E.

Example 11.4.6. Give an example of a simple, positively oriented 2-cell which has as its trace the unit disc $D = \{(x, y) : x^2 + y^2 \leq 1\}$.

Solution: There are many ways to do this. One way is to use the polar coordinate parameterization:

$$E(r, t) = (r \cos(2\pi t), r \sin(2\pi t)) \quad \text{for} \quad (r, t) \in I^2.$$

This is illustrated in Figure 11.3 B. We have

$$dE = \begin{pmatrix} \cos(2\pi t) & -2\pi r \sin(2\pi t) \\ \sin(2\pi t) & 2\pi r \cos(2\pi t) \end{pmatrix}$$

and this has determinant $2\pi r$, which is positive on the interior of I^2. This parameterization is clearly one-to-one on the interior of I^2. Hence, E is a simple, positively oriented 2-cell.

Note that part of the trace of the boundary ∂E of this cell does not actually lie on the boundary of the trace of E, but in its interior, and this part of the trace of ∂E is traversed twice – once in each direction. Also, over part of its parameter interval, ∂E is constant (the part corresponding to the side $r = 0$ of ∂I^2). Our definition of a simple cell does not rule out this kind of behavior.

Integration Over a Cell

Just as we defined the integral of a 1-form over a path in Section 11.1, we may now define the integral of a 2-form over a 2-cell.

Definition 11.4.7. If E is a 2-cell in $\mathbb{R}^2$ and $\omega = f\, dx \wedge dy$ is a 2-form defined on the trace of E, then we define the integral of ω over E to be

$$\int_E \omega = \int_{I^2} E^*(\omega).$$

Note that the integral on the right in this definition exists. To see this, let $\omega = f\, dx \wedge dy$ and $E(u, v) = (e_1(u, v), e_2(u, v))$, then

$$E^*(\omega) = f \circ E \left(\frac{\partial e_1}{\partial u} \frac{\partial e_2}{\partial v} - \frac{\partial e_1}{\partial v} \frac{\partial e_2}{\partial u} \right) du \wedge dv.$$

By the definition of a 2-cell, the function multiplying $du \wedge dv$ in this expression is continuous on I^2.

Integration Over a Simple Cell

The image of the interior of I^2 under a simple cell E is an open subset of the trace of E by Exercise 9.6.8. It follows that the boundary of the trace of E is contained in the trace of ∂E. A path has zero area (Exercise 11.4.7). This implies that the trace of ∂E has zero area and, hence, that the trace of E and the image under E of the interior of I^2 are Jordan regions which differ by a set of area 0. Furthermore, a simple cell, restricted to the interior of I^2, satisfies the conditions of the change of variables formula given in Theorem 11.4.4. This leads to:

Theorem 11.4.8. *If E is a simple, positively oriented 2-cell with trace $A = E(I^2)$, and if $\omega = f\, dx \wedge dy$ is a 2-form defined on A, then*

$$\int_E \omega = \int_A \omega = \int_A f\, dV(x, y).$$

Proof. This follows immediately from Theorem 11.4.4.   □

Thus, in this case – the case of greatest interest– the integral of the form ω over the cell E is just the integral of a function f over a Jordan region A.

Change of Parameter

Just as with integrals of 1-forms, there is a sense in which the integral of a 2-form over a 2-cell is independent of the parameterization of the 2-cell. If E and F are 2-cells, then we will say that F is related to E by a smooth change of parameter if if there is a smooth one-to-one function H from the interior of I^2 to itself, with non-singular differential, such that $F = E \circ H$ on the interior of I^2. The smooth change of parameter H is said to be positively oriented if $\det(dH) > 0$ on the interior of I^2 and negatively oriented if $\det(dH) < 0$.

Theorem 11.4.9. *If E : and F are 2-cells which are related by a smooth change of parameter H in the above sense, then*

$$\int_F \omega = \int_E \omega.$$

if H is positively oriented and ω is any 2-form defined on $E(I^2)$. This equation holds with the right side replaced by its negative if H is negatively oriented.

Proof. We have

$$\int_F \omega = \int_{I^2} (E \circ H)^*(\omega) = \int_{I^2} H^*(E^*(\omega)) = \int_{I^2} E^*(\omega) = \int_E \omega,$$

by Theorem 11.3.14 and Theorem 11.4.4. $\square$

Green's Theorem on a Cell

We can now extend Green's Theorem to integrals over a two cell.

Theorem 11.4.10. Green's Theorem *If E is a 2-cell in $\mathbb{R}^2$ and ϕ is a smooth 1-form on a neighborhood of the trace of E, then*

$$\int_{\partial E} \phi = \int_E d\phi.$$

Proof. We have

$$\int_{\partial E} \phi = \int_{\partial I^2} E^*(\phi) = \int_{I^2} dE^*(\phi) = \int_{I^2} E^*(d\phi) = \int_E d\phi,$$

by Green's Theorem on a rectangle and Theorem 11.3.10(c). $\square$

Remark 11.4.11. The cell E in the above version of Green's Theorem is not required to be positively oriented or even simple. Thus, the path ∂E may not be positively oriented and the integral of $\phi = f\, dx \wedge dy$ over E may not be the usual 2-dimensional integral of f over the trace of E (it will be its negative if E is negatively oriented). On the other hand, if E is simple, then the integral on the right is the usual 2-dimensional Riemann integral of f over the trace of E by Theorem 11.4.8.

Remark 11.4.12. In actually computing one side or the other of the equality in Green's Theorem, it may be convenient to switch to a different parameterization. For example, if E is simple and orientation preserving, then the integral over E may be replaced by the integral over an equivalent cell F, or by the Riemann integral over the trace A of E, or by an integral over another set which is related to the trace of E through a smooth change of variables as in Theorem 11.4.4. Similarly, we may replace ∂E by an equivalent path γ. The new path γ and the new way of parameterizing A may not be related to each other in the same way that E and ∂E are related.
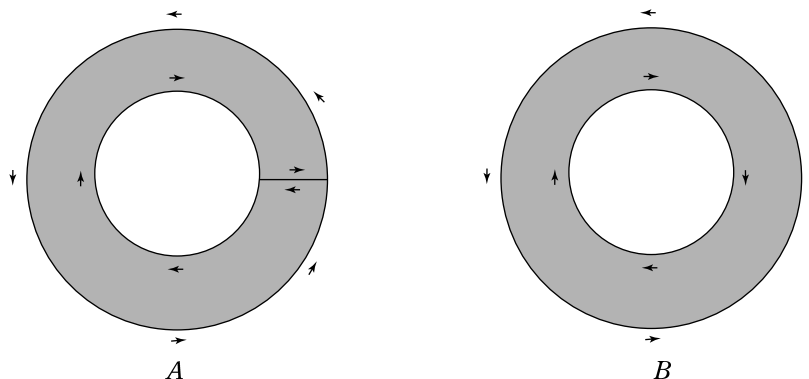
Figure 11.4: The Annulus as a Cell

Example 11.4.13. Let A be the compact set bounded by the ellipse described parametrically by $\gamma(t) = (a \cos t, b \sin t)$, $0 \le t \le 2\pi$. Use Green's Theorem to find the area of A.

Solution: The 2-form $dx \wedge dy$ is $d\phi$, where $\phi = xdy$. Along γ, $x = a \cos t$, and $dy = b \cos t \, dt$. Thus, by Green's Theorem, the area we seek is

$$\int_A dx \wedge dy = \int_\gamma xdy = \int_0^{2\pi} ab \cos^2 t \, dt = \pi ab.$$

Note that the set A is the trace of a 2-cell (Exercise 11.4.3), but we do not need to explicitly find the cell $E : I^2 \to A$ that expresses it as such. If we did find such an E, it is unlikely that the path γ that we used here would be exactly equal to ∂E. However, γ and ∂E will necessarily be equivalent paths, provided E is chosen so that ∂E is a path which traverses ∂A once in the positive direction.

Often the topological boundary of the trace of a cell E is not ∂E and, in fact, it may not even be the trace of a single path. It could be the union of the traces of several paths. Properly interpreted, Green's Theorem still applies, but the integral over the boundary is the sum of integrals over these several paths. The annulus in the following example illustrates this fact, among other things.

Example 11.4.14. For the annulus

$$A = \{(x, y) : 1 \le x^2 + y^2 \le 4\},$$

show that the integral over ∂A of the 1-form

$$\phi = -\frac{y}{x^2 + y^2} dx + \frac{x}{x^2 + y^2} dy$$

is 0 by using Green's Theorem. Then directly calculate the integral of ϕ over the circle $x^2 + y^2 = 4$. Why doesn't Green's Theorem also imply that this integral is 0?

Solution: Figure 11.4A illustrates how to express the annulus as the trace of a cell (finding an explicit parameterization that does this is Exercise 11.4.5). The boundary path of this cell has two overlapping horizontal sections that are oriented in opposite directions. The integrals along these sections will cancel each other, leaving only the integrals around the two circles which comprise the topological boundary of A. One of these is traversed counterclockwise and the other clockwise (Figure 11.4B). To calculate the resulting integral of ϕ along ∂A, we note that

$$d\phi = \frac{y^2 - x^2}{(x^2 + y^2)^2} dy \wedge dx + \frac{y^2 - x^2}{(x^2 + y^2)^2} dx \wedge dy = 0.$$

Thus, by Green's Theorem,

$$\int_{\partial A} \phi = \int_A d\phi = 0.$$

On the other hand, a direct calculation of the integral of ϕ over the outer circle $x^2 + y^2 = 4$ can be done using the parameterization $\gamma(t) = (2\cos t, 2\sin t)$ of this curve on $[0, 2\pi]$. The result is

$$\int_\gamma \left(-\frac{y}{x^2 + y^2} dx + \frac{x}{x^2 + y^2} dy \right) = \int_0^{2\pi} \left(\sin^2 t + \cos^2 t \right) dt = 2\pi.$$

If Green's Theorem applied, the integral would be 0, since $d\phi = 0$. The reason Green's Theorem does not apply in this case is that the circle $x^2 + y^2 = 4$ is not the boundary of a set on which ϕ is a smooth 1-form. The form ϕ has a singularity at $(0, 0)$. On the other hand, the point $(0, 0)$ is not in the annulus A and so it does not cause a problem in applying Green's Theorem to A and ∂A.

Classical Version of Green's Theorem

If $\phi = P\,dx + Q\,dy$ is a differential 2-form and $\gamma = (\gamma_1, \gamma_2) : I \to \mathbb{R}^2$ a path in the domain of ϕ, then

$$\int_\gamma \phi = \int_I \phi \circ \gamma(t) \cdot \gamma'(t)\,dt = \int_I [P(\gamma(t))\gamma_1'(t) + Q(\gamma(t))\gamma_2'(t)]\,dt.$$

Classical notation for this integral is as follows: The differential form ϕ has component vector field $F = (P, Q)$. The tangent vector to the curve γ is $T = \gamma'/\|\gamma'\|$. We write

$$\int_\gamma \phi = \int_\gamma F \cdot T\,ds,$$

where $ds = \|\gamma'(t)\| dt$ is the differential of length along the path γ.

By Remark 11.3.8, if ϕ is a 1-form in $\mathbb{R}^3$ with component vector field F, then $d\phi = \operatorname{curl} F$. The same statement holds in $\mathbb{R}^2$ if the curl of a vector field (P, Q) is understood to be $\partial Q/\partial x - \partial P/\partial y$.

With this notation, the classical version of Green's Theorem is.

Theorem 11.4.15. *Let A be a closed Jordan region in $\mathbb{R}^2$ with topological boundary which is the image of a path ∂A, positively oriented with respect to A. If F is a smooth vector field on $\overline{A}$, T is the vector function which is the tangent vector to ∂A at each point of ∂A, and ds is the differential of arc length along ∂A, then*

$$\int_{\partial A} F \cdot T ds = \iint_A \operatorname{curl} F \, dV.$$

In this section, we have essentially proved this version of the theorem in the case where $\overline{A}$ is the trace of a simple, positively oriented cell. Our proof also yields a proof of the above theorem in the case where $\overline{A}$ can be cut up into finitely many pieces which are traces of simple oriented cells (see Exercise 11.4.13).

Example 11.4.16. If $F(x,y) = (\cos(\ln |x|) + y, xy^2)$, find $\int_C F \cdot T ds$ if $C = \{(x,y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$.

Solution: Green's Theorem tells us that the above integral is the same as

$$\int_{B_1(0,0)} \left(\frac{\partial}{\partial x} xy^2 - \frac{\partial}{\partial y}(\cos(\ln x) + y) \right) dV(x,y) = \int_{B_1(0,0)} (y^2 - 1) \, dV(x,y).$$

We calculate the latter integral using polar coordinates. The result is

$$\int_0^{2\pi} \int_0^1 (r^3 \sin^2 \theta - r) \, dr d\theta = -3\pi/4.$$

Exercise Set 11.4

1. If R is a rectangle of width a and height b, then use Green's Theorem to find $\int_{\partial R} x \, dy$.

2. Use Green's Theorem to find $\int_{\partial I^2}(y^2 x \, dx + x^2 y \, dy)$.

3. Show that $x = a\cos(\pi t), y = b((2s-1)\sin(\pi t)), (s,t) \in I^2$ gives an explicit parameterization as a simple, orientation preserving 2-cell E for the ellipse A of Example 11.4.13. Show that ∂E traverses ∂A once in the positive direction. Explain why this path yields the same integral for a 1-form on A as does the path γ of the example.

4. Using the parameterization E given in the preceding exercise, calculate the area of the ellipse of Example 11.4.13 by directly calculating $\int_E dx \wedge dy$.

5. Find an explicit parameterization for the 2-cell in Figure 11.4A that has the annulus of Example 11.4.14 as its trace.

6. Use Green's Theorem to find $\int_{\partial A}(y^3 \, dx - x^3 \, dy)$ if A is the annulus of the previous exercise.

7. Prove that the image of a path in $\mathbb{R}^2$ is a set of area zero (see Exercises 10.2.6 and 10.2.8).

8. Verify the claim, made in the discussion of orientation for paths, that an ordered pair $\{v, w\}$ of vectors in $\mathbb{R}^2$ forms a positively oriented basis if and only if the matrix with v as first column and w as second column has positive determinant.

9. Prove that a 2×2 matrix takes a positively oriented basis to a positively oriented basis if and only if it has positive determinant.

10. Use Green's Theorem to calculate $\int_{\partial D}(xy\, dx + (x + \ln(2 + y))\, dy)$, where D is the unit disc.

11. If E is a simple positively oriented cell in $\mathbb{R}^2$, with trace A, find a formula which expresses the area of A as an integral around ∂E. Is there more than one way to do this?

12. Use the result of the previous exercise to find the area of the region in $\mathbb{R}^2$ enclosed by the path $x = \cos t$, $y = \sin 2t$, $-\pi/2 \le t \le \pi/2$.

13. Suppose A and ∂A satisfy the hypotheses of Theorem 11.4.15. Suppose that A may be written as the union of finitely many sets, of the form $B_j = \mathrm{Im}(E_j)$ where each E_j is a simple positively oriented cell and any two of the sets B_j intersect only at common boundary points. Explain why it is reasonable to think that the sum of the integrals of a 1-form ϕ along the paths ∂E_j is equal to the integral of ϕ along ∂A.

14. Let U be an open set in $\mathbb{R}^2$ and let a and b be points of U. We say that two paths γ_0 and γ_1 both of which begin at a and end at b, are *homotopic* in U if there is a cell $E : I^2 \to U$ such that $E(s, 0) = a$, $E(s, 1) = b$, $E(0, t) = \gamma_0(t)$, and $E(1, t) = \gamma_1(t)$ for all $s, t \in [0, 1]$. Show that if ϕ is a 1-form with $d\phi = 0$ on U, then

$$\int_{\gamma_0} \phi = \int_{\gamma_1} \phi,$$

whenever γ_0 and γ_1 are homotopic paths joining a to b. Conclude that if any two paths joining the same two points of U are homotopic and $d\phi = 0$, then $\int_\gamma \phi$ depends only on the endpoints of γ and not on the path joining these endpoints.

15. Show that if U is a convex open subset of $\mathbb{R}^2$ and a and b are points of U, then any two smooth paths joining a to b are homotopic in U (see the previous exercise).

11.5 Surface Integrals and Stokes's Theorem

This section is devoted to to the study of integration on 2-dimensional surfaces in $\mathbb{R}^d$ and to generalizations of Green's Theorem to this context.

We begin with a discussion of integration over parameterized surfaces. We discuss the concepts of surface area and orientation for parameterized surfaces and prove that these notions are essentially independent of the choice of parameterization. We then specialize to the case where the parameterized surface is a 2-cell in $\mathbb{R}^d$ and prove Stokes's Theorem. This is a generalization of Green's Theorem to the case where the 2-cell has its trace in $\mathbb{R}^d$ for $d \geq 3$.

In the next section we will generalize Green's Theorem to the case of a 3-cell in $\mathbb{R}^3$ (Gauss's Theorem) or, more generally, a 3-cell in $\mathbb{R}^d$ for $d \geq 3$.

These results do not require many new ideas. Most of what we need has already been encountered in our study of Green's Theorem in the previous section.

Not every geometric object that we might wish to integrate over can be expressed as the trace of a cell. To exploit the full power of these theorems, we will need to consider objects which are constructed by piecing together cells – much as we dealt with piecewise smooth paths in previous sections. This will be done in the final section of this chapter.

Integration Over a Parameterized Surface

A smoothly parameterized surface is the 2-dimensional analogue of a smooth path.

Definition 11.5.1. A parameterized 2-surface in $\mathbb{R}^d$ is a continuous function $H : U \to \mathbb{R}^d$ from an open set $U \subset \mathbb{R}^2$ into $\mathbb{R}^d$. It is a smoothly parameterized surface if H is one-to-one and smooth, with a differential dH which has rank 2 at each point of U. The image of a smoothly parameterized 2-surface is called its *trace*.

The definition given here differs slightly from Definition 9.4.7 in that, here, a specific parameter function H is part of the definition.

The integral of a 2-form over a smoothly parameterized surface follows the pattern of the definitions of integration of 1-forms over paths and of 2-forms over 2-cells in $\mathbb{R}^2$.

Definition 11.5.2. If U is a Jordan region in $\mathbb{R}^2$, $H : U \to \mathbb{R}^d$ $(d \geq 2)$ is a smoothly parameterized surface, in $\mathbb{R}^d$, and ω is a 2-form defined on $A = H(U)$ with $H^*(\omega)$ bounded on U, then we define the integral of ω over H to be

$$\int_H \omega = \int_U H^*(\omega).$$

The condition that $H^*(\omega)$ be bounded in the above definition is needed to ensure that the integral on the right exists (the continuity of ω and smoothness of H ensure that $H^*(\omega)$ is continuous). Note that, if F is the component vector field of ω, then $H^*(\omega)$ is a 2-form on U which is the inner product of $F \circ H$ with a vector consisting of determinants of 2×2 submatrices of dH (see Example 11.3.12). It follows that the condition that $H^*(\omega)$ be bounded in the above definition will be satisfied if dH and ω are both bounded.

Remark 11.5.3. Often the parameter function H will actually be defined and continuous on a compact Jordan region A with U as its interior and the 2-form ω will be continuous on $H(A)$. This guarantees that ω is bounded on $H(U)$. If dH extends to be continuous on the compact set A then we are also guaranteed that dH will be bounded. Note that, in this case, it does not matter whether or not the integral on the right above is taken over $U = A^\circ$ or over A, since ∂A has area 0 (A is a Jordan region).

Example 11.5.4. If $\Delta = \{(x, y) : x > 0, y > 0, x + y < 1\}$ and the parameterized surface $H : \Delta \to \mathbb{R}^3$ is defined by $H(x, y) = (x, y, x - 2y + 5)$, then find $\int_H \omega$ if $\omega = -y\, dy \wedge dz + x\, dz \wedge dx$,

   Solution: In this example, the parameterization H actually expresses the surface as the graph of a function defined on the triangle Δ. That is, H expresses the variables (x, y, z) in terms of (x, y) by $x = x$, $y = y$, $z = x - 2y + 5$. Under H^*, the differentials dx and dy remain unchanged, while $H^*(dz) = dx - 2dy$. Thus,

$$H^*(\omega) = -y\, dy \wedge (dx - 2dy) + x\,(dx - 2dy) \wedge dx = (2x + y)\, dx \wedge dy.$$

Thus,

$$\int_H \omega = \int_U (2x + y)\, dx \wedge dy = \int_0^1 \int_0^{1-y} (2x + y)dxdy = 1/2$$

Parameter Independence

Definition 11.5.5. Let $H : U \to \mathbb{R}^d$ and $J : V \to \mathbb{R}^d$ be smoothly parameterized surfaces. If $P : V \to U$ is a smooth one-to-one function with $\det(dP)$ either strictly positive or strictly negative on V, then we will say that P is a smooth parameter change from H to J provided $H = J \circ P$. If $\det(dP) > 0$ we will say that P is positively oriented, while if $\det dP < 0$ we will say that P is negatively oriented.

   Note that if there is a smooth parameter change from H to J, then $H(U) = J(V)$. That is, H and J have the same trace.

   The theorem on independence of parameterization (Theorem 11.4.9) holds in this more general context. The proof is the same.

Theorem 11.5.6. *Let $H : U \to \mathbb{R}^d$ and $J : V \to \mathbb{R}^d$ be smoothly parameterized surfaces. If there is a smooth parameter change from H to J, then*

$$\int_H \omega = \pm \int_J \omega.$$

if ω is any bounded 2-form on $H(U) = J(V)$. The sign in this identity is positive if P is positively oriented and it is negative if P is negatively oriented.

   This theorem often allows us to simplify an integration problem by choosing a more convenient parameterization than the one given.

Example 11.5.7. Find a smooth parameter change which expresses the integral in Example 11.5.4 as an integral over a square rather than a triangle. Then do the integration.

Solution: We set $P(u,v) = (u, (1-u)v)$. Then P is a one-to-one function from the interior of the square I^2 onto the open triangle Δ and its differential is

$$dP = \begin{pmatrix} 1 & 0 \\ -v & 1-u \end{pmatrix}$$

which has determinant $1 - u$. This is positive on the interior of I^2 and so it determines a positively oriented smooth parameter change. Since $H(x,y) = (x, y, x - 2y + 5)$, the new parameterized surface $J = H \circ P$ is

$$J(u,v) = (u, (1-u)v, u - 2(1-u)v + 5) = (u, v - uv, u - 2v + 2uv + 3).$$

That is, the surface obtained by setting $x = u$, $y = v - uv$, $z = u - 2v + 2uv + 3$. Then $dx = du$, $dy = -v\,du + (1-u)\,dv$, and $dz = (1+2v)\,du + 2(u-1)\,dv$. Since $\omega = -y\,dy \wedge dz + x\,dx \wedge dz$, this implies

$$\begin{aligned}
J^*(\omega) &= -(v - uv)(-v\,du + (1-u)\,dv) \wedge ((1+2v)\,du + 2(u-1)\,dv) \\
&\quad - u\,du \wedge ((1+2v)\,du + 2(u-1)\,dv) \\
&= ((v-2)u^2 + 2(1-v)u + v)\,du \wedge dv
\end{aligned}$$

The integral of ω over J is then

$$\int_J \omega = \int_{I^2} J^*(\omega) = \int_0^1 \int_0^1 ((v-2)u^2 + 2(1-v)u + v)\,du\,dv = 1/2$$

This is not a case where changing the parameterization simplifies the integration.

Orientation

A smoothly parameterized surface E comes equipped with a natural *orientation*. What do we mean by this? It will turn out to be important.

We begin by discussing the concept of orientation for $\mathbb{R}^2$. The ordered pair of vectors $(1,0), (0,1)$ is an ordered basis for this vector space. If we choose another ordered pair of basis vectors $(a,b), (c,d)$, then

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and

$$\begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a & c \\ b & d \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Thus, the matrix

$$A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$$

transforms the ordered basis $(1,0), (0,1)$ to a new ordered basis $(a,b), (c,d)$.

Now the matrix A must be non-singular since (a, b) and (c, d) are linearly independent. This means that $\det A \neq 0$. However, $\det A$ may be positive or it may be negative. This means that the possible ordered bases for $\mathbb{R}^2$ fall into two classes – those for which $\det A$ is positive and those for which $\det A$ is negative. A pair of ordered bases that fall into the same class are said to have the *same orientation* while a pair which fall into different classes are said to have *opposite orientation*. If we fix an ordered basis, then any other ordered basis is said to have *positive orientation* or *negative orientation* (relative to the fixed ordered basis) depending on whether or not it has the same or the opposite orientation of that of the original basis.

Example 11.5.8. For the following ordered pairs of basis vectors, tell which have the positive orientation and which have negative orientation with respect to the standard ordered basis $(1, 0), (0, 1)$:

1. $(0, 1), (1, 0)$;

2. $(0, -1), (1, 0)$;

3. $(1, 1), (-1, 1)$.

   Solution: We have

$$\det \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = -1, \quad \det \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = 1, \quad \det \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} = 2.$$

Thus, the first pair has negative orientation while the second and third pairs have the positive orientation with respect to the standard pair.

Of course specifying a coordinate system for the plane as well as a choice of ordering of the coordinate axes is the same as specifying an ordered basis. Thus, an orientation of the plane is determined by a choice of an ordered coordinate system.

Specifying an orientation on the plane is also equivalent to specifying a positive direction of rotation about a point. A non-zero rotation of magnitude less than $\pi/2$ is *positive* if it moves the positive x-axis toward the positive y-axis.

Surfaces and Orientation

A smooth p-surface S in $\mathbb{R}^q$ is a subset of $\mathbb{R}^q$ which is *locally* a smoothly parameterized p-surface. This means that at each point $s \in S$ there is a neighborhood U of s in $\mathbb{R}^q$ such that $S \cap U$ has a smooth parameterization.

Our main concern in this section is with 2-surfaces. They will be referred to simply as *surfaces*.

A smoothly parameterized 2-surface has a natural orientation. That is, if $H : U \to R^p$ is the map which parameterizes the surface S and $a \in U, b = H(a)$, then the linear transformation $dH : \mathbb{R}^2 \to \mathbb{R}^p$ maps $\mathbb{R}^2$ onto a 2-dimensional linear subspace L of $\mathbb{R}^p$ and it maps the standard basis $(1, 0), (0, 1)$ onto an ordered basis for L. Note that $b + L$ is the tangent space of S at the point b.
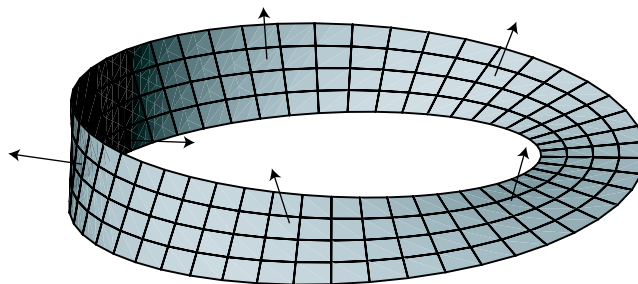
Figure 11.5: A Möbius Band

This ordered basis defines an orientation on L. This is what we mean by the orientation of the surface S at the point $b = H(a)$. Because H is smooth, the space L and the ordered pair of basis vectors vary in a continuous fashion as the point b moves about the surface S.

Suppose $H_1 : \mathbb{R}^p \to \mathbb{R}^q$ is another smoothly parameterized surface, with image S_1 which is equal to S in some neighborhood U of b. Then $U \cap S = U \cap S_1$ are surfaces with two different parameterizations. These parameterizations may determine the same orientation for the surface at b or opposite orientations. That is, the notion of orientation of a surface at a point depends on the choice of parameterization for the surface in a neighborhood of this point. This discussion leads to the following definition.

Definition 11.5.9. An orientation of a smooth surface S at at point $b \in S$ is the orientation class of a pair of basis vectors for the vector space L, where $b + L$ is the tangent space of S at b. An orientation for S itself is a choice of orientation for S at each of its points b in such a way that ordered basis vectors defining this orientation may be chosen in a continuously varying fashion as b moves over S. An orientable surface is one which may be given an orientation.

Surfaces in 3-Space

If H is a smoothly parameterized 2-surface S in $\mathbb{R}^3$ with parameter set U and trace S, then the images under dH of the basis vectors $(1, 0)$ and $(0, 1)$ are the first and second rows of the matrix dH. They may also be described as the vectors $\partial H / \partial u$ and $\partial H / \partial v$. They constitute an ordered pair of basis vectors for the vector space L such that $H(u, v) + L$ is tangent space of S at $H(u, v)$. As (u, v) range over U, they determine an orientation of S. The cross product of these vectors $\partial H / \partial u \times \partial H / \partial v$ is often called the normal vector to the parameterized surface and denoted N_H. This is a vector orthogonal to the vectors $\partial H / \partial u$ and $\partial H / \partial v$ and it varies continuously with the point $(u, v) \in U$. The cross product of any ordered basis of vectors in L will have the same or opposite direction as N_H depending on whether or not the ordered basis determines the same orientation as $(\partial H / \partial u, \partial H / \partial v)$. In other words, the direction of N_H at a point

on the surface determines the orientation of the ordered pair $(\partial H/\partial u, \partial H/\partial v)$ and, hence, the orientation of S at that point. The following theorem follows from this observation.

Theorem 11.5.10. *An orientation on a surface S in $\mathbb{R}^3$ is determined by a continuous function which assigns to each point of S a vector orthogonal to the tangent space of S at that point. There exists such a function if and only if the surface is orientable.*

Most of the common surfaces we deal with in $\mathbb{R}^3$ are orientable. This includes spheres, cylinders, tori, and any smoothly parameterized surface. However, not all surfaces in $\mathbb{R}^3$ are orientable, as the next example shows.

Example 11.5.11. Find a surface in $\mathbb{R}^3$ which is not orientable.

Solution: Such a surface is the *Möbius band*, illustrated in Figure 11.5. Note that an attempt to continuously assign a normal vector to the points of this surface, beginning at the left and proceeding in the counterclockwise direction, results in the vectors pointing in the opposite of the original direction once we return to the starting point.

A physical example of a Möbius band may be constructed by taking a long, thin rectangular strip of paper, twisting one end through 180 degrees and then glueing it to the opposite end.

Surface Integrals in 3-Space

Let H be a smoothly parameterized 2-surface in $\mathbb{R}^3$ with trace S. The unit normal to the surface A is defined to be $N = N_H/||N_H||$. This appears to depend on the parameterization H and not just one its trace S and, in fact, by definition, it is a function on the parameter set U of H. However, at a given point of S, there are only two unit vectors which are orthogonal to the tangent plane of S and they point in opposite directions. Thus, if two parameterizations of S give it the same orientation, then they must determine the same normal vector at each point (this also follows from Exercise 11.5.7). In other words, for a smooth oriented surface, there is a a uniquely defined unit normal vector at each point of the surface. For this reason, we consider the unit normal vector to be a function of points (x, y, z) on the surface S, rather than a function of points (u, v) in the parameter set U. Given a parameterization H of the surface, we recover N_H as

$$N_H(u, v) = ||N_H(x, y)||N(H(u, v)) \quad \text{or} \quad N_H = ||N_H||N \circ H.$$

Surface Area

Just as we defined the arc length s of a path and the integral over a path with respect to the differential ds of arc length, we may define the area of a parameterized surface and the integral of a function with respect to the differential of surface area.

Definition 11.5.12. If $H : U \to \mathbb{R}^3$ is a smoothly parameterized surface, then we define the surface area of the trace S of H to be

$$\sigma(S) = \int_U ||N_H(u,v)|| \, du \wedge dv.$$

If f is a continuous function defined on S, then we define its integral with respect to surface area on S to be

$$\int_S f \, d\sigma = \int_H f \, d\sigma = \int_U (f \circ H)(u,v)||N_H(u,v)|| \, du \wedge dv,$$

This is independent of the parameterization in the sense that if G is another smoothly parameterized surface which is related to H by a smooth parameter change P, then the integrals in the above definition are unchanged if we replace H by $G = H \circ P$. This is due to the change of variables theorem (Theorem 11.4.4) and the fact that $N_{H \circ P} = \det(dP)N_H \circ P$ (Exercise 11.5.7). This shows, in particular, that the surface area of the trace S of H is independent of the parameterization.

Let $(x,y,z) = H(u,v)$, $(u,v) \in U$ be a smooth parameterization of a 2-surface S in $\mathbb{R}^3$, as above, and let $\phi = f_1 dy \wedge dz + f_2 dz \wedge dx + f_3 dx \wedge dy$ be a 2-form defined on a neighborhood of the trace of H. By Example 11.3.12, if we let $F = (f_1, f_2, f_3)$ be the component vector field of ϕ, then

$$H^*(\phi) = (F \circ H) \cdot \left[\frac{\partial H}{\partial u} \times \frac{\partial H}{\partial v}\right] du \wedge dv = (F \circ H) \cdot N_H \, du \wedge dv, \quad (11.5.1)$$

If we use the notation, $d\sigma = ||N_H||du \wedge dv$, this allows us to express the integral of the 2-form ϕ over a smoothly parameterized surface H in its classical form as an integral with respect to surface area over the trace S of H.

$$\int_H \phi = \int_S F \cdot N \, d\sigma. \quad (11.5.2)$$

This has physical interpretations in certain situations. For example, if F is the velocity field of a fluid moving in $\mathbb{R}^3$, then the integral represents the *flux* or rate of flow of the fluid across the surface S.

Integration over a 2-Cell in $\mathbb{R}^d$

In the previous section, we defined 2-cells in $\mathbb{R}^d$ (Definition 11.4.5).

We may think of a simple 2-cell in $\mathbb{R}^d$ as smoothly parameterized 2-surface in $\mathbb{R}^d$ along with a path ∂E which runs around the edge of this surface. The boundary, ∂E, of a 2-cell E is, as before, the path which is the composition of E with the path ∂I^2 in $\mathbb{R}^2$. In general, this will not be the same as the topological boundary of $E(I^2)$. In particular, in dimensions higher than 2, the trace $E(I^2)$ has no interior and is, therefore, its own topological boundary, whereas ∂E is just a path in $\mathbb{R}^d$ which runs around the edge of $E(I^2)$.

Considered as a smoothly parameterized 2-surface defined on the interior of I^2, a 2-cell E satisfies the conditions of Definition 11.5.2. Similarly, a 2-form ϕ defined on a set containing the trace $E(I^2)$ is continuous, hence bounded, on $E(I^2)$ and so it also satisfies the conditions of Definition 11.5.2. Hence, the integral

$$\int_E \omega = \int_{I^2} E^*(\omega).$$

exists. It is this surface integral over a 2-cell E that we use in formulating Stokes's Theorem.

Stokes' Theorem

Stokes' Theorem for two dimensional surfaces is much like Green's Theorem. The difference is that two dimensional surfaces lying in $\mathbb{R}^d$ for $d \geq 3$ replace regions in $\mathbb{R}^2$. The result is still stated in terms of 2-cells, but now they are 2-cells in dimension higher than 2. We will be primarily concerned with 2-cells in $\mathbb{R}^3$.

Theorem 11.5.13. Stokes' Theorem *Let $E : I^2 \to \mathbb{R}^d$ be a 2-cell and ϕ a smooth 1-form defined on an open set in $\mathbb{R}^3$ containing $E(I^2)$. Then*

$$\int_{\partial E} \phi = \int_E d\phi.$$

The proof is identical to the proof of Green's Theorem (Theorem 11.4.10).

Remark 11.5.14. As with Green's Theorem, although Stoke's Theorem is stated in terms of a cell E and its boundary ∂E, in practice the integrals over E and ∂E may be computed using convenient parameterizations which may have little to do with each other.

Example 11.5.15. Use Stokes' Theorem to calculate the integral of the 2-form $(x + y)\, dz$ around the boundary of the surface

$$z = x^2 - y^2 - 2x + 2y, \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1,$$

where the boundary path is traversed in the counterclockwise direction when seen from above the surface (positive z axis points up).

Solution: We parameterize the surface by setting $x = u, y = v, z = u^2 - v^2 - 2u + 2v$. That is, we represent the surface as the trace of the 2-cell $E(u, v) = (u, v, u^2 - v^2 - 2u + 2v)$, $(u, v) \in I^2$. Traversing ∂I^2 in the counterclockwise direction causes $E(u, v)$ to traverse the boundary of our surface in the required direction. Since $d((x + y)\, dz) = dy \wedge dz - dz \wedge dx$, Stokes' Theorem implies that

$$\int_{\partial E} (x + y)\, dz = \int_E (dy \wedge dz - dz \wedge dx).$$

We have $dx = du$, $dy = dv$ and $dz = (2u - 2)\, du - (2v - 2)\, dv$ for the parameterization determined by E. Thus,

$$\int_{\partial E} (x + y)\, dz = \int_{I^2} (4 - 2u - 2v)\, du \wedge dv$$

$$= \int_0^1 \int_0^1 (4 - 2u - 2v)\, du\, dv = 2$$

Example 11.5.16. Let $\omega = x\, dy \wedge dz - y\, dz \wedge dx - 2y\, dx \wedge dy$. Find the integral of the 2-form ω over the torus T described as follows: For each point on the circle $A = \{(x, y, 0) \in \mathbb{R}^3 : x^2 + y^2 = 4\}$, let $C_{x,y}$ be a circle in R^3, of radius 1, which is centered at $(x, y, 0)$ and lies in the plane through the origin perpendicular to the circle A. Then T is the union of all the circles $C_{x,y}$ (see Figure 11.6). Note that T is a smooth 2-dimensional surface.

   Solution: We may parameterize T as follows:

$$x = (2 + \cos 2\pi t) \cos 2\pi s;$$
$$y = (2 + \cos 2\pi t) \sin 2\pi s;$$
$$z = \sin 2\pi t,$$

with $0 \le s \le 1$ and $0 \le t \le 1$. In other words, T is the trace of the 2-cell $E : I^2 \to \mathbb{R}^3$ given by

$$E(s, t) = ((2 + \cos 2\pi t) \cos 2\pi s, (2 + \cos 2\pi t) \sin 2\pi s, \sin 2\pi t).$$

Now the 2-form ω is $d\phi$ where ϕ is the 1-form $\phi = y^2\, dx + xy\, dz$. Thus, by Stokes' Theorem

$$\int_E \omega = \int_E d\phi = \int_{\partial E} \phi. \tag{11.5.3}$$

However, ∂E is made up of four parameterized circles. Two of them are

$$\gamma_1(t) = ((2 + \cos 2\pi t), 0, \sin 2\pi t), \quad \text{and} \quad \gamma_2(s) = (3 \cos 2\pi s, 3 \sin 2\pi s, 0).$$

and the other two are $\gamma_3(t) = \gamma_1(1 - t)$ and $\gamma_4(s) = \gamma_2(1 - s)$ – that is, γ_3 and γ_4 are just γ_1 and γ_2 traversed in the reverse direction. It follows that the contributions of the integrals over these paths cancel and, hence, that the integrals in (11.5.3) are all 0.

Classical Form of Stokes's Theorem

If $\phi = f_1\, dx + f_2\, dy + f_3\, dz$ is a 1-form and F is the component vector field $F = (f_1, f_2, f_3)$, then by Remark 11.3.4, $d\phi$ has $\mathrm{curl}(F)$ as its component vector field. Using this and (11.5.2) yields the classical form of Stokes's Theorem.

Theorem 11.5.17. *Let E be a simple 2-cell in $\mathbb{R}^3$ with trace S and let $\phi = f_1\, dx + f_2\, dy + f_3\, dz$ a 1-form defined on the trace of E. With N the normal*
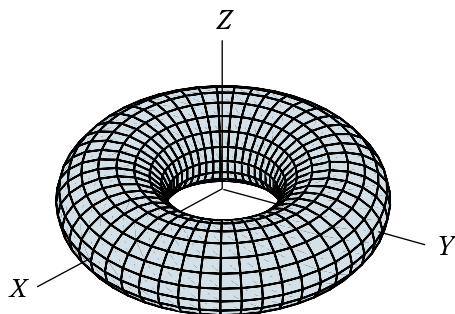
Figure 11.6: The Torus of Example 11.6

vector for E as defined above, T the tangent vector to the path ∂E, and F the vector field $F = (f_1, f_2, f_3)$, we have

$$\int_{\partial E} F \cdot T \, ds = \int_S \operatorname{curl}(F) \cdot N d\sigma$$

Proof. The integral on the left is just the path integral $\int_{\partial E} \phi$ interpreted as in Theorem 11.4.15. By Stokes's Theorem, Remark 11.3.4, and (11.5.2) this is equal to

$$\int_E d\phi = \int_S \operatorname{curl}(F) \cdot N \, d\sigma.$$

□

Exercise Set 11.5

1. For the part of the surface $x + y + z = 1$ that lies in the first octant, find a smooth parameterization H for which the normal vector points up, and then compute the integral of the 2-form $\omega = x^2 \, dy \wedge dz$ over H.

2. For the surface $z = 1 - x^2 - y^2$, $z > 0$, find a smooth parameterization H, with normal vector pointing up, and then compute the integral of the 2-form $\omega = x \, dy \wedge dz + y \, dz \wedge dx + z \, dx \wedge dy$ over this surface.

3. For the smoothly parameterized surface in $\mathbb{R}^3$ defined by

$$H(u, v) = (5u, \cos 2\pi v, \sin 2\pi v), \quad (u, v) \in I^2,$$

describe the trace of H and then compute the integral over H of the 2-form $\omega = y \, dy \wedge dz - x \, dz \wedge dx + 2 \, dx \wedge dy$.

4. Find the integral over the sphere $x^2 + y^2 + z^2 = 1$ of the 2-form $dy \wedge dz - 2dz \wedge dx$.

5. If H is a parameterized 2-surface in $\mathbb{R}^3$, with $(x, y, z) = H(u, v)$, and if $N_H = (g_1, g_2, g_3)$ is its normal vector field, show that $H^*(dy \wedge dz) = g_1 \, du \wedge dv$, $H^*(dz \wedge dx) = g_2 \, du \wedge dv$, and $H^*(dx \wedge dy) = g_3 \, du \wedge dv$.

6. Find the normal N_H and unit normal N for the parameterized torus of Example 11.5.16.

7. Show that if $H : U \to \mathbb{R}^3$ is a smoothly parameterized 2-surface and $P : V \to U$ is a smooth parameter change, then the normal vectors of H and $H \circ P$ are related by $N_{H \circ P} = \det(dP) N_H \circ P$.

8. Let H be a parameterized surface in $\mathbb{R}^3$ with trace S and let $N = (\eta_1, \eta_2, \eta_3)$ be the unit normal vector field on S. Show that the area of S is $\int_H \eta$, where $\eta = \eta_1 dy \wedge dz + \eta_2 dz \wedge dx + \eta_3 dx \wedge dy$. Hint: use (11.5.2).

9. Use Stokes's Theorem to compute the integral of the 2-form

$$\omega = y \, dy \wedge dz + z \, dz \wedge dx + \, dx \wedge dy$$

over the hemisphere $x^2 + y^2 + z^2 = 1$, $z > 0$, oriented so that the normal vector points up. Hint: $\omega = d\phi$ for a certain 1-form ϕ.

10. If $F(x, y, z) = (xy, yz, xz)$ and S is that part of the plane $x + y + z = 1$ which lies in the first octant, oriented such that the normal vector N points up, use the classical form of Stokes's Theorem to compute $\int_A \mathrm{curl}(F) \cdot N \, d\sigma$.

11. If $\phi = z \, dx + 3x \, dy - y \, dz$, use Stokes's Theorem to compute the integral of the 1-form ϕ over the ellipse which is the intersection of the cylinder $x^2 + y^2 = 9$ with the plane $z = x$. Hint: the ellipse is the boundary of the surface consisting of that part of the plane $z = x$ which lies inside the cylinder.

12. Show why the integral of $d\phi$ over the sphere

$$S = \{(x, y, z) : x^2 + y^2 + z^2 = 1\}$$

is 0 for every smooth 1-form ϕ on S.

11.6   Gauss's Theorem

In this section, we generalize Green's Theorem to the case of a 3-cell in $\mathbb{R}^3$. The result is Gauss's Theorem. It relates the integral of a 3-form ϕ over a 3-cell with the integral of $d\phi$ over the boundary of the 3-cell. We begin with a brief discussion of integrals of 3-forms in $\mathbb{R}^3$.

The Integral of a 3-Form

A 3-form in $\mathbb{R}^3$ has the form $\phi = f \, dx \wedge dy \wedge dz$ for some continuous function f. As with 2-forms in $\mathbb{R}^2$, we define the integral of such a thing over a Jordan region U to be

$$\int_U \phi = \int_U f \, dV. \tag{11.6.1}$$

Just as it did with integrals of 2-forms, the change of variables theorem leads to change of variables theorem for integration of 3-forms. The proof is the same as the proof of the 2-dimensional version in Theorem 11.4.4.

Theorem 11.6.1. *Let H be a smooth transformation from the open Jordan region U in $\mathbb{R}^3$ to another Jordan region in $\mathbb{R}^3$ and suppose H is one-to-one with non-singular differential on U. If ϕ is a bounded 3-form on $H(U)$ and $H^*(\phi)$ is bounded on U, then*

$$\int_{H(U)} \phi = \int_U H^*(\phi)$$

provided $\det(dH) > 0$ *everywhere on U. If* $\det(dH) < 0$ *on U, equality holds if the right side of the equation is replaced by its negative.*

A transformation H which satisfies the above conditions will be called a *smooth parameter change.*

Example 11.6.2. Find the integral of the 3-form $z(x^2 + y^2) \, dx \wedge dy \wedge dz$ over the truncated cone $C = \{(x, y, z) : x^2 + y^2 < z^2, \ 1 < z < 2.$

Solution: We could do this problem as an ordinary triple integral in rectangular coordinates. However, we choose to parameterize C using something like cylindrical coordinates (conical coordinates, actually). That is, we let R be the rectangle defined by $0 < r < 1$, $0 < \theta < 2\pi$, and $1 < z < 2$ and define $H : R \to C$ by

$$H(r, \theta, z) = (rz \cos \theta, rz \sin \theta, z).$$

That is, we make the change of variables

$$x = rz \cos \theta, \quad y = rz \sin \theta, \quad z = z.$$

Then

$$dx = z \cos \theta \, dr - rz \sin \theta d\theta + r \cos \theta \, dz$$
$$dy = z \sin \theta \, dr + rz \cos \theta d\theta + r \sin \theta \, dz$$
$$dz = dz,$$

so that $dx \wedge dy \wedge dz = rz^2 \, dr \wedge d\theta \wedge dz$, while $z(x^2 + y^2) = r^2 z$. Thus,

$$H^*(\phi) = r^3 z^3 \, dr \wedge d\theta \wedge dz.$$

and

$$\int_H \phi = \int_R H^*(\phi) = \int_1^2 \int_0^{2\pi} \int_0^1 r^3 z^3 \, dr \, d\theta \, dz = \frac{15\pi}{8}.$$

The Boundary of a Cube

Our next task is to prove Gauss's Theorem on the standard cube I^3 in $\mathbb{R}^3$. In order to formulate the theorem we need to fix an orientation on the boundary of the cube.

The boundary of a cube is not a smooth surface. It consists of six squares, which are smooth surfaces, joined together along their sides. We choose to orient each of these in such a way that a corresponding normal vector points away from the cube. That is, an ordered pair of vectors in one of the sides has the correct orientation if the cross product of these vectors points to the exterior of the cube.

One way to parameterize the six faces is as follows: We let (s, t) be the coordinates of a point on the standard square I^2. Then

$$F^{10}(s,t) = (0, s, t) \quad \text{and} \quad F^{11}(s,t) = (1, s, t)$$

parameterize the two faces perpendicular to the x-axis, while

$$F^{20}(s,t) = (s, 0, t) \quad \text{and} \quad F^{21}(s,t) = (s, 1, t)$$
$$F^{30}(s,t) = (s, t, 0) \quad \text{and} \quad F^{31}(s,t) = (s, t, 1)$$

parameterize the faces perpendicular to the y and z axes, respectively. Unfortunately, three of these have the wrong orientation. For example, F^{10} and F^{11} each send the standard basis in $\mathbb{R}^2$ to a pair of vectors in $\mathbb{R}^3$ with cross product pointing in the positive x direction. Hence, they don't both point to the exterior of the cube. In fact, for F^{10} this cross product vector points to the interior of the cube. In general, the orientation of $F^{i\sigma}$ is correct if $i + \sigma$ is even and is incorrect if $i + \sigma$ is odd. Thus, an integral over a face with $i + \sigma$ odd will have the wrong sign. We can fix this by multiplying the integral by -1. This idea leads to an interpretation of the boundary of the cube I^3 as a formal sum

$$\partial I^3 = \sum_{i\sigma} (-1)^{i+\sigma} F^{i\sigma} \tag{11.6.2}$$

where i runs from 1 to 3 and σ from 0 to 1. We then define the integral of a 2-form ϕ over ∂I^3 to be

$$\int_{\partial I^3} \phi = \sum_{i\sigma} (-1)^{i+\sigma} \int_{F^{i\sigma}} \phi \tag{11.6.3}$$

We would get the same result if we just reversed the orientation of each face $F^{i\sigma}$ with $i + \sigma$ odd and then took the sum of the integrals over the resulting parameterized surfaces. There is an advantage to writing the integral as in (11.6.3) which will become apparent in the next section.

With these conventions established, we may state and prove Gauss's Theorem for the standard cube in $\mathbb{R}^3$.

Gauss's Theorem on a Cube

The proof of Gauss's Theorem on a cube is not materially different from the proof of Green's Theorem on a square.

Theorem 11.6.3. *Suppose ϕ is a smooth 2-form defined on I^3. Then*

$$\int_{\partial I^3} \phi = \int_{I^3} d\phi$$

Proof. We first show that the theorem holds for ϕ of the form $\phi = f \, dy \wedge dz$. With ∂I^3 represented as in (11.6.2), we have

$$\int_{\partial I^3} \phi = \sum_{i\sigma}(-1)^{i+\sigma} \int_{F_{i\sigma}} \phi = \sum_{i\sigma}(-1)^{i+\sigma} \int_{F_{i\sigma}} f \, dy \wedge dz.$$

The integral on the right in this equation will vanish if either y or z is constant on the face $F_{i\sigma}$. Thus, only the integrals of $f \, dy \wedge dz$ over the faces F_{10} and F_{11} may be non-zero. This implies

$$\int_{\partial I^3} \phi = \int_0^1 \int_0^1 f(1,s,t)dsdt - \int_0^1 \int_0^1 f(0,s,t)dsdt$$

$$= \int_0^1 \int_0^1 \int_0^1 \frac{\partial f}{\partial x}(x,s,t)dxdsdt = \int_{I^3} d\phi,$$

by the Fundamental Theorem of Calculus applied to the integral in the x direction.

If ϕ has the form $g \, dy \wedge dz$ or $h \, dx \wedge dz$, the proof is the same with the variables and the value of i interchanged. Since every smooth 2-form is a sum of forms for which the theorem is true and since the integrals involved are linear functions of the forms in the integrand, the theorem is true in general. $\square$

Gauss's Theorem for a 3-Cell

Definition 11.6.4. A 3-cell in $\mathbb{R}^3$ is a smooth function $E : I^3 \to \mathbb{R}^3$. A 3-cell is simple if it is one-to-one with non-singular differential on the interior of I^3. A simple cell E is positively oriented if $\det(dE) > 0$ on the interior of E.

As in the definition of 2-cell, the meaning of *smooth* requires some comment, since I^2 is not an open set. Along each face or edge of I^2 some of the partial derivatives of the coordinate functions of E must be interpreted as one-sided derivatives, while at interior points of I^2 these are the usual 2-sided derivatives. The resulting functions on I^2 are then required to be continuous.

The faces of E are the functions $E^{i\sigma} = E \circ F^{i\sigma}$, where $F^{i\sigma}$ is the $i\sigma$ face of I^3 as defined at the beginning of this section. Thus, $E^{10}(s,t) = E(0,s,t)$, $E^{11}(s,t) = E(1,s,t)$, $E^{20}(s,t) = E(s,0,t)$, etc. It follows from the above definition that each face is a 2-cell.

The boundary of a 3-cell E is defined to be

$$\partial E = \sum_{i\sigma} (-1)^{i+\sigma} E^{i\sigma},$$

where, as in (11.6.3), this means that the integral of a 2-form ϕ over ∂E is defined to be

$$\int_{\partial E} \phi = \sum_{i\sigma} (-1)^{i+\sigma} \int_{E^{i\sigma}} \phi.$$

The following is Gauss's Theorem for a 3-cell.

Theorem 11.6.5. *If E is a smooth 3-cell in $\mathbb{R}^3$ and ϕ a smooth 2-form on the trace of E, then*

$$\int_{\partial E} \phi = \int_E d\phi.$$

Proof. This is just like the proof of Green's Theorem for a 2-cell. We have

$$\int_{\partial E} \phi = \int_{\partial I^3} E^*(\phi) = \int_{I^3} dE^*(\phi)$$

$$= \int_{I^3} E^*(d\phi) = \int_E d\phi,$$

by Theorem 11.6.3 and Theorem 11.3.10(c). $\square$

Example 11.6.6. Find the integral of the 2-form

$$\phi = (x^2 + y)\, dy \wedge dz + (2xz - y)\, dx \wedge dz + (xy^2 + z)\, dx \wedge dy$$

over the boundary of the solid A defined by the inequalities $0 \le z \le 1 - x^2 - y^2$.

Solution: We use Gauss's Theorem, which tells us that the integral we seek is equal to $\int_A d\phi$. We will parameterize A using cylindrical coordinates

$$x = r\cos t, \quad y = r\sin t, \quad z = z \quad \text{with} \quad 0 \le z \le 1 - r^2,\ 0 \le r \le 1,\ 0 \le t \le 2\pi.$$

Since $d\phi = (2x + 2)\, dx \wedge dy \wedge dz = 2r(r\cos t + 1)dr \wedge dt \wedge dz$, we have

$$\int_A d\phi = \int_0^1 \int_0^{1-r^2} \int_0^{2\pi} 2(r^2 \cos t + r)dt\, dz\, dr = \int_0^1 \int_0^{1-r^2} 4\pi r\, dz\, dr = \pi.$$

Example 11.6.7. For $0 \le b \le 1$, let B be that part of the solid sphere of radius one, centered at the origin, that lies between the planes $z = -b$ and $z = b$. Compute the volume of B in two ways – first, as an integral over B and, second, as a surface integral over ∂B.

Solution: The volume we seek is $\int_B dx \wedge dy \wedge dz$. We parameterize B using cylindrical coordinates. Then $x = r\cos\theta$, $y = r\sin\theta$ and $z = z$ with $0 \le r \le 1$,
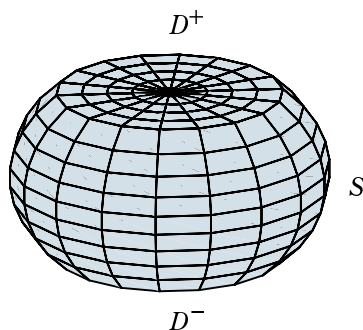
Figure 11.7: Horizontal Slice of a Sphere

$0 \leq \theta \leq 2\pi$, and $-b \leq z \leq b$. We know $dx \wedge dy \wedge dz = r \, dr \wedge d\theta \wedge dz$ and $r = \sqrt{1 - z^2}$ at points on the surface of the sphere. Thus,

$$\int_B dx \wedge dy \wedge dz = \int_{-b}^b \int_0^{2\pi} \int_0^{\sqrt{1-z^2}} r \, dr \, d\theta \, dz$$

$$= \int_{-b}^b \pi(1 - z^2) \, dz = 2\pi(b - b^3/3).$$

This is the result of the calculation of the volume integral.

To compute the volume of B as a surface integral we use Gauss's Theorem. Since $d(z \, dx \wedge dy) = dz \wedge dx \wedge dx = dx \wedge dy \wedge dz$, Gauss's Theorem tells us that

$$\int_B dx \wedge dy \wedge dz = \int_{\partial B} z \, dx \wedge dy = \int_{\partial B} zr \, dr \wedge d\theta$$

where the latter integral results from switching to cylindrical coordinates.

The surface ∂B is made up of three parts: a section S of of the sphere defined by the conditions $r = \sqrt{1 - z^2}$, $-b \leq z \leq b$, and top and bottom horizontal discs D^+ and D^- defined by $z = \pm b$, $0 \leq r \leq \sqrt{1 - b^2}$.

The horizontal discs each have radius $\sqrt{1 - b^2}$ and so the contribution of the top disc D^+ to the integral $\int_{\partial B} zr \, dr \wedge d\theta$ is $b(1 - b^2)\pi$. The bottom disc D^- appears, at first glance, to yield the negative of this since everything appears to be the same except that $z = b$ on D^+ and $z = -b$ on D^-. However, this is not correct. As part of ∂B, the bottom disc D^- has negative orientation relative to the standard x, y coordinates in the plane while D^+ has positive orientation. The negative orientation of D^- reverses the direction of integration with respect to θ and, hence, reverses the sign of the integral. This leads to a result which is identical to that computed for D^+. Thus, the combined contribution of D^- and D^+ to $\int_{\partial B} zr \, dr \wedge d\theta$ is $2\pi b(1 - b^2)$.

To compute the contribution of the spherical section S, we use the z and θ

coordinates to parameterize S. Then $r = \sqrt{1 - z^2}$ on S

$$\int_{\partial S} zr \, dr \wedge d\theta = \int_0^{2\pi} \int_{-b}^{b} z\sqrt{1 - z^2} \, dz \, d\theta = 4b^3\pi/3.$$

Adding the various contributions gives us

$$\int_B dx \wedge dy \wedge dz = \int_{\partial B} zr \, dr \wedge d\theta = 4/3 \, b^3\pi + 2b(1 - b^2)\pi = 2\pi(b - b^3/3).$$

Fortunately, this is the same answer as before.

Classical Form of Gauss's Theorem

If $\phi = f_1 \, dy \wedge dz + f_2 \, dz \wedge dx + f_3 \, dx \wedge dy$ is a 2-form in $\mathbb{R}^3$ and we let $F = (f_1, f_2, f_3)$ be its component vector field, then

$$d\phi = \operatorname{div} F \, dx \wedge dy \wedge dz,$$

where $\operatorname{div} F = \partial f_1/\partial x + \partial f_2/\partial y + \partial f_3/\partial z$. If we combine this with (11.5.2) and Theorem 11.6.5, the result is the classical form of Gauss's Theorem:

Theorem 11.6.8. *If E is a 3-cell in $\mathbb{R}^3$ with trace A. Suppose ∂E has trace equal to the topological boundary ∂A of A, and F is a smooth vector function defined on the trace A of E, then*

$$\int_{\partial A} F \cdot N \, d\sigma = \int_A \operatorname{div} F \, dV.$$

In a fluid flow problem, where F is the velocity field of the flow, this has the following interpretation. The left side represents the *flux* or rate of flow of fluid out of the region A, while the right side is the integral over A of a function $\operatorname{div} F$ which represents, at each point of A, the tendency of the fluid to move away from (diverge from) the point.

The Integral over a 3-Surface in $\mathbb{R}^d$

A smoothly parameterized 3-surface in $\mathbb{R}^d$ is a smooth function $H : U \to \mathbb{R}^d$ such that U is an open subset of $\mathbb{R}^3$ and dH is non-singular on U. The trace of H is its image in $\mathbb{R}^d$.

   Just as an ordered basis for a 2-dimensional vector space determines an orientation for the vector space, an ordered basis for a vector space of dimension 3 or higher also determines an orientation for the vector space. Two ordered bases determine the same orientation if and only if the determinant of the matrix which transforms the first basis to the second is positive.

   As before, H determines an orientation on its trace $S = H(I^3)$. That is, $dH(a)$ sends the standard basis in $\mathbb{R}^3$ to an ordered basis for the linear subspace of $\mathbb{R}^d$ whose translate by $b = H(a)$ is the tangent space to S at b. A 3-surface

in $\mathbb{R}^d$ is a subset which, in a neighborhood of each of its points, may be given a smooth parameterization – that is, its intersection with this neighborhood is the trace of a smoothly parameterized 3 surface. A 3-surface is *orientable* if there is a smooth function which assigns an ordered basis, above, to each point of the surface.

We define the integral of a 3-form over a smoothly parameterized 3-surface in $\mathbb{R}^d$ in the same way we defined the integral of a 2-form over a smoothly parameterized 2-surface.

Definition 11.6.9. If U is an open Jordan region, $H : U \to \mathbb{R}^d$ is a smoothly parameterized 3-surface, and ϕ is a 3-form on $H(U)$ such that $H^*(\phi)$ is bounded on U, we set

$$\int_H \phi = \int_U H^*(\phi).$$

This defines the integral on the left.

As before, this integral, though defined through the parameterization H is actually independent of parameterization in the sense that the integral is unchanged if H is replaced by $J = H \circ P$, where $P : V \to U$ is any positively oriented smooth parameter change, provided V and J also satisfy the conditions of the above definition. The integral does depend on the orientation of H and if this is reversed, then the integral changes sign. Here, a smooth parameter change $P : V \to U$ between open Jordan regions in $\mathbb{R}^3$ is a smooth one-to-one map with non-singular differential $dP > 0$. It is positively oriented if $\det dP > 0$.

Stokes's Theorem for 3-Cells in $\mathbb{R}^d$

The definition of a 3-cell in $\mathbb{R}^d$ is the same as that of a 3-cell in $\mathbb{R}^3$ except that the trace of the cell lies in $\mathbb{R}^d$ rather than $\mathbb{R}^3$. Since, on the interior of I^3, a 3-cell is a smoothly parameterized surface, we may integrate a 3-form over it. With no extra work, we have Stokes's Theorem for a 3-cell in $\mathbb{R}^d$, for any $d \geq 3$. Its proof is the same as the proof of Gauss's Theorem.

Theorem 11.6.10. *If $E : I^3 \to \mathbb{R}^d$ is a 3-cell in $\mathbb{R}^d$, and ϕ is a 3-form defined on the trace of E, then*

$$\int_{\partial E} \phi = \int_E d\phi.$$

In the next section, we will state the general form of Stokes's Theorem, which involves integrals over p-cells in $\mathbb{R}^q$ for any $q \geq p$.

Exercise Set 11.6

1. Suppose E is a positively oriented simple 3-cell in $\mathbb{R}^3$ with trace A. Show that the volume of A is

$$V(A) = \int_{\partial E} \frac{1}{3}(x\,dy \wedge dz + y\,dz \wedge dx + z\,dx \wedge dy).$$

2. Let C be the solid defined by

$$C = \{(x, y, z) \in \mathbb{R}^3 : x^2 + y^2 \leq z \leq 1\}.$$

Use Gauss's Theorem to find the integral over the boundary of C of the 2-form

$$\phi = (x + y \sin^5 z)\, dy \wedge dz + (y - \cos zx)\, dz \wedge dx + (3z^2 + \ln(1 + xy))\, dx \wedge dy.$$

3. Show how to construct a 3-cell with trace equal to

$$A = \{(x, y, z) \in \mathbb{R}^3 : a^2 \leq x^2 + y^2 + z^2 \leq b^2\}.$$

4. For a 3-cell E as in the previous exercise, and a 2-form ϕ on A, show that

$$\int_E d\phi = \int_{C_b} \phi - \int_{C_a} \phi$$

where C_a and $\mathbb{C}_b$ are the spheres of radius a and b, respectively, oriented so that the normal vectors point to the exterior of the sphere. If $d\phi = 0$, what do you conclude.

5. Show how to extend the result of the previous exercise to more general situations where one surface is the boundary of a solid A and the second surface is the boundary of a second solid B which is contained in the interior of A.

6. Let F be a $\mathcal{C}^1$ vector field on an open set $U \subset \mathbb{R}^3$. If $a \in U$, use Gausses Theorem to prove that

$$\operatorname{div} F(a) = \lim_{r \to 0} \frac{1}{V(B_r(a))} \int_{\partial B_r(a)} F \cdot N \, d\sigma.$$

7. Let U be an open set in $\mathbb{R}^3$ such that $\overline{U}$ is the trace of a 3-cell E and let $F = (f_1, f_2, f_3)$ a vector field on the trace $\overline{U}$. There is a 1-form ϕ with F as component vector field and a 2-form ϕ^* with F as component vector field. That is,

$$\phi = f_1\, dx + f_2\, dy + f_3\, dz \quad \text{and} \quad \phi^* = f_1\, dy \wedge dz + f_2\, dz \wedge dx + f_3\, dx \wedge dy.$$

Show that

(a) $\phi \wedge \phi^* = F \cdot F\, dx \wedge dy \wedge dz = ||F||^2\, dx \wedge dy \wedge dz$;

(b) if $\phi = dg$ for some continuous function g on $\overline{U}$ which is $\mathcal{C}^2$ on U, then $d\phi^* = \Delta g$, where $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ is the Laplacian;

(c) If g is harmonic (i. e. if $\Delta g = 0$ on U), then $\int_U ||F||^2 dV = \int_{\partial E} g\phi^*$.

(d) if g is harmonic and $g = 0$ on the trace of ∂E, then g is identically 0 on U.

8. Let $r : \mathbb{R}^3 \backslash \{0\} \to R$ be the function $r(x, y, z) = \sqrt{x^2 + y^2 + z^2}$. Using the notation of the previous exercise, compute dr, show that $d(1/r) = -dr/r^2$ and $(d(1/r))^* = dr^*/r^2$. Show that $d(dr^*/r^2) = 0$ and, hence, that $1/r$ is harmonic on $R^3 \backslash \{0\}$.

9. The gravitational force field due to a mass at the origin is a constant k times the component vector field of the 2-form dr^*/r^2 of the previous exercise. Show that if S is a solid sphere in $\mathbb{R}^3$, centered at the origin, then the flux across ∂S due to this field is

$$\int_{\partial S} k \frac{dr^*}{r^2} = -4k\pi.$$

Hint: for the surface ∂S, show that N is the component vector field of dr^* restricted to ∂S. Then use the classical expression for a surface integral (11.5.2).

10. Use Gauss's Theorem to show that the integral in the previous exercise does not change if the sphere S is replaced by any reasonable solid A with 0 in its interior. What reasonable assumptions on A will make this true?

11.7 Chains and Cycles

Much of what we have done with Green's, Stokes's and Gauss's Theorems in the previous section involves involves integration over cells. However, in some cases, we have worked with integrals over objects which are sums of cells in a certain sense. In particular, an integral over the boundary of a cell is not an integral over a cell, but a sum of integrals over the several cells which form the boundary. In the previous section we came to think of the boundary of a 3-cell as a formal linear combination (11.6.2) of 2-cells corresponding to the faces of I^3. This suggests that, for any natural number k, we think of the boundary of a k-cell as a formal linear combination of the cells which consist of restrictions of the cell to the various faces of I^k. This will require a theory of integration, not just over cells, but over formal linear combinations of cells. Expanding on this idea leads to some very powerful and far reaching concepts in mathematics. In this section, we will give a brief introduction to this formalism and then use it to restate Green's, Stokes's and Gauss's Theorem in their modern form.

We begin with an introduction to this idea in the context of paths. Here the objects we wish to introduce are 1-chains and 1-cycles.

1-Chains

A path γ in $\mathbb{R}^d$ is piecewise smooth, which means that it may be thought of as several smooth paths $\gamma_1, \cdots, \gamma_n$ joined together end to end to form a single path. The integral of a function over γ is then the sum of the integrals over the paths γ_j. we may reparameterize each of these paths so as to have parameter

interval $I = [0, 1]$ without effecting the integral (Exercise 11.1.4). The formal sum of the paths γ_j is then a *1-chain* in the sense of the following definition.

Definition 11.7.1. A *1-chain* in $\mathbb{R}^d$ is a formal finite linear combination, with integral coefficients,

$$\Gamma = \sum_{j=1}^{p} m_j \gamma_j, \tag{11.7.1}$$

of smooth paths in $\mathbb{R}^d$.

Note that (11.7.1) is not a linear combination of the γ_j as functions on $[0, 1]$ – that is, the multiplication by integers and the sums are not pointwise sums of $\mathbb{R}^d$ valued functions. It is purely a formal expression and cannot be simplified or manipulated until we impose some rules for manipulating such expressions. We do this below.

We agree that if the individual terms $m_j \gamma_j$ in a chain are rearranged, so that they appear in a different order, then the chain does not change. We agree that the chain does not change if we drop summands $m_j \gamma_j$ with $m_j = 0$, and we agree that two summands $m_j \gamma_j$ and $m_k \gamma_k$ with $\gamma_j = \gamma_k$ may be combined to yield $(m_j + m_k) \gamma_j$. The empty chain – that is the chain with no summands is denoted by 0. We add two chains in the obvious way: the sum of two formal linear combinations of paths is another formal linear combination of paths. The operation of addition, so defined, is clearly associative and commutative.

The set of 1-chains, as defined above, forms a commutative group – that is, it has an operation $(+)$ which is associative and commutative, there is a zero element (the linear combination with no summands) and each element has an additive inverse (just replace each coefficient m_j by $-m_j$).

Definition 11.7.2. The expression (11.7.1) for a chain Γ is said to be in *reduced form* if the γ_j are distinct paths and all the m_j are non-zero. Note that each chain may be expressed in reduced form. We define the *trace* of a chain Γ to be

$$\Gamma(I) = \bigcup_{j=1}^{p} \gamma_j(I),$$

where (11.7.1) is an expression of the chain in reduced form.

1-Cycles

A 0-chain in $\mathbb{R}^d$ is a formal linear combination, with integral coefficients, of singleton subsets of $\mathbb{R}^d$ – that is, a formal sum of the form

$$\sum_{j=1}^{p} m_j \{x_j\}.$$

with each x_j in $\mathbb{R}^d$.

Here, the sum is not a sum of vectors in $\mathbb{R}^d$. It is a purely formal sum and can only be manipulated using the rules we set down: Again, terms may be rearranged in the sum without changing the 0-chain. Terms with 0 as coefficient are dropped, and terms with the same $\{x_j\}$ may be combined by adding their coefficients. The empty chain is denoted by 0. Addition is defined as before and the result is another commutative group. We must be careful here: the addition operation, defined this way, has nothing to do with the operation of addition in the vector space $\mathbb{R}^d$. The following example illustrates this fact.

Example 11.7.3. For the 0-chains C_1 and C_2 in $\mathbb{R}^1$ defined by $C_1 = 1\{2\} + 3\{3.5\} - 2\{0\}$ and $C_2 = 1\{4.9\} + 4\{0\} - 3\{3.5\}$, find $C_1 + C_2$ and simplify it as much as possible.

Solution: We have

$$
\begin{aligned}
C_1 + C_2 &= 1\{2\} + 2\{3.5\} - 2\{0\} + 1\{4.9\} + 4\{0\} - 2\{3.5\} \\
&= 1\{2\} + (2\{3.5\} - 2\{3.5\}) + (-2\{0\} + 4\{0\}) + 1\{4.9\} \\
&= 1\{2\} + (2 - 2)\{3.5\} + (-2 + 4)\{0\} + \{4.9\} \\
&= \{2\} + 0\{3.5\} + 2\{0\} + 1\{4.9\} = 1\{2\} + 2\{0\} + 1\{4.9\}.
\end{aligned}
$$

Note this does *not* further simplify to $\{2 + 2 \cdot 0 + 4.9\} = \{6.9\}$. In the group of 0-chains in $\mathbb{R}^1$ it is not true that $2\{0\} = 0$ or that $1\{2\} + 1\{4.9\} = 1\{6.9\}$. We will however, commonly drop the coefficient 1 in front of a path or a singleton point. Then the result of the above computation becomes $\{2\} + 2\{0\} + \{4.9\}$.

Note that if 1-chains are replaced by 0-chains in Definition 11.7.2 we have a notion of *reduced form* for 0-chains. Each 0-chain can be put in reduced form. Once it is expressed in reduced form, the trace of a 0-chain is just the union of the points of $\mathbb{R}^d$ that appear in this expression. Note that in the previous example, the last expression in the series of equalities is an expression for $C_1 + C_2$ in reduced form.

Definition 11.7.4. We define a map ∂ from 1-chains in $\mathbb{R}^d$ to 0–chains in $\mathbb{R}^d$ by

$$
\partial \left(\sum_{j=1}^{p} m_j \gamma_j \right) = \sum_{j=1}^{p} (m_j \{\gamma_j(1)\} - m_j \{\gamma_j(0)\}).
$$

A 1-chain Γ in U is called a *1-cycle* if $\partial \Gamma = 0$.

The map ∂ from 1-chains to 0–chains is a group homomorphism. This means that, for any two 1-chains Γ and Λ. $\partial(\Gamma + \Lambda) = \partial\Gamma + \partial\Lambda$.

A smooth path with parameter interval $[0, 1]$ is, itself, a 1-chain (a 1-chain where there is only one summand and its coefficient is 1). Also, as mentioned earlier, a path γ which is not smooth can also be used to produce a 1-chain Γ by breaking the path up into smooth pieces and reparameterizing the pieces so that they have $[0, 1]$ as parameter interval. If this is done, then it turns out that γ is a closed path if and only if $\partial \Gamma = 0$ (Exercise 11.7.13).
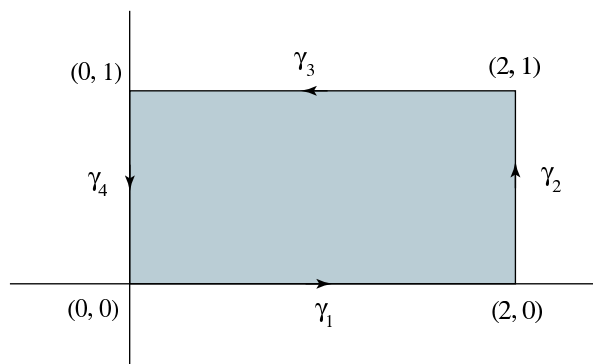
Figure 11.8: Boundary of a Rectangle as a Cycle

Example 11.7.5. Consider the rectangle R in $\mathbb{R}^2$ with vertices the points $(0,0)$, $(2,0)$, $(2,1)$, and $(0,1)$ (Figure 11.8). Represent its boundary as a cycle.

Solution: We set $\gamma_1(t) = (2t,0)$, $\gamma_2(t) = (2,t)$, $\gamma_3(t) = (2-2t,1)$, and $\gamma_4(t) = (0,1-t)$. Then $\Gamma = \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$. Note that $\Gamma(I) = \partial R$ and

$$
\begin{aligned}
\partial\Gamma &= \gamma_1(1) - \gamma_1(0) + \gamma_2(1) - \gamma_2(0) + \gamma_3(1) - \gamma_3(0) + \gamma_3(1) - \gamma_3(0) \\
&= \{(2,0)\} - \{(0,0)\} + \{(2,1)\} - \{(2,0)\} + \{(0,1)\} - \{(2,1)\} \\
&\quad + \{(0,0)\} - \{(0,1)\} = 0.
\end{aligned}
$$

and so Γ is a cycle.

Note, we could also represent the boundary of R as a single path which joins together the smooth paths $\gamma_1, \gamma_2, \gamma_3$, and γ_4. As we shall see below, for the purposes of integration, the two ways of representing the boundary of R are equivalent.

The boundary of a reasonably nice bounded subset of the plane may be represented as the union of a number of smooth curves. The rectangle in Figure 11.8 is one such set. When this is true, we would like to represent the boundary by a certain cycle. In Figure 11.8 this was the cycle of the previous example. The next example describes another such situation.

Example 11.7.6. In Figure 11.9 , the region S in the plane consists of points inside the large circle but outside the union of the two smaller circles. Represent ∂S by a cycle.

Solution: Smooth curves which trace each of the three circles are:

$$
\begin{aligned}
\gamma_1(t) &= (4\cos(2\pi t), 4\sin(2\pi t)), \\
\gamma_2(t) &= (2 + \cos(2\pi t), \sin(2\pi t)), \\
\gamma_3(t) &= (-2 + \cos(2\pi t), \sin(2\pi t)).
\end{aligned}
$$

Each circle is traced once in the counterclockwise direction by the corresponding curve. We represent the boundary ∂S of S by the cycle $\Gamma = \gamma_1 - \gamma_2 - \gamma_3$.
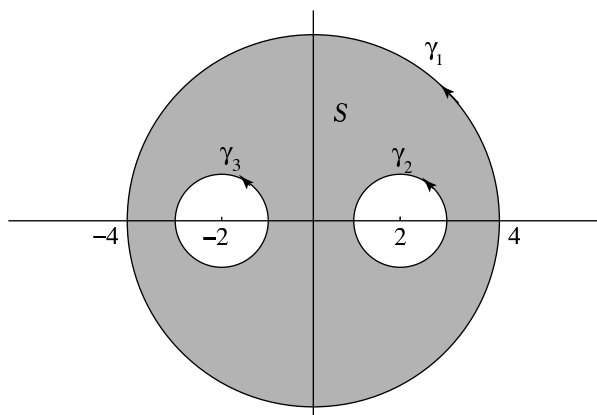
Figure 11.9: Boundary of S as the cycle $\gamma_1 - \gamma_2 - \gamma_3$

Why do we choose to multiply γ_1 and γ_2 by -1 in the sum defining Γ? It is due to the following: while the circle γ_1 has positive orientation relative to S, the circles γ_2 and γ_3 have negative orientation relative to S and multiplying by -1 compensates for this. For the meaning of this statement see the discussion on orientation of paths in Section 11.4.

p-Chains and p-Cycles

For any non-negative integer p, we will define a p-chain in $\mathbb{R}^d$ to be a formal linear combination of p-cells in $\mathbb{R}^d$. First we need to define what we mean by a p-cell in $\mathbb{R}^d$. We have defined 2-cells and 3-cells in previous sections. A 1-cell in $\mathbb{R}^d$ will be a smooth path in $\mathbb{R}^d$ parameterized on $I = [0, 1]$. A 0-cell is just a singleton set $\{x\}$ in $\mathbb{R}^d$.

Definition 11.7.7. We define d-cells just as we defined 2-cells and 3-cells. A d-cell in R^p is a smooth function $E : I^d \to \mathbb{R}^p$. A d-cell is simple if it is one-to-one with non-singular differential on the interior of I^d.

As before, in defining for such a function to be smooth on the compact set I^d, on the boundary some partial derivatives must be interpreted as one-sided derivatives .

Definition 11.7.8. A p-chain C in $\mathbb{R}^d$ is a formal linear combination

$$C = \sum_{J=1}^{n} m_j E_j \tag{11.7.2}$$

of p-cells with integer coefficients.

As we did with 1-chains, we agree that if the individual terms $m_j E_j$ in a chain are rearranged, so that they appear in a different order, then the chain does not change. We agree that the chain does not change if we drop summands $m_j E_j$ with $m_j = 0$, and we agree that two summands $m_j E_j$ and $m_k E_k$ with $E_j = E_k$ may be combined to yield $(m_j + m_k)E_j$. The empty chain – that is the chain with no summands is denoted by 0. We add two chains in the obvious way: the sum of two formal linear combinations of p-cells is another formal linear combination of p-cells. The operation of addition, so defined, is clearly associative and commutative.

As before, the set of p-chains in $\mathbb{R}^d$, as defined above, forms a commutative group.

The expression (11.7.2) for a chain C is said to be in *reduced form* if the E_j are distinct paths and all the m_j are non-zero. Note that each chain may be expressed in reduced form. We define the *trace* of a chain C to be the union of the traces of the E_j in an expression of the chain in reduced form.

Boundaries

If $E : I^p \to \mathbb{R}^d$ is a continuous function, then for $j = 1, \cdots, p$ we consider the $2p$ functions of $p - 1$ variables defined by

$$E^{j0} = E(x_1, \cdots, x_{j-1}, 0, x_j, \cdots, x_p) \text{ and}$$
$$E^{j1} = E(x_1, \cdots, x_{j-1}, 1, x_j, \cdots, x_p)$$

Each of these is a continuous function from I^{p-1} to $\mathbb{R}^d$. We will call these the $p - 1$ dimensional faces of E.

Definition 11.7.9. If E is a p-cell, then its boundary, ∂E, is the $p - 1$-chain defined by

$$\partial E = \sum_{i,\sigma} (-1)^{i+\sigma} E^{i\sigma}$$

Where i ranges over $1, \cdots, p$ and σ ranges over $0, 1$.

If $C = \sum_j E_j$ is a p-chain, then we define its boundary ∂C to be the $p - 1$ chain $\sum_j \partial E_j$. We say that C is a p-cycle if $\partial C = 0$.

Recall that the above definition of ∂E, is the way we defined the boundary of a 3-cell in the previous section. It is, not quite the same, but is equivalent to the way we defined the boundary of a 2-cell in section 11.4.

Theorem 11.7.10. *If C is a p-chain, then $\partial^2 C = 0$.*

Proof. It is enough to prove this in the case where C is a single cell E. Then

$$\partial^2 E = \partial(\partial E) = \sum_{i\sigma} \sum_{j\tau} (-1)^{i+j+\sigma+\tau} (E^{i\sigma})^{j\tau}.$$

Note that, if $i \leq j$, then $(E^{i\sigma})^{j\tau} = (E^{j+1\sigma})^{i\tau}$. Since these two terms appear with opposite signs in the above sum, they cancel each other out. But every term in the above sum is of one of these two types. Hence, the sum is 0. $\qquad\square$

The previous theorem tells us that the boundary of a chain is always a cycle. In particular, the boundary of a p-cell is a $p-1$-cycle.

Example 11.7.11. Express the solid square of Example 11.7.5 and Figure 11.8 as the trace of a 2-cell E and calculate ∂E.

    Solution: We set $E(s,t) = (2s,t)$ for $(s,t) \in I^2$. This has the rectangle of Figure 11.8 as trace. By Definition 11.7.9,

$$\partial E = E^{20} + E^{11} - E^{21} - E^{10}$$

where, in terms of the paths γ_j of Example 11.7.5,

$$\begin{aligned}
E^{20}(s) &= E(s,0) = (2s,0) = \gamma_1(s) \\
E^{11}(s) &= E(1,s) = (1,s) = \gamma_2(s) \\
E^{21}(s) &= E(s,1) = (2s,1) = \gamma_3(1-s) \\
E^{10}(s) &= E(0,s) = (0,s) = \gamma_4(1-s).
\end{aligned}$$

Note that E^{21} and $E^{10}(s)$ are γ_3 and γ_4 with orientation reversed. This is why they each occur with a factor of (-1) in the cycle ∂E. This compensates for the orientation reversal when we do integration over ∂E and ensures that, for the purposes of integration, the cycle ∂E and the cycle $\Gamma = \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$ are equivalent.

Integration Over Chains and Cycles

Chains exist so that we may integrate over them. We define the integral of a p-form over a p-chain below. First we define the integral of a p-form over a p-cell. This is no different than the definitions for integrals of forms in dimensions $1, 2$ or 3. We use the transformation law for how a p form ϕ in $\mathbb{R}^q$ transforms to a p-form $E^*(\phi)$ on I^p under a cell $E : I^p \to \mathbb{R}^q$. This is defined exactly as in Definition 11.3.9.

Definition 11.7.12. If $E : I^p \to \mathbb{R}^q$ is a p-cell and ϕ a p-form defined on a set containing $E(I^p)$, then we define the integral of ϕ over E by

$$\int_E \phi = \int_{I^p} E^*(\phi).$$

    We define the integral of a p- form over a p-chain as follows:

Definition 11.7.13. Let

$$C = \sum_{j=1}^p m_j E_j$$

be a p-chain in $\mathbb{R}^d$, expressed in reduced form. If ϕ is a p-form defined on the trace of C, then we set

$$\int_C \phi = \sum_{j=1}^p m_j \int_{E_j} \phi. \tag{11.7.3}$$

It is a consequence of this definition that if C_1 and C_2 are two p-chains and ϕ is a p-form defined and continuous on a set containing both the trace of C_1 and the trace of C_2, then

$$\int_{C_1 + \mathbb{C}_2} \phi = \int_{C_1} \phi + \int_{C_2} \phi \qquad (11.7.4)$$

The proof of this fact is left to the exercises (Exercise 11.7 .12).

Definition 11.7.14. Suppose C_1 and C_2 are cycles in $\mathbb{R}^d$. We will say that C_1 and C_2 are *equivalent* if they have the same trace and

$$\int_{C_1} \phi = \int_{C_2} \phi$$

for every p-form ϕ on the trace of C_1.

In general, a p-cell E is equivalent to any p-cell F for which there is a is a positively oriented smooth parameter change P such that $F = E \circ P$. If P is negatively oriented, then the chain $(-1)E$ is equivalent to F. In the case of a 1-cell γ (a smooth path) this is illustrated by the fact that $(-1)\gamma$ is equivalent to $^-\gamma$, the path γ traversed in the reverse direction (see Exercise 11.1.8).

Example 11.7.15. Show that if γ_1 and γ_2 are two paths with parameter interval $I = [0,1]$, and if $\gamma_1(1) = \gamma_2(0)$ (so that γ_2 starts where γ_1 ends), then the chain $\Gamma = \gamma_1 + \gamma_2$ is equivalent to the chain consisting of the single path γ which is γ_1 and γ_2 spliced together, that is

$$\gamma(t) = \begin{cases} \gamma_1(2t) & \text{if} \quad 0 \le t \le 1/2 \\ \gamma_2(2t-1) & \text{if} \quad 1/2 \le t \le 1 \end{cases} .$$

Solution: Note that $\gamma(I) = \gamma_1(I) \cup \gamma_2(I) = \Gamma(I)$. On $[0, 1/2]$ γ is obtained from γ_1 by a smooth parameter change $t \to 2t$, while on $[1/2, 1]$ γ is obtained from γ_2 by the smooth parameter change $t \to 2t - 1$. Thus, for any 1-form on the trace of γ,

$$\int_\gamma \phi = \int_0^1 \phi(\gamma(t))\gamma'(t)\,dt = \int_0^{1/2} \phi(\gamma(t))\gamma'(t)\,dt + \int_{1/2}^1 \phi(\gamma(t))\gamma'(t)\,dt$$

$$= \int_{\gamma_1} \phi + \int_{\gamma_2} \phi = \int_\Gamma \phi$$

and, hence, γ and Γ are equivalent chains.

The General Stokes Theorem

Theorem 11.7.16. *If ϕ is a smooth $p - 1$-form defined on I^p, then*

$$\int_{\partial I^p} \phi = \int_{I^p} d\phi.$$

We won't go through the proof here. It is very much like the proof of the $p = 3$ version of the theorem, which was proved earlier (Theorem 11.6.3). It is a simple application of the Fundamental Theorem of Calculus.

This leads us to the general version of Stokes's Theorem.

Theorem 11.7.17. *Let C be a p-chain in $\mathbb{R}^q$ and ϕ a smooth $p-1$-form defined on the trace of C. Then*

$$\int_{\partial C} \phi = \int_C d\phi \qquad (11.7.5)$$

Proof. If C is a single cell E, then this follows, as with earlier versions, from the previous theorem and the identities

$$\int_{\partial E} \phi = \int_{\partial I^p} E^*(\phi) \quad \text{and} \quad \int_E d\phi = \int_{I^p} dE^*(\phi).$$

The proof for general chains now follows from the fact that both sides of (11.7.5) are linear in C. That is, if C is a certain linear combination of cells E_j, then the integrals in the formula are the corresponding linear combinations of the integrals with C replaced by E_j. ☐

If C is a single cell, then the above theorem is Green's Theorem when $p = 2$ and $q = 2$, the dimension 2 Stokes's Theorem when $p = 2$ and $q > 2$, Gauss's Theorem when $p = 3$ and $q = 3$, and the dimension 3 Stokes's Theorem when $p = 3$ and $q > 3$. In the case where $p = 1, q = 1$ it is the Fundamental Theorem of Calculus, and when $p = 1, q > 1$ it is the Fundamental Theorem of Calculus for path integrals.

The following are simple corollaries of the general Stokes Theorem. The proofs are left to the exercises.

Corollary 11.7.18. *If C is a p-cycle and ϕ is a smooth $p-1$ form on the trace of C, then*

$$\int_C d\phi = 0.$$

Corollary 11.7.19. *If C is a p-chain and ϕ is a smooth closed p-form on the trace of C, then*

$$\int_{\partial C} \phi = 0.$$

Exercise Set 11.7

1. If E_1, E_2, and E_3 are three distinct p-cells in $\mathbb{R}^d$, express the sum of the chains $2E_1 + E_2 - 3E_3$ and $-5E_1 - E_2 + E_3$ in reduced form.

2. Express the sum of the 0-chains $C_1 = 2\{-3\} - 4\{1\} + \{2\}$ and $\mathbb{C}_2 = 3\{1\} - \{2\}$ in reduced form.

3. For $t \in [0, 1]$ let $\gamma_1(t) = (2t - 1, 0)$, $\gamma_2(t) = (1 - t, t)$, $\gamma_3(t) = (t - 1, t)$. Which of the following 1-chains in $\mathbb{R}^2$ is a cycle?

(a) $\gamma_1 + \gamma_2 + \gamma_3$;

(b) $\gamma_1 + \gamma_2 - \gamma_3$;

(c) $\gamma_1 + 2\gamma_2 - 3\gamma_3$.

4. Let $E(r, \theta) = (r \cos 2\pi\theta, r \sin 2\pi\theta)$ for $(r, \theta) \in I^2$. Show that E is a simple cell and explicitly describe the cycle ∂E.

5. Let Δ be the triangle in $\mathbb{R}^2$ with vertices at $(0,0)$, $(1,0)$, and $(0,2)$. Express this triangle as the trace of a 2-cell E and find the cycle ∂E.

6. Find the integral of the 1-form $2xy^3 \, dx + 3x^2y^2 \, dy$ over the 1-cycle of the previous exercise.

7. For $t \in [0,1]$, let $\gamma_1(t) = (2t - 1, 0)$, $\gamma_2(t) = (\cos(\pi t), \sin(\pi t))$, and define a 1-chain Γ in $\mathbb{R}^2$ by $\Gamma = \gamma_1 + \gamma_2$. For the 1-form $\phi(x, y) = 3x^2 \, dx + 2y \, dy$, find $\int_\Gamma \phi$.

8. Find $\int_\Gamma \psi$ if Γ is the 1-chain of the previous exercise and $\psi = x \, dy$.

9. Define 2-cells in $\mathbb{R}^2$ as follows. For $(s, t) \in I^2$,

$$E(s, t) = ((1 + s) \cos \pi t, (1 + s) \sin \pi t)$$
$$F(s, t) = ((1 + s) \cos \pi t, -(1 + s) \sin \pi t).$$

If C is the 2-chain $E - F$, then find ∂C and $\int_{\partial C} (e^{x^2} \, dx + \sin(y^2)) \, dy$.

10. In $\mathbb{R}^2$ define a smooth path $\gamma_r(t) = (r \cos(2\pi t), r \sin(2\pi t))$ for each $r > 0$. If ϕ is a 1-form on $\mathbb{R}^2 \setminus \{0\}$ such that $d\phi = 0$, then show that $\int_{\gamma_r} \phi$ is independent of r. Hint: for $0 < s < r$, consider the cycle $\Gamma = \gamma_r - \gamma_s$. Is this ∂E for some 2-cell E?

11. Let ϕ be a smooth 2-form on $\mathbb{R}^3 \setminus \{0\}$ which satisfies $d\phi = 0$. Show that $\int_{\partial E} \phi$ is the same number for all simple 3-cells E such that 0 is in the image of the interior of I^3 under E, but not in the image of ∂I^2 under E. On the other hand, if 0 is not in the trace of E at all, then this integral is 0. Hint: for the first part, show that for any such E, the integral over the boundary of E is the same as the integral over any sufficiently small hollow sphere centered at 0.

12. Prove (11.7.4).

13. Suppose γ is a path and $\Gamma = \gamma_1 + \cdots + \gamma_n$ is the 1-chain made by breaking γ up into smooth paths and reparameterizing each of them so that it has parameter interval $[0, 1]$. Show that Γ is a cycle if and only if γ is a closed path.

14. Prove that if Γ is a 1-cycle, then Γ is equivalent to a 1-cycle with all of its summands closed paths. Hint: use repeated application of the following idea: if one path begins where another one ends, then the two can be joined together to form a single path which is equivalent to the sum of the two individual paths.

15. Let ϕ be a smooth 2-form on $\mathbb{R}^3 \setminus Z$, where Z is the set of integers on the x-axis. Let γ be a positively oriented parameterization of the sphere of radius $n + 1/2$ centered at the origin, and for $j = -n, \cdots n$ let γ_j be a positively oriented parameterization of the sphere of radius $1/3$ centered at j. Then let Γ be the cycle $\Gamma = \sum_{j=-n}^{n} \gamma_j$. If $d\phi = 0$ show that

$$\int_\gamma \phi = \int_\Gamma \phi.$$