

Math 521 – Spring 2010

521

May 12, 2010

Contents

1	Pep Talk	2
2	Notation	5
3	Logic	6
4	Sets	8
5	Functions and Maps	11
6	Composition and Inverse	13
7	Mathematical Induction	15
8	Infinite Sets	16
9	Axioms for the Real Numbers	18
10	Distance	22
11	Limits	24
12	Convergence of Sequences	26
13	Continuity	32
14	Open Sets and Closed Sets	34

15 Connected Sets	37
16 Compact Sets	39
17 Derivatives	43
18 The Integral	44
19 Taylor's Formula	49
20 Series	50
21 Uniform Convergence	56
22 Power Series - I	59
23 Analytic Functions	61
24 Power Series - II	64
25 The Heat Equation	66
26 Fourier Series - I	68
27 Fourier Series - II	68
A Uniqueness of the Real Numbers	68
B Additional Problems	71

1 Pep Talk

What does not kill me, makes me stronger.

Friedrich Nietzsche (1844 - 1900)

in The Twilight of the Idols (1899)

It is generally acknowledged in the math department that the two hardest undergraduate courses we teach are 222 and 521. The former is hard because

for many students it is their first college math course, and although they did well in high school, they are unprepared for what is expected in college. The latter is hard for many reasons. For many students it is the first course where they are expected to understand and write proofs. (See 3.4.) There are a lot of hard words like 'if', 'for all', 'there exists', (see Chapter 3) and the order in which they appear is crucial (see 13.3). Inequalities (see 9.2) play a crucial role. The language of set theory (see Chapter 4) is used heavily and some students have a hard time visualizing what set is. On top of all this there is some strange lingo like 'connected', 'compact', 'open cover', etc. The first time you think you understand and discover that I disagree you may well decide that it is easier to get hit by a truck.

The subject *is* hard. Calculus was invented by Leibniz and Newton in the seventeenth century, but the ideas presented in these notes were not fully developed until the early twentieth century. Some of the best mathematical minds in human history contributed to the subject. But you have one advantage over these luminaries: a teacher who understands the subject and wants to help you understand.

Given the task before you, why should you make the effort? The above quote of Nietzsche may inspire you. The ability to reason carefully will serve you well in any future endeavor. You may be driven (like I am) to understand the patterns in mathematical reasoning. These patterns help us see the similarities between apparently different problems and enable us to use our understanding of one problem to solve the others. Finally, you may actually need to use the material someday. Many mathematical problems have no explicit solution: we must prove that there is a solution and use that proof both to guide us to an approximate solution and to estimate the accuracy of that solution. If we get to the end of these notes in this course you will get a glimpse of how this works when we solve the heat equation in problem 25.2.

Here are some tips:

1. Read. It is a bad idea to try to learn just from lectures. A really good lecturer can make the material look easy, but just listening is rarely enough to lead to true understanding.
2. Focus on definitions. It is impossible to understand why a compact set is closed and bounded if you don't know what the words mean. Use the index and cross references in these notes or the text you are

reading to recall definitions and previous theorems. If you are reading these notes on a computer and have a sufficiently recent version of Adobe Acrobat Reader or some other pdf viewer, you will notice that this document contains embedded hypertext links. Some viewers even allow text searches and links to URLs. (If you are online try clicking on the link to the official description below.) This makes the task of jumping around the text much easier.

3. Be active not passive. As you read or listen, make up examples to test your understanding. When you see a definition make up an example of something which satisfies it and something which doesn't.
4. Ask questions. If you can formulate a question when you are confused, you may discover that you become unconfused and don't need to ask the question. But don't hesitate to ask me questions in class or office hours.
5. Be concise when you write. Excess verbiage is hard to follow and may conceal confusion.
6. Don't fall behind. If you postpone understanding till the night before the exam, you may find that you are still lost. For each lecture there will be a short prequiz on **Moodle** (see below) designed to get you to read the material before I lecture on it. In addition there will be a short post quiz almost every day (I hope).

We will use an online course management system called **Moodle**. To use it go to <https://www.math.wisc.edu/moodle> and click on Math 521. You will be prompted to authenticate. Use your net id and net password the same as if you were logging on the MyUW or WiscMail. If you are not enrolled in the course (according to the registrar) you will not be allowed in.

The text references in these notes are to the following texts:

- (**Buck**) R.Creighton Buck, *Advanced Calculus*, 3rd edition, Waveland Press.
- (**Lang**) Serge Lang, *Undergraduate Analysis*, 2nd edition, Springer Undergraduate Texts in Mathematics.
- (**Morgan**) Frank Morgan, *Real Analysis*, American Mathematical Society

(**Rudin**) Walter Rudin, *Principles of Mathematical Analysis*, McGraw Hill.

There are three sections of 521 this semester and each has chosen one of the first three. The instructor I am replacing chose Buck. I used this book the last time I taught 521. The terminology, content, even the title is are somewhat dated. (The math department will soon change the name of 521 from *Advanced Calculus* to *Real Analysis I*.) Lang is good but somewhat ponderous. It contains more material than can be comfortably treated in a two semester course. My favorite is from the first three is Morgan. It is student friendly and contains almost exactly the material I hope to cover. (See the official description of 521 at <http://www.math.wisc.edu/521-advanced-calculus>.) It is also the least expensive. The Rudin text is the best for someone who already thinks like a mathematician, but is difficult for the apprentice.

I have arranged these notes so you can use them with any of these texts – there are remarks and footnotes which explain slight differences in terminology – but the notes are designed with the Morgan text in mind. Possibly you won't even need to buy a text: all three texts are on reserve in the Math Library. The prequizzes will only require you to read the course notes.

2 Notation

The following standard notations are used:¹

$\mathbb{Z}^+$	the set of positive integers (natural numbers).
$\mathbb{N}$	the set of nonnegative integers.
$\mathbb{Z}$	the set of integers.
$\mathbb{Q}$	the set of rational numbers.
$\mathbb{R}$	the set of real numbers.
$\mathbb{C}$	the set of complex numbers.

The usual **interval notation** is used, e.g.

$$[a, b) := \{x \in \mathbb{R} : a \leq x < b\}, \quad (-\infty, b] := \{x \in \mathbb{R} : x \leq b\}, \quad \text{etc.}$$

We write $:=$ to indicate that two objects are equal by definition. We also signal definitions by writing *iff* instead of *if and only if*.

¹ Morgan uses $\mathbb{N}$ denote the positive integers while Lang uses $\mathbb{N}$ to denote the nonnegative integers.

3 Logic

3.1. In these notes we shall often use the abbreviations

$\implies$	for	“implies”,
$\iff$	for	“if and only if”,
$\exists$	for	“there exists”,
$\forall$	for	“for all”.

These abbreviations help clarify the logical structure of the definitions and theorems. They usually aren’t used in textbooks but are commonly used in lectures.

3.2. The **logical operations** are defined by the following table:

P	Q	P and Q	P or Q	P $\implies$ Q	P $\iff$ Q	not P
T	T	T	T	T	T	F
T	F	F	T	F	F	
F	T	F	T	T	F	T
F	F	F	F	T	T	

It is particularly important to notice that the fifth column asserts that ‘*false implies anything*’ is true. Note also that the use of the word *or* is inclusive as in ‘*either $x > 0$ or else $x < 5$* ’ (true) not exclusive as in ‘*soup or salad?*’ (not both).

3.3. The following are synonymous:

- $P \implies Q$.
- P implies Q.
- If P, then Q.
- If not Q, then not P.
- Q, if P.
- P only if Q.

Do not confuse the statement ‘*if Q , then P* ’ with its **converse** ‘*if P , then Q* ’: it can happen that one is true and the other is not as in $x = 2 \implies x^2 = 4$. The statement ‘*If not Q , then not P* ’ is called the **contrapositive** of the statement ‘*if P , then Q* ’. A statement and its contrapositive are equivalent. This provides the justification for **proof by contradiction**.

3.4. One of the principal aims of this course is to teach you how to read and write proofs. A **proof** is an argument intended to convince the reader that a general principle is true in all situations. The amount of detail that an author supplies in a proof should depend on the audience. Too little detail leaves the reader in doubt; too much detail may leave the reader unable to see the forest for the trees. As a general principle, the author of a proof should be able to supply the reader with additional detail on demand. When a student writes a proof for a teacher, the aim is usually not to convince the teacher of the truth of some general principle (the teacher already knows that), but to convince the teacher that the student understands the proof and can write it clearly.

The “theorems” below show the proper format for writing a proof. In each of them you are supposed to imagine that the theorem to be proved has the indicated form. Notice how the key words *choose*, *assume*, *let*, and *therefore* are used in the proof. In these sample formats, the phrase “Blah Blah Blah” indicates a sequence of steps, each one justified by earlier steps.

Theorem *If P , then Q .*

Proof. Assume P . Blah Blah Blah. Therefore Q . □

Theorem *P if and only if Q .*

Proof. Assume P . Blah Blah Blah. Therefore Q . Conversely, assume Q . Blah Blah Blah. Therefore P . □

Theorem *$P(x)$ for all x .*

Proof. Choose x . Blah Blah Blah. Therefore $P(x)$. □

Theorem *There is an x such that $P(x)$.*

Proof. Let $x = \dots$. Blah Blah Blah. Therefore $P(x)$. □

4 Sets

4.1. A set A divides the mathematical universe into two parts: those objects x that *belong* to A and those that don't. The notation $x \in A$ means x belongs to A . The notation $x \notin A$ means that x does not belong to A . The objects that belong to A are sometimes called the *elements* of A but we will often call them *points* or *numbers*. Other words roughly synonymous with the word *set* are *class*, *collection*, and *aggregate*. These longer words are generally used to avoid using the word *set* twice in one sentence. The situation typically arises when an author wants to talk about sets whose elements are themselves sets. One might write “the collection of all finite sets of integers”, rather than “the set of all finite sets of integers”.

4.2. For two sets A and B , the notation $A \subseteq B$ means that A is a **subset** of B , i.e. for all x we have $x \in A \implies x \in B$. By definition, two sets are **equal** if each is a subset of the other:

$$A = B \iff A \subseteq B \text{ and } B \subseteq A.$$

The notation $\{x : P(x)\}$ denotes the set of all x for which the property $P(x)$ is true. The notation $\{x \in A : P(x)\}$ denotes the set of all $x \in A$ for which the property $P(x)$ is true. Finite sets may be defined by enumerating their elements as in

$$x \in \{a_1, a_2, \dots, a_n\} \iff x = a_1 \text{ or } x = a_2 \text{ or } \dots \text{ or } x = a_n$$

and often infinite sets as well as in

$$\mathbb{N} = \{0, 1, 2, \dots\}, \quad \mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}, \quad \mathbb{Z}^+ = \{1, 2, 3, \dots\}.$$

4.3. If A and B are sets, then the sets

$$A \cup B := \{x : x \in A \text{ or } x \in B\}, \quad A \cap B := \{x : x \in A \text{ and } x \in B\},$$

are called respectively the **union** and **intersection** of A and B . The **empty set** is denoted $\emptyset$: For all x it is true that

$$x \notin \emptyset.$$

Two sets are **disjoint** iff they have no elements in common, i.e. iff $A \cap B = \emptyset$. The set

$$X \setminus A := \{x \in X : x \notin A\}$$

is called the **complement** of A in X .² The set

$$A \times B := \{(x, y) : x \in A \text{ and } y \in B\}$$

of all ordered pairs (x, y) with $x \in A$ and $y \in B$ is called the **Cartesian product** of A and B . The term **direct product** is a synonym. We also use the notation

$$A^n := \underbrace{A \times A \times \cdots \times A}_n$$

In particular,

$$\mathbb{R}^n := \{(x_1, x_2, \dots, x_n) : x_i \in \mathbb{R}\}$$

denotes the vector space of all n tuples of real numbers, so $\mathbb{R}^1 = \mathbb{R}$, $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, $\mathbb{R}^3 = \mathbb{R} \times \mathbb{R} \times \mathbb{R}$, etc.³

Remark 4.4. Do not confuse an **n tuple** (finite sequence of length n) with a finite set. For the former order is important: $\{3, 7\} = \{7, 3\}$ but $(3, 7) \neq (7, 3)$; for the latter repetitions don't matter: $\{2, 2, 3\} = \{2, 3\}$ but $(2, 2, 3) \neq (2, 3)$.

4.5. An **indexed family of sets** is a function which assigns a set A_i to each element i of a set I . The set I is called the **index set** of the family and the family is usually denoted $(A_i)_{i \in I}$. The **union** and **intersection** of the indexed family are defined by

$$\begin{aligned} x \in \bigcup_{i \in I} A_i &\iff \exists i \in I \text{ such that } x \in A_i. \\ x \in \bigcap_{i \in I} A_i &\iff \forall i \in I \text{ we have } x \in A_i. \end{aligned}$$

The notation $\exists i \in I$ is an abbreviation for “there exists $i \in I$ ”, and the notation $\forall i \in I$ is an abbreviation for “for all $i \in I$ ”. In these definitions the set I can be infinite. For finite sets I we recover the earlier definitions, e.g. for $I = \{1, 2\}$ we have

$$\bigcup_{i \in \{1, 2\}} A_i = A_1 \cup A_2, \quad \bigcap_{i \in \{1, 2\}} A_i = A_1 \cap A_2.$$

This illustrates the logical principle that $\exists$ is like an “infinite or” and $\forall$ is like an “infinite and”.

² Morgan uses the notation A^c for $\mathbb{R}^n \setminus A$ (when A is a subset of $\mathbb{R}^n$).

³ Buck uses the term **n space** as a synonym for $\mathbb{R}^n$.

Remark 4.6. Set theory is simply a way of formalizing logic. Simple set theoretic identities may be proved by **truth tables**. For example, consider the following “distributive law”

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

To show that $x \in (A \cup B) \cap C \iff x \in (A \cap C) \cup (B \cap C)$ we can simply consider all the possibilities:

A	B	C	$(A \cup B) \cap C$	$(A \cap C) \cup (B \cap C)$
T	T	T	T	T
T	T	F	F	F
T	F	T	T	T
T	F	F	F	F
F	T	T	T	T
F	T	F	F	F
F	F	T	F	F
F	F	F	F	F

It is usually not necessary to show so much detail in your written work, but it will be hard for you to decide just how much detail *is* appropriate. A good rule of thumb is that you should be prepared to supply more detail if challenged.

4.7. From logic we know that

$$\text{not } \exists \iff \forall \text{ not,} \quad \text{and} \quad \text{not } \forall \iff \exists \text{ not,}$$

so for any set X and any indexed family $(A_i)_{i \in I}$ we have

$$X \setminus \bigcup_{i \in I} A_i = \bigcap_{i \in I} (X \setminus A_i), \quad X \setminus \bigcap_{i \in I} A_i = \bigcup_{i \in I} (X \setminus A_i).$$

Logicians call these **De Morgan’s Laws**. In particular, for $I = \{1, 2\}$ we have

$$\begin{aligned} X \setminus (A_1 \cup A_2) &= (X \setminus A_1) \cap (X \setminus A_2), \\ X \setminus (A_1 \cap A_2) &= (X \setminus A_1) \cup (X \setminus A_2). \end{aligned}$$

Also by logic we have set theoretic distributive laws

$$X \cap \bigcup_{i \in I} A_i = \bigcup_{i \in I} (X \cap A_i), \quad X \cup \bigcap_{i \in I} A_i = \bigcap_{i \in I} (X \cup A_i).$$

In particular, for $I = \{1, 2\}$ we have

$$\begin{aligned} X \cup (A_1 \cap A_2) &= (X \cup A_1) \cap (X \cup A_2), \\ X \cap (A_1 \cup A_2) &= (X \cap A_1) \cup (X \cap A_2). \end{aligned}$$

The latter was proved above in Remark 4.6.

5 Functions and Maps

5.1. A **function** is a rule which assigns a value $f(x)$ to every point x from a set called the **domain** of the function. The set

$$\text{graph}(f) := \{(x, y) : y = f(x)\}$$

of all pairs (x, y) such that $y = f(x)$ is called the **graph** of the function f . Two functions are equal iff they have the same graph.

5.2. Let X and Y be sets. We say that f is a **map** from X to Y and write $f : X \rightarrow Y$ when f is a function which assigns a point $y = f(x) \in Y$ to each point $x \in X$. Two maps $f : X \rightarrow Y$ and $f' : X' \rightarrow Y'$ are said to be **equal** when $X = X'$, $Y = Y'$, and $f(x) = f'(x)$ for all $x \in X$. Thus if f and f' are equal maps, then $\text{graph}(f) = \text{graph}(f')$ but not conversely (because $Y = Y'$ is part of the definition of equality for maps). However most authors would say that two functions are equal iff they have the same graph.

Remark 5.3. Some authors use the notation $x \mapsto f(x)$ to define a map. This allows them to avoid introducing a name for the map. Thus instead of writing

$$\text{Consider the map } f : \mathbb{R} \rightarrow \mathbb{R} \text{ defined by } f(x) = x^5 + x.$$

they may write

$$\text{Consider the map } \mathbb{R} \rightarrow \mathbb{R} : x \mapsto x^5 + x.$$

5.4. When $A \subseteq X$, $B \subseteq Y$, and $f : X \rightarrow Y$, the sets

$$\begin{aligned} f(A) &:= \{y \in Y : \exists x \in A \text{ such that } y = f(x)\}, \\ f^{-1}(B) &:= \{x \in X : f(x) \in B\}, \end{aligned}$$

are called respectively the **image** of A by f and **inverse image** of B by f .

Remark 5.5. The sets X and Y are sometimes called the **source** and **target** of a map $f : X \rightarrow Y$. The image $f(X)$ of the source is what is called the **range** of the function f in calculus. Thus the domain of a map is the same as its source while the range is a subset of its target.

Remark 5.6. There are slight variations in terminology among authors. Morgan avoids the word *map* and Lang sometimes uses the word *mapping*. Buck uses the term the **preimage** instead of *inverse image*. Lang (page 4) uses the $\mapsto$ notation but the other authors apparently avoid it. For Lang a function is a map whose target is $\mathbb{R}$. In precalculus courses a function is usually defined by an expression and the domain is implicitly taken to be the largest set of numbers for which the expression is meaningful, but in advanced mathematics authors usually make the domain explicit. Examples 6.7 below shows why the distinction is important.

Problem 5.7. Let $f : X \rightarrow Y$, $(A_i)_{i \in I}$ be a family of subsets of X , and $(B_i)_{i \in I}$ be a family of subsets of Y . Which of the following are (always) true? Hint: When there is a counterexample, there is a counterexample with the index set $I = \{1, 2\}$.

$$\begin{array}{ll} f^{-1} \left(\bigcup_{i \in I} B_i \right) \subseteq \bigcup_{i \in I} f^{-1}(B_i)? & \bigcup_{i \in I} f^{-1}(B_i) \subseteq f^{-1} \left(\bigcup_{i \in I} B_i \right)? \\ f^{-1} \left(\bigcap_{i \in I} B_i \right) \subseteq \bigcap_{i \in I} f^{-1}(B_i)? & \bigcap_{i \in I} f^{-1}(B_i) \subseteq f^{-1} \left(\bigcap_{i \in I} B_i \right)? \\ f \left(\bigcup_{i \in I} A_i \right) \subseteq \bigcup_{i \in I} f(A_i)? & \bigcup_{i \in I} f(A_i) \subseteq f \left(\bigcup_{i \in I} A_i \right)? \\ f \left(\bigcap_{i \in I} A_i \right) \subseteq \bigcap_{i \in I} f(A_i)? & \bigcap_{i \in I} f(A_i) \subseteq f \left(\bigcap_{i \in I} A_i \right)? \end{array}$$

Problem 5.8. Proof or counterexample:

- (1) If $f : X \rightarrow Y$ and $A_1, A_2 \subseteq X$, then $f(A_1 \setminus A_2) \subseteq f(A_1) \setminus f(A_2)$?
- (2) If $f : X \rightarrow Y$ and $A_1, A_2 \subseteq X$, then $f(A_1) \setminus f(A_2) \subseteq f(A_1 \setminus A_2)$?
- (3) If $f : X \rightarrow Y$ and $B_1, B_2 \subseteq Y$, then $f^{-1}(B_1 \setminus B_2) \subseteq f^{-1}(B_1) \setminus f^{-1}(B_2)$?

(4) If $f : X \rightarrow Y$ and $B_1, B_2 \subseteq Y$, then $f^{-1}(B_1) \setminus f^{-1}(B_2) \subseteq f^{-1}(B_1 \setminus B_2)$?

Problem 5.9. Proof or counterexample:

(1) If $f : X \rightarrow Y$ and $A \subseteq X$, then $f^{-1}(f(A)) \subseteq A$?

(2) If $f : X \rightarrow Y$ and $A \subseteq X$, then $A \subseteq f^{-1}(f(A))$?

(3) If $f : X \rightarrow Y$ and $B \subseteq Y$, then $f(f^{-1}(B)) \subseteq B$?

(4) If $f : X \rightarrow Y$ and $B \subseteq Y$, then $B \subseteq f(f^{-1}(B))$?

6 Composition and Inverse

6.1. If $f : X \rightarrow Y$ and $g : Y \rightarrow Z$, then the **composition** of f and g is the map $g \circ f : X \rightarrow Z$ defined by

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

for $x \in X$. For any set X the **identity map** of X is the map $\text{id}_X : X \rightarrow X$ defined by

$$\text{id}_X(x) = x$$

for $x \in X$. Clearly

$$\text{id}_Y \circ f = f \text{ and } f \circ \text{id}_X = f$$

for $f : X \rightarrow Y$.

6.2. A map $g : Y \rightarrow X$ is said to be a **left inverse** for the map $f : X \rightarrow Y$ iff $g \circ f = \text{id}_X$, i.e. iff $g(f(x)) = x$ for all $x \in X$. A map $g : Y \rightarrow X$ is said to be a **right inverse** for the map $f : X \rightarrow Y$ iff $f \circ g = \text{id}_Y$, i.e. iff $f(g(y)) = y$ for all $y \in Y$. A map $g : Y \rightarrow X$ is said to be a (two sided) **inverse** to the map $f : X \rightarrow Y$ iff it is both a left inverse and a right inverse to f . If g is a left inverse to f and g' is a right inverse to f then $g = g'$. (Proof: $g = g \circ \text{id}_X = g \circ f \circ g' = \text{id}_Y \circ g' = g'$.) In this case there is a unique inverse and it is denoted f^{-1} . So if $f : X \rightarrow Y$ has an inverse $f^{-1} : Y \rightarrow X$, then

$$y = f(x) \iff x = f^{-1}(y)$$

for $x \in X$ and $y \in Y$.

Definition 6.3. A map $f : X \rightarrow Y$ is said to be **injective** iff

$$\forall x_1, x_2 \in X [f(x_1) = f(x_2) \implies x_1 = x_2]$$

and it is said to be **surjective** iff $Y = f(X)$, i.e.

$$\forall y \in Y \exists x \in X y = f(x).$$

A map is **bijective** iff it is both injective and surjective. Thus

- (1) A map is injective if and only if it has a left inverse;
- (2) A map is surjective if and only if it has a right inverse;
- (3) A map is bijective if and only if it has a (two-sided) inverse.

In the lingo used in Remark 5.5, a map is surjective if and only if its range equals its target.

Remark 6.4. Both Lang and Morgan use the more modern terms injective, surjective, bijective, but Buck and Rudin use the older terminology **one-one** instead of injective, **onto** instead of *surjective*, and **one-one onto** instead of bijective. The older terminology is used in teaching calculus.

Remark 6.5. The 'only if' part of item (2) is called the **Axiom of Choice**. It was once controversial because one can imagine a situation where one can prove that a map f is surjective but where one cannot give an explicit formula for a right inverse.

Example 6.6. The assertions (1-3) in Definition 6.3 are false if continuity (defined later) is required: There is a continuous bijective map whose inverse is not continuous (and hence continuous injective map which does not have a continuous left inverse, and a continuous surjective map which does not have a continuous right inverse). Let $S := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$ denote the unit circle in $\mathbb{R}^2$ and define $f : [0, 2\pi) \rightarrow S$ by $f(\theta) = (\cos \theta, \sin \theta)$. Then f is bijective and continuous but f^{-1} is not continuous. In Theorem 16.9 below we will see that the inverse of a continuous bijective map is continuous if its domain is compact (defined later), but $[0, 2\pi)$ is not compact.

Example 6.7. Consider the four maps

$$f_1 : \mathbb{R} \rightarrow \mathbb{R}, \quad f_2 : [0, \infty) \rightarrow \mathbb{R}, \quad f_3 : \mathbb{R} \rightarrow [0, \infty), \quad f_4 : [0, \infty) \rightarrow [0, \infty)$$

defined by $f_i(x) = x^2$. Then f_1 is not injective and not bijective, f_2 is injective but not surjective, f_3 is surjective but not injective, and f_4 is bijective. Any map $g_2 : \mathbb{R} \rightarrow [0, \infty)$ such that $g_2(y) = \sqrt{y}$ for $y \geq 0$ is a left inverse to f_2 , and any map $g_3 : [0, \infty) \rightarrow \mathbb{R}$ such that $g_3(y) = \pm\sqrt{y}$ (the $\pm$ can depend on y) is a right inverse to f_3 . The inverse map to f_4 is $f_4^{-1}(y) = \sqrt{y}$.

Problem 6.8. Which (if any) of the false formulas in problem 5.7 become true if we assume that the map f is injective? surjective? (Proof or counter example.)

7 Mathematical Induction

7.1. The principle of **mathematical induction** is the following axiom.

Let $S \subseteq \mathbb{N}$ be a set of nonnegative integers. Assume that $0 \in S$ and that $n \in S \implies n + 1 \in S$. Then $S = \mathbb{N}$.

It is usually taught in College Algebra as a means of proving identities like

$$\sum_{k=0}^n k = \frac{n(n+1)}{2} \quad \text{and} \quad \sum_{k=0}^n k^2 = \frac{n(n+1)(2n+1)}{6}.$$

For example to prove the first identity let S denote the set of all $n \in \mathbb{N}$ for which the statement is true. Then $0 \in S$ and if $n \in S$ then

$$\sum_{k=0}^{n+1} k = \left(\sum_{k=0}^n k \right) + (n+1) = \frac{n(n+1)}{2} + (n+1) = \frac{(n+1)(n+2)}{2}$$

so $n+1 \in S$. Hence by the principle of mathematical induction, $S = \mathbb{N}$, i.e. the identity is true for all $n \in \mathbb{N}$.

7.2. We will often use **inductive definitions** where a sequence is defined by specifying the first few values and then giving a “recurrence relation” for determining the remaining values from the earlier ones. For example, the **Fibonacci numbers** are defined by

$$F_0 = 0, \quad F_1 = 1, \quad F_n = F_{n-1} + F_{n-2}.$$

Thus $F_2 = 1$, $F_3 = 3$, $F_4 = 5$, $F_5 = 8$, $F_9 = 55$ etc.

Problem 7.3. Prove the sum of squares formula in 7.1

8 Infinite Sets

Definition 8.1. A set S is said to be

- **infinite** iff there is a injective map $f : \mathbb{N} \rightarrow S$,
- **finite** iff it is not infinite,
- **countable** iff there is a bijective map $J \rightarrow S$ for some $J \subseteq \mathbb{N}$, and
- **uncountable** iff it is not countable.⁴

Proposition 8.2. *If there is a bijective map $\{1, 2, \dots, m\} \rightarrow \{1, 2, \dots, n\}$, then $m = n$. Hence if S is finite, there is a unique n (called the **cardinality** of S) such that there is a bijective map $\{1, 2, \dots, n\} \rightarrow S$.*

Problem 8.3. Suppose that S and T are finite sets of cardinality s and t respectively? What is the cardinality of $S \times T$? Of the set T^S of maps from S to T ? Of the set 2^S of subsets of S ? Of $S \cup T$ (assuming that $S \cap T = \emptyset$)?

The following four propositions are proved in Chapter 2 of Morgan. See also Lang pages 12-15. The main value of these propositions is to give the student experience in constructing maps. The constructions are not needed in the rest of the course, but learning them will give you some confidence.

Proposition 8.4. *A subset of a countable set is countable.*

Proof. A subset of a finite set is finite so it is enough to consider a subset $S \subset \mathbb{N}$ of the nonnegative integers. Define s_n inductively (see 7.2) by $s_{n+1} =$ the least element of $S \setminus \{s_1, \dots, s_n\}$. If for some n the set $S \setminus \{s_1, \dots, s_n\}$ is empty then the induction stops and the set S is finite of cardinality n . Otherwise, the map $\mathbb{N} \rightarrow S : n \mapsto s_n$ is injective. But it is also surjective: if $k \in S$ then $k = s_n$ for some $n \leq k$. $\square$

Proposition 8.5. *If S and T are countable, so is $S \times T$.*

⁴This is the terminology used by Morgan. For Buck and Rudin *countable* means *countable and infinite* which is equivalent to saying that there is a bijective map $\mathbb{N} \rightarrow S$. Lang uses the terms **denumerable** and **nondenumerable** for *countable and infinite* and *uncountable*.

Proof. It is enough to prove that $\mathbb{N} \times \mathbb{N}$ is countable. Consider the following enumeration of $\mathbb{N} \times \mathbb{N}$:

$$\begin{array}{cccccc}
 f(0) = (0,0) & f(2) = (0,1) & f(5) = (0,2) & f(9) = (0,3) & f(14) = (0,4) & \cdots \\
 f(1) = (1,0) & f(4) = (1,1) & f(8) = (1,2) & f(13) = (1,3) & \cdot & \cdots \\
 f(3) = (2,0) & f(7) = (2,1) & f(12) = (2,2) & \cdot & \cdot & \cdots \\
 f(6) = (3,0) & f(11) = (3,1) & \cdot & \cdot & \cdot & \cdots \\
 f(10) = (4,0) & \cdot & \cdot & \cdot & \cdot & \cdots \\
 \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
 \end{array}$$

This defines a bijection $f : \mathbb{N} \rightarrow \mathbb{N} \times \mathbb{N}$. (Compare with Morgan page 10.) On each diagonal the sum of the coordinates is constant and the input n to the function $f(n)$ increases as you go from left to right up the diagonal. $\square$

Problem 8.6. For f as in Proposition 8.5, find $f(1307)$ and $f^{-1}(58, 19)$. Hint: The formula for the sum of the first n natural numbers from 7.1 will be helpful.

Proposition 8.7. *The set $\mathbb{Q}$ of rational numbers is countable.*

Proposition 8.8. *The set $\mathbb{R}$ of real numbers is uncountable.*

Theorem 8.9. *The set $2^{\mathbb{N}}$ of all subsets of $\mathbb{N}$ is uncountable.*

Proof. Suppose not. Then there is a bijective map $\mathbb{N} \rightarrow 2^{\mathbb{N}} : n \mapsto A_n$. Let $B = \{n \in \mathbb{N} : n \notin A_n\}$. Then $B = A_k$ for some k . Either $k \in A_k$ or $k \notin A_k$. If the former, then $k \notin B$ so $A_k \neq B$. If the latter, $k \in B$, so again $A_k \neq B$. Either way we have a contradiction to the assumption that the map $n \mapsto A_n$ is bijective. $\square$

Theorem 8.10 (Dedekind). *A set S is infinite if and only if there is an injective map from S to itself which is not surjective.*

Theorem 8.11 (Cantor Schroeder Bernstein). *If there is an injective map from S to T and there is an injective map from T to S , then there is a bijective map from S to T .*

Proof. Buck page 552. $\square$

9 Axioms for the Real Numbers

We state here the axioms for the real number system $\mathbb{R}$. We shall accept these axioms without proof but it can be proved (from more general axioms) that there is an essentially unique structure satisfying them.

9.1. Algebraic Axioms. The set $\mathbb{R}$ of **real numbers** is equipped with two operations

$$\mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} : (a, b) \mapsto a + b, \quad \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R} : (a, b) \mapsto a \cdot b$$

such that the usual laws of grade school arithmetic hold:

(Commutative Laws) $a + b = b + a$ and $a \cdot b = b \cdot a$.

(Associative Laws) $(a + b) + c = a + (b + c)$ and $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.

(Distributive Law) $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$.

(Zero, One) There are (necessarily unique) distinct elements $0, 1 \in \mathbb{R}$ such that $a + 0 = a$ and $a \cdot 1 = a$ for all $a \in \mathbb{R}$.

(Inverses) For every $a \in \mathbb{R}$ there is a (necessarily unique) element $-a$ such that $a + (-a) = 0$. For every $a \in \mathbb{R} \setminus \{0\}$ there is a (necessarily unique) element a^{-1} such that $a \cdot a^{-1} = 1$.

The standard notations from high school algebra are used: in particular, $ab := a \cdot b$, $a - b := a + (-b)$, and $a/b = a \cdot b^{-1}$.

9.2. Order Axioms. The set $\mathbb{R}$ has an order relation denoted $a < b$ satisfying the following laws for all $a, b, c \in \mathbb{R}$:

(Trichotomy) Exactly one of the alternatives $a < b$, $a = b$, $b < a$, holds.

(Transitivity) $a < b, b < c \implies a < c$.

(Addition) $a < b \implies a + c < b + c$.

(Multiplication) $0 < a, b \implies 0 < ab$.

The other order notations are defined as usual, i.e. $a < b \iff b > a$ and $a \leq b \iff b \geq a \iff$ either $a < b$ or $a = b$.

Remark 9.3. All the rules of algebra used in College Algebra (Math 112) follow from the Algebraic Axioms 9.1 and Order Axioms 9.2. For example, $(a + b)^2 = a^2 + 2ab + b^2$, $a^2 \geq 0$, etc.

Problem 9.4. Does it follow from the axioms in 9.1 and 9.2 that $a^{-1} > 0$ if $a > 0$? Explain.

Definition 9.5. A set S of real numbers is said to be **bounded above** iff there is a number $b \in \mathbb{R}$ such that $x \leq b$ for all $x \in S$; the number b is then called an **upper bound** for S . A number $b \in \mathbb{R}$ is called a **least upper bound** for S iff it is an upper bound for S and $b \leq b'$ for every other upper bound b' for S . Similarly the set S is said to be **bounded below** iff there is an number $a \in \mathbb{R}$ such that $a \leq x$ for all $x \in S$; the element b is then called a **lower bound** for S . An element $a \in \mathbb{R}$ is called a **greatest lower bound** iff it is an lower bound for S and $a' \leq a$ for every other lower bound a' for S . The words **infimum** and *greatest lower bound* are synonymous as are the words **supremum** and *least upper bound*. The least upperbound of the set S will be denoted $\sup(S)$ and the greatest lower bound of the set S will be denoted $\inf(S)$. We write $\sup(S) = \infty$ when S is not bounded above and $\inf(S) = -\infty$ when S is not bounded below.

9.6. Completeness Axiom. Every set S of real numbers which is bounded above has a least upper bound, i.e.

if $x \leq b$ for all $x \in S$, then $x \leq \sup(S) \leq b$ for all $x \in S$.

Because multiplication by -1 reverses the order it is the same to say that every set which is bounded below has a greatest lower bound. Thus

if $a \leq x$ for all $x \in S$, then $a \leq \inf(S) \leq x$ for all $x \in S$.

Remark 9.7. For Morgan the completeness axiom is a theorem (see page 44 of his book) but most authors of undergraduate texts don't prove it. Of course it can't be proved until we give a precise construction (definition) of the real numbers. Morgan takes the view that a real number is a number with a decimal expansion, but avoids certain subtle issues connected with this view. Actually making his definition precise would involve defining how to do the arithmetic operations with decimal expansions and would be a bit tedious. To see why, imagine that $a + b = c$ and that

$$a = \sum_{n=1}^{\infty} a_n 10^{-n}, \quad b = \sum_{n=1}^{\infty} b_n 10^{-n}, \quad c = \sum_{n=1}^{\infty} c_n 10^{-n},$$

where the coefficients, a_n, b_n, c_n are integers between 0 and 9 and try to express c_n in terms the a_n 's and the b_n 's. In an appendix to his book, Buck sketches a construction the real numbers by something called **Dedekind cuts** but leaves out many details. Rudin explains cuts in an appendix to his chapter 1. (The book *Foundations of Analysis* by Edmund Landau gives all the details.) The completeness axiom is an easy consequence of this construction. One can also define the real numbers using 'equivalence classes of Cauchy sequences' (see 12.15) and again the completeness axiom is a consequence; if we get to the topic of completion of metric spaces, I'll explain it then.

The crucial point is the uniqueness 'up to isomorphism' which we will prove in Appendix A. It means that if we start from the axioms, it doesn't matter what definition of the real numbers we use. We will not use the completeness axiom until the proof of Theorem 12.7 and the exposition is arranged in such a way that it will be clear that the reasoning is not circular. The following exercises will help you understand some of the issues.

Problem 9.8. Prove the following **Archimedean Property** of the real numbers: There is neither an infinite real number nor an infinitesimal real number. More precisely,

- (1) There is no real number which is larger than every integer.
- (2) For every positive real number $\varepsilon > 0$ there is a positive integer n such that $1/n < \varepsilon$.

Hint: If $\omega > n$ for every integer n what about $\omega - 1$? The proof will use the completeness axiom.

Problem 9.9. Let $\mathcal{R}$ denote the set of real valued rational functions, i.e. $f \in \mathcal{R}$ iff $f(x) = p(x)/q(x)$ where $p(x)$ and $q(x)$ are polynomials with real coefficients (and $q(x)$ is not the zero polynomial). For $f, g \in \mathcal{R}$ define an order relation by the condition that $f > g$ iff there exists an M such that $f(x) > g(x)$ for all $x > M$. Then the set $\mathcal{R}$ satisfies the algebraic axioms and order axioms given above. View $\mathbb{R}$ (and hence $\mathbb{Z}$) as a subset of $\mathcal{R}$ by identifying the real number c with the constant function whose value is always c . Exhibit (in the lingo of Problem 9.8) an infinite element $f \in \mathcal{R}$ and an infinitesimal element $g \in \mathcal{R}$. Hint: What is $\lim_{x \rightarrow \infty} \frac{p(x)}{q(x)}$?

Problem 9.10. Prove that there is a rational number in every nonempty open interval $(a, b) \subset \mathbb{R}$.

Solution: By the Archimedean property from Problem 9.8 there are integers m and n with $m < a < b < n$. Again by the Archimedean property there is a positive integer k with $0 < 1/k < b - a$. Define $q_j = m + j/k$ for $j = 0, 1, \dots, k(n - m)$. Then $q_j = m$ when $j = 0$, $q_j = n$ when $j = k(n - m)$, and $q_j - q_{j-1} = 1/k < b - a$. If j is the largest integer such that $q_{j-1} \leq a$ then $a < q_j < b$ as $q_j - q_{j-1} = 1/k < b - a$.

Problem 9.11. Let $\mathbb{Q}(\sqrt{2})$ denote the set of all numbers of form $x = a + b\sqrt{2}$ where a and b are rational. Show that $\mathbb{Q}(\sqrt{2})$ is closed under the algebraic operations, i.e if $x, y \in \mathbb{Q}(\sqrt{2})$, then $x \pm y \in \mathbb{Q}(\sqrt{2})$, $xy \in \mathbb{Q}(\sqrt{2})$, and (if $y \neq 0$) $x/y \in \mathbb{Q}(\sqrt{2})$. Show further that $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$. Does $\mathbb{Q}(\sqrt{2})$ satisfy the completeness axiom?

Problem 9.12. Assume that

$$x = \sum_{k=1}^{\infty} x_k 10^{-k}$$

is the decimal expansion of a real number x so that the x_k are integers between 0 and 9. How large must n be to ensure that

$$\left| x - \sum_{k=1}^n x_k 10^{-k} \right| < 10^{-7}?$$

Assume that $a + b = c$ as in Remark 9.7 and that

$$A = \sum_{k=1}^n a_k 10^{-k}, \quad B = \sum_{k=1}^n b_k 10^{-k}.$$

How large must n be to ensure that $|c - (A + B)| < 10^{-7}$?

Solution: For any n we have

$$\left| x - \sum_{k=1}^n x_k 10^{-k} \right| = \sum_{k=n+1}^{\infty} x_k 10^{-k} \leq \sum_{k=n+1}^{\infty} 9 \cdot 10^{-k} = 10^{-n}$$

We have equality if $x_k = 9$ for $k > n$ so to be sure that the strict inequality holds we should take $n = 8$. For the second part we have

$$|(a + b) - (A + B)| \leq |a - A| + |b - B| \leq 2 \cdot 10^{-8}$$

by the first part and the triangle inequality. Since $2 \cdot 10^{-8} < 10^{-7}$ we have that $n = 8$ works here as well. Note: The point of this problem is to convince you that it is awkward to work with decimal expansion if you want to be super careful. Also the wording of the problem (“How large must n be to ensure that”) suggests that I want the smallest possible n that works. Usually I don’t care about that. I could have replaced the phrase ‘How large must n be to ensure that’ by “Find N so that for $n \geq N$ we have”. In that case the answer $n = 13472$ would be correct (but overkill).

10 Distance

10.1. The **distance** $d(p, q)$ between two points $p = (x_1, x_2, \dots, x_n)$ and $q = (y_1, y_2, \dots, y_n)$ in $\mathbb{R}^n$ is defined by⁵

$$d(p, q) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}.$$

The distance $d(v, 0)$ from a vector $v \in \mathbb{R}^n$ to the origin is called the **norm** of v denoted by $|v|$ so

$$d(p, q) := |p - q|.$$

The norm satisfies the following laws for $v, w \in \mathbb{R}^n$:

(zero norm) $|v| = 0 \iff v = 0$,

(homogeneity) $|av| = a|v|$ if $a > 0$,

(symmetry) $|-v| = |v|$,

(triangle inequality) $|v + w| \leq |v| + |w|$.

The zero norm law holds because a sum of squares vanishes only if each summand vanishes and the triangle inequality is proved in problem 10.3 below.⁶ The laws for the norm imply that the distance function satisfies the following:

(zero distance) $d(p, q) = 0 \iff p = q$,

⁵ In Chapters I-V Buck writes $|p - q|$ instead of $d(p, q)$ but uses the notation $d(p, q)$ starting in Chapter VI (page 304) in a more general setting.

⁶ See also the Corollary on page 14 of Buck or Theorem 2.1 on page 134 of Lang.

(**symmetry**) $d(p, q) = d(q, p)$,

(**triangle inequality**) $d(p, r) \leq d(p, q) + d(q, r)$.

These are proved by reading $v = p - q$ and $w = q - r$ in the corresponding law for the norm.

10.2. Define the **inner product** of two vectors in $\mathbb{R}^n$ by

$$\langle u, v \rangle = u_1v_1 + u_2v_2 + \cdots + u_nv_n$$

for $u = (u_1, u_2, \dots, u_n)$, $v = (v_1, v_2, \dots, v_n)$. In freshman calculus you called this the **dot product** and learned that the angle θ between the two vectors u and v satisfies

$$\langle u, v \rangle = \cos \theta |u| |v|.$$

Since the cosine takes values between -1 and +1, it follows that

$$|\langle u, v \rangle| \leq |u| |v|.$$

This inequality is called the **Schwarz inequality**.

Problem 10.3. In this problem you will prove the triangle inequality without using trigonometry in two steps.

1. Prove the Schwarz inequality. (Hint: The quadratic polynomial $f(t) = |u + tv|^2 = At^2 + Bt + C$ is always nonnegative so its discriminant $B^2 - 4AC$ is nonpositive.)
2. Derive the triangle inequality as a corollary. (Hint: Compare $|u \pm v|^2$ and $(|u| + |v|)^2$.)

Definition 10.4. For $p \in \mathbb{R}^n$ and $\delta > 0$ the set

$$B(p, \delta) := \{q \in \mathbb{R}^n : d(p, q) < \delta\}$$

is called the **open ball** centered at p with radius δ . When $X \subseteq \mathbb{R}^n$ we often use the abbreviation

$$B_X(p, \delta) := B(p, \delta) \cap X := \{q \in X : d(p, q) < \delta\}$$

Remark 10.5. When $n = 1$ the open ball is an open interval:

$$\begin{aligned} B(a, \delta) &= \{x \in \mathbb{R} : |x - a| < \delta\} \\ &= \{x \in \mathbb{R} : a - \delta < x < a + \delta\} \\ &= (a - \delta, a + \delta) \end{aligned}$$

for $a \in \mathbb{R}$

Definition 10.6. A set S is **bounded** iff it is contained in some large ball, i.e. iff there exists $M > 0$ such that $|p| < M$ for all $p \in S$. Thus a set of real numbers is bounded (see 9.5) if and only if it is bounded above and bounded below.

11 Limits

The intuitive idea of the notation

$$\lim_{p \rightarrow p_0} F(p) = L$$

is that $F(p)$ is very close to L when p is very close to p_0 . Some authors write $F(p) \rightarrow L$ as $p \rightarrow p_0$; others write $F(p) \approx L$ when $p \approx p_0$. In this chapter we give a more precise definition. The following lingo is helpful.

11.1. A set U is called a **neighborhood** of the point p if U contains some open ball $B(p, \delta)$ centered at p . A **punctured neighborhood** of p is a set of form $U \setminus \{p\}$ where U is a neighborhood of p . A point p is a **accumulation point**⁷ of a set S iff every punctured neighborhood of p contains a point of S . The following equivalent definition appears in some books.

Proposition 11.2. *A point p is an accumulation point of the set S if and only if every neighborhood of p contains infinitely many points of S .*

Proof. "If" is easy: an infinite set is nonempty and at most one of the points in an infinite set is p . For "only if" assume p is an accumulation point of the set S and choose a neighborhood U of p . By definition there is a point $p_0 \in U \cap S \setminus \{p\}$. Let $\delta_0 = |p - p_0|$. For $n > 0$ define $\delta_n > 0$ and $p_n \in S$ inductively (see 7.2) by $\delta_n = \min(|p_n - p|, 1/n)$ and $p_{n+1} \in B(p, \delta_n) \cap S$. The

⁷ Buck uses the term **cluster point** and some authors use the term **limit point**.

map $n \mapsto p_n$ is injective as $|p_n - p| < |p_m - p|$ for $n > m$. Choose $\delta > 0$ so that $B(p, \delta) \subset U$ (by the definition of neighborhood). Then $\delta_n < \delta$ for $1/n < \delta$ so $p_n \in B(p, \delta) \cap S \subseteq U \cap S$. Hence U contains the infinite set $\{p_n : n > 1/\delta\}$. $\square$

11.3. Let p_0 be a accumulation point of a set S and F be a function defined on S (but possibly not at p_0). The notation

$$\lim_{p \rightarrow p_0} F(p) = L$$

means that for every neighborhood V of L of there is a punctured neighborhood $U \setminus \{p\}$ of p_0 such that $f(S \cap U \setminus \{p\}) \subset V$. When $p_0 \in S$ and p_0 is a accumulation point of S we have that a function f defined on S is continuous at p_0 (see Definition 13.1 below) if and only if

$$\lim_{p \rightarrow p_0} f(p) = f(p_0)$$

(and the function is trivially continuous at a point $p_0 \in S$ which is not an accumulation point of S). However, the limit notation is usually used in situations where (p_0 is a accumulation point of S but) $p_0 \notin S$. For example, the **derivative** of a real valued function $f : I \rightarrow \mathbb{R}$ defined on an open interval $I \subseteq \mathbb{R}$ is defined by

$$f'(x_0) := \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

The ratio in the limit is undefined when $x = x_0$ but is defined for nearby values of x .

11.4. For a real valued function f defined on a subset of $\mathbb{R}$ we can extend the definition of the notation $\lim_{x \rightarrow a} F(x) = L$ to include the cases where $a = \pm\infty$ and/or $L = \pm\infty$ as follows. Let

$$\hat{\mathbb{R}} := \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$$

consist of the set of real numbers together with two additional points which we think of as located at infinity. The set $\hat{\mathbb{R}}$ is sometimes called the set of **extended real numbers**. Extend the usual order relation on $\mathbb{R}$ to $\hat{\mathbb{R}}$ in the obvious way. For $a \in \hat{\mathbb{R}}$, a set $U \subseteq \hat{\mathbb{R}}$ is called **neighborhood** of a iff

either $a \in \mathbb{R}$ and U contains an open interval $(a - \delta, a + \delta)$ for some $\delta > 0$,

or else $a = \infty$ and U contains an interval $(M, \infty]$ for some $M > 0$,

or else $a = -\infty$ and U contains an interval $[-\infty, -M)$ for some $M > 0$.

Because $B(a, \delta) = (a - \delta, a + \delta)$ this definition agrees with the definition in 11.3 for $a \in \mathbb{R}$. A point $a \in \hat{\mathbb{R}}$ is called a **accumulation point** of a subset $S \subseteq \mathbb{R}$ iff every punctured neighborhood of a intersects S . If $f : S \rightarrow \mathbb{R}$ and a is a accumulation point of S , then the notation

$$\lim_{x \rightarrow a} F(x) = L$$

means that for every neighborhood V of L there is a punctured neighborhood U of a such that $f(U \cap S) \subseteq V$.

Remark 11.5. Unraveling the above definitions we see that for $a, L \in \mathbb{R}$ we have that $\lim_{x \rightarrow a} F(x) = L$ iff $\forall \varepsilon > 0 \exists \delta > 0$ such that $\forall x \in S$ we have that $0 < |x - a| < \delta \implies |F(x) - L| < \varepsilon$. Also $\lim_{x \rightarrow \infty} F(x) = L$ iff $\forall \varepsilon > 0 \exists M > 0$ such that $\forall x \in S$ we have that $M < x \implies |F(x) - L| < \varepsilon$ with similar definitions for the other cases where $a, L \in \{\pm\infty\}$. The various definitions given in 11.3 and 11.4 are easier to understand because the lingo makes them look the same and because there aren't so many symbols. This is why the terminology was invented.

12 Convergence of Sequences

12.1. A **sequence** is a function defined on a subset of the integers. (Usually this subset is the set $\mathbb{Z}^+ := \{n \in \mathbb{Z} : n > 0\}$ of positive integers or the set $\mathbb{N} := \{n \in \mathbb{Z} : n \geq 0\}$ of nonnegative integers.) It is customary to denote the value of a sequence at an integer n with a subscript rather than with parentheses and to denote a sequence with a notation like $(p_n)_n$ or $(p_n)_{n \in \mathbb{Z}^+}$.

Definition 12.2. The sequence $(p_n)_n$ of points of $\mathbb{R}^m$ is said to **converge** to the point $p \in \mathbb{R}^m$ iff

$$\lim_{n \rightarrow \infty} p_n = p$$

This is sometimes abbreviated as $p_n \rightarrow p$ as $n \rightarrow \infty$. We say a sequence **converges** or is **convergent** iff it converges to p for some $p \in \mathbb{R}^m$. A sequence is said to **diverge** when it does not converge. (A sequence in $\mathbb{R}$ whose limit is infinite is also said to diverge.)

Remark 12.3. Using the lingo introduced in 11.4 this may be stated as $\lim_{n \rightarrow \infty} p_n = p$ iff for every $\varepsilon > 0$ there exists $N = N(\varepsilon) > 0$ such that $n \geq N \implies |p_n - p| < \varepsilon$.

Theorem 12.4. Assume that $(a_n)_n$ and $(b_n)_n$ are convergent sequences of real numbers:

$$\lim_{n \rightarrow \infty} a_n = a, \quad \lim_{n \rightarrow \infty} b_n = b.$$

Then

$$\lim_{n \rightarrow \infty} a_n + b_n = a + b, \quad \lim_{n \rightarrow \infty} a_n b_n = ab.$$

Moreover, if $b \neq 0$ then $b_n \neq 0$ for sufficiently large n and

$$\lim_{n \rightarrow \infty} \frac{a_n}{b_n} = \frac{a}{b}.$$

Proof. **(I)** We prove $\lim_{n \rightarrow \infty} a_n + b_n = a + b$. Choose $\varepsilon > 0$. By hypothesis there exists N_1 and N_2 such that

$$n > N_1 \implies |a_n - a| < \frac{\varepsilon}{2}, \quad n > N_2 \implies |b_n - b| < \frac{\varepsilon}{2}.$$

Let $N = \max(N_1, N_2)$. Then for $n > N$ we have

$$\begin{aligned} |(a_n + b_n) - (a + b)| &= |(a_n - a) + (b_n - b)| \\ &\leq |a_n - a| + |b_n - b| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon \end{aligned}$$

as required.

(II) We prove $\lim_{n \rightarrow \infty} a_n b_n = ab$. Choose $\varepsilon > 0$. By hypothesis there exists N_0, N_1, N_2 such that

$$\begin{aligned} n > N_0 &\implies |a_n - a| < 1, \\ n > N_1 &\implies |b_n - b| < \frac{\varepsilon}{2(|a| + 1)}, \\ n > N_2 &\implies |a_n - a| < \frac{\varepsilon}{2|b|}. \end{aligned}$$

Let $N = \max(N_0, N_1, N_2)$. Then for $n > N$ we have

$$\begin{aligned}
|a_n b_n - ab| &= |a_n(b_n - b) + (a_n - a)b| \\
&\leq |a_n| |b_n - b| + |a_n - a| |b| \\
&< (|a| + 1) |b_n - b| + |a_n - a| |b| \\
&< (|a| + 1) \frac{\varepsilon}{2(|a| + 1)} + \frac{\varepsilon}{2|b|} |b| \\
&= \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

(III) We prove $\lim_{n \rightarrow \infty} 1/b_n = 1/b$ if $b \neq 0$. Choose $\varepsilon > 0$. By hypothesis there exists N_1 such that for $n > N_1$ we have that $|b_n - b| < \frac{1}{2}|b|$ (and hence that $\frac{1}{2}|b| < |b_n|$) and there exists N_2 such that $|b_n - b| < \varepsilon \frac{1}{2}|b|^2$. Then for $n > \max(N_1, N_2)$ we have

$$\left| \frac{1}{b_n} - \frac{1}{b} \right| = \frac{|b_n - b|}{|b b_n|} < \frac{|b_n - b|}{\frac{1}{2}|b|^2} < \varepsilon.$$

(IV) That $\lim_{n \rightarrow \infty} a_n/b_n = a/b$ follows immediately from (II) and (III) by substituting $1/b_n$ for b_n and $1/b$ for b $\square$

Definition 12.5. A real valued function f defined on a subset of the real numbers $\mathbb{R}$ is called

increasing iff $x_1 < x_2 \implies f(x_1) < f(x_2)$,

decreasing iff $x_1 > x_2 \implies f(x_1) > f(x_2)$,

monotonic iff it is either increasing or decreasing.

Remark 12.6. If $<$ is replaced by $\leq$ in this definition, the meaning changes: a constant function satisfies the modified definition. When we want to use the weaker form we use the term **nondecreasing** instead of *increasing*, the term **nonincreasing** instead of *decreasing*, and the term **weakly monotonic** instead of *monotonic*. Thus in freshman calculus it is correct to say that *a differentiable function is nondecreasing if and only if its derivative is everywhere nonnegative* but incorrect to say that *a differentiable function is increasing if and only if its derivative is everywhere positive*. (If $f(x) = x^3$, then f is increasing but $f'(0) = 0$.) To avoid this confusion some authors insert the word **strictly** in 12.5.

Theorem 12.7. *A bounded weakly monotonic sequence is convergent. In fact*

$$\lim_{n \rightarrow \infty} a_n = \sup_n a_n$$

if the sequence $(a_n)_n$ is nondecreasing, and

$$\lim_{n \rightarrow \infty} a_n = \inf_n a_n$$

if the sequence $\{a_n\}$ is nonincreasing. (See 9.5.)

Proof. (Compare Buck page 47 or Lang page 35 or Morgan page 38.) In this proof we use the completeness axiom 9.6 for the first time.

Assume that the sequence $(a_n)_n$ is nondecreasing and let $a = \sup\{a_n : n \in \mathbb{N}\}$. Then $a_n \leq a$ for all n as a is an upperbound for the set $\{a_n : n \in \mathbb{N}\}$. Choose $\varepsilon > 0$. Then $a - \varepsilon < a$ so $a - \varepsilon$ is not an upperbound for the set $\{a_n : n \in \mathbb{N}\}$. Hence there is an N with $a - \varepsilon < a_N$. For $n > N$ we have $a_N \leq a_n$ as the sequence $(a_n)_n$ is nondecreasing so

$$a - \varepsilon < a_N \leq a_n \leq a < a + \varepsilon$$

so $|a_n - a| < \varepsilon$ for $n > N$ as required. The nonincreasing case is proved by an analogous argument, or alternatively by applying the nondecreasing case to the sequence $(-a_n)_n$. $\square$

12.8. We introduce some handy notation. For any sequence $(a_k)_k$ of real numbers we have $\{a_k : k \geq n\} \subseteq \{a_k : k \geq m\}$ for $m < n$. If the sequence $(a_k)_k$ is bounded above, then the sequence $s_n := \sup\{a_k : k \geq n\}$ is nonincreasing. The limit of the latter sequence is denoted

$$\limsup_{n \rightarrow \infty} a_n := \lim_{n \rightarrow \infty} \sup\{a_k : k \geq n\}.$$

Similarly for a sequence which is bounded below,

$$\liminf_{n \rightarrow \infty} a_n := \lim_{n \rightarrow \infty} \inf\{a_k : k \geq n\}.$$

Definition 12.9. When $n_1 < n_2 < n_3 < \dots$ is an increasing sequence of positive integers, the sequence $(p_{n_k})_k$ is called a **subsequence** of the sequence $(p_n)_n$.

Remark 12.10. If the sequence $(p_n)_n$ converges to p , then every subsequence $(p_{n_k})_k$ also converges to p . This follows immediately from the definition of convergence: $n_k \geq k$ so if $N = N(\varepsilon)$ satisfies $|p_n - p| < \varepsilon$ for $n > N(\varepsilon)$ then in particular we have $|p_{n_k} - p| < \varepsilon$ for $k > N(\varepsilon)$.

Theorem 12.11 (Bolzano-Weierstrass). *Every bounded sequence in $\mathbb{R}^m$ has a convergent subsequence.*

Proof. Compare Buck Theorem 22 page 62 or Lang page 38-39 or Morgan page 38. We first do the case $m = 1$. Let $(a_n)_n$ be a bounded sequence of real numbers. Then there is a number M such that $-M \leq a_n \leq M$ for all n . For each n define

$$A_n := \{a_{n+1}, a_{n+2}, \dots\}, \quad b_n := \inf A_n.$$

Since $A_n \subset A_{n+1}$ we have that $b_n \leq b_{n+1}$ so the sequence $(b_n)_n$ converges to its supremum $b := \sup\{b_n\} := \liminf_n a_n$. We will show that a subsequence of the sequence $(a_n)_n$ also converges to b . For $n \in \mathbb{N}$ we have that $b_n < b_n + n^{-1}$ and b_n is the greatest lower bound for A_n so $b_n + n^{-1}$ is not a lower bound for A_n so there is a $c_n \in A_n$ with $b_n \leq c_n < b_n + n^{-1}$. As b_n converges to b by Theorem 12.7 we have that for every $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that $|b_n - b| < \varepsilon/2$ for $n > N(\varepsilon)$. Hence for $n > \max(2/\varepsilon, N(\varepsilon))$ we have that $|c_n - b| \leq |c_n - b_n| + |b_n - b| < n^{-1} + \varepsilon/2 < \varepsilon$ which shows that $(c_n)_n$ converges to b .

Now $c_n \in A_n$ so $c_n = a_j$ for some $j = j(n) \geq n + 1$, but we aren't quite done because the definition of subsequence requires that the subscripts $j(n)$ increase and there is no reason for that to be true. However we can extract a further subsequence by induction. Namely if $n_1 < n_2 < \dots < n_k$ have been defined, define n_{k+1} by $n_{k+1} = j(n_k)$. Then $n_{k+1} = j(n_k) \geq n_k + 1 > n_k$ as required. (The further subsequence still converges by Remark 12.10.)

Now we prove the theorem for a sequence of points in $\mathbb{R}^m$ by induction on m . Assume the theorem holds for $\mathbb{R}^m$ and choose a bounded sequence $(p_n)_n$ of points in $\mathbb{R}^{m+1}$. Then $p_n = (q_n, a_n)$ where $q_n \in \mathbb{R}^m$ and $a_n \in \mathbb{R}$. That the sequence $(p_n)_n$ is bounded means that there is an M such that $|p_n| \leq M$ for all n . As $|p_n|^2 = |q_n|^2 + a_n^2$ it follows that $|q_n| \leq M$ and $|a_n| \leq M$ for all n , i.e. the sequence $(q_n)_n$ and $(a_n)_n$ are also bounded. By the inductive hypothesis the sequence $(q_n)_n$ has a subsequence $(q_{n_k})_k$ converging to q . By replacing the sequence $(p_n)_n$ by the sequence $(p_{n_k})_k$ we may assume that the

sequence $(q_n)_n$ converges. (Remark 12.10.) Now by the case $m = 1$ (already proved) the sequence $(a_n)_n$ contains a convergent subsequence $(a_{n_k})_k$. Hence

$$\lim_{k \rightarrow \infty} q_{n_k} = q, \quad \lim_{k \rightarrow \infty} a_{n_k} = a.$$

By the triangle inequality $|(q', a') - (q, a)| \leq |q' - q| + |a' - a|$ so

$$\lim_{k \rightarrow \infty} p_{n_k} = \lim_{k \rightarrow \infty} (q_{n_k}, a_{n_k}) = (q, a)$$

as required. $\square$

Corollary 12.12. *For a subset $S \subseteq \mathbb{R}^m$ the following conditions are equivalent.*

- (1) *For every sequence $(p_n)_n$ of points of S there is a subsequence $(p_{n_k})_k$ which converges to $p \in S$.*
- (2) *The set S closed and bounded.*

Proof. Assume (1). The certainly every convergent sequence $p_n \in S$ has limit in S so S is closed by Theorem 14.11. Also S is bounded as otherwise for every n there would be a point $p_n \in S$ with $|p_n| > n$ and this sequence cannot have a convergent subsequence. Conversely assume (2) and let $(p_n)_n$ be a sequence of points in S . As S is bounded there is a convergent subsequence (by Theorem 12.11) and as S is closed the limit of this subsequence is a point of S (by Theorem 14.11). $\square$

Corollary 12.13. *Every bounded infinite subset of $\mathbb{R}^m$ has an accumulation point.*

Proof. By Theorem 12.11 and Proposition 11.2. $\square$

Remark 12.14. The image⁸ of the sequence $(p_n)_{n \in \mathbb{N}}$ (when viewed as a map $n \mapsto p_n$) is the set

$$S = \{p_n : n \in \mathbb{N}\}.$$

The set S can be finite. For example for the sequence $p_n = (-1)^n$, the set S is the two element set $S = \{-1, 1\}$. If the image of a sequence is finite then there must be at least one constant subsequence and a constant subsequence is trivially convergent. By definition only an infinite set can have an accumulation point.

⁸ Buck calls the set S the **trace** of the sequence, but that terminology is uncommon.

Definition 12.15. A sequence $\{p_n\}$ is called **Cauchy** iff

$$\lim_{m,n \rightarrow \infty} |p_n - p_m| = 0$$

i.e. iff for every $\varepsilon > 0$ there exists $N > 0$ such that $|p_n - p_m| < \varepsilon$ for $n, m \geq N$.

Theorem 12.16 (Cauchy Convergence Criterion). *A sequence in $\mathbb{R}^n$ converges if and only if it is a Cauchy sequence.*

Proof. (Buck Theorem 23 and its corollary on pages 62-63.) □

13 Continuity

Throughout this chapter $f : X \rightarrow Y$ where $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$.

Definition 13.1. The map f is said to be **continuous** at a point $p \in X$ iff for every $\varepsilon > 0$ there exists $\delta > 0$ such that $f(B_X(p, \delta)) \subseteq B(f(p), \varepsilon)$.

Theorem 13.2. *The map f is continuous at $p \in X$ if and only if for every sequence $\{p_n\}$ of points in X we have*

$$\lim_{n \rightarrow \infty} p_n = p \implies \lim_{n \rightarrow \infty} f(p_n) = f(p). \quad (1)$$

Proof. We prove ‘only if’. Assume f is continuous at p . Choose sequence $\{p_n\}$ of points in X . Assume

$$\lim_{n \rightarrow \infty} p_n = p. \quad (2)$$

Choose $\varepsilon > 0$. Because f is assumed to be continuous at p there is a $\delta > 0$ such that for all $q \in X$

$$|q - p| < \delta \implies |f(q) - f(p)| < \varepsilon. \quad (3)$$

By (2) there is an N such that $|p_n - p| < \delta$ for $n > N$. Hence by (3) $|f(p_n) - f(p)| < \varepsilon$ for $n > N$. This proves

$$\lim_{n \rightarrow \infty} f(p_n) = f(p). \quad (4)$$

as required.

We prove ‘if’. Assume that f is not continuous at $p \in X$. Then there is an $\varepsilon > 0$ such that for every $\delta > 0$ there is a $q \in X$ such that

$$|q - p| < \delta \text{ but } |f(q) - f(p)| \geq \varepsilon.$$

In particular, for each $n \in \mathbb{Z}^+$ there is a q_n such that

$$|q_n - p| < \frac{1}{n} \text{ but } |f(q_n) - f(p)| \geq \varepsilon.$$

But then (2) holds but (4) fails. This proves that (1) is false as required. $\square$

Definition 13.3. The map f is said to be **continuous** iff it is continuous at every point of X , i.e. iff

$$\forall p \in X \forall \varepsilon > 0 \exists \delta > 0 \text{ such that } f(B(p, \delta)) \subseteq B(f(p), \varepsilon).$$

The map f is said to be **uniformly continuous** iff

$$\forall \varepsilon > 0 \exists \delta > 0 \text{ such that } \forall p \in X \text{ we have } f(B(p, \delta)) \subseteq B(f(p), \varepsilon).$$

(For continuity $\delta = \delta(p, \varepsilon)$; for uniform continuity $\delta = \delta(\varepsilon)$.)

Proposition 13.4. If $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ are both continuous, then so is the composition $g \circ f : X \rightarrow Z$.

Proof. Choose $p_0 \in X$ and $\varepsilon > 0$. As g is continuous there exists $\eta > 0$ such that $|g(q) - g(f(p_0))| < \varepsilon$ whenever $q \in Y$ and $|q - f(p_0)| < \eta$. As f is continuous, there exists $\delta > 0$ such that $|f(p) - f(p_0)| < \eta$ whenever $p \in X$ and $|p - p_0| < \delta$. For $p \in X$ we have $q = f(p) \in Y$ so

$$|p - p_0| < \delta \implies |f(p) - f(p_0)| < \eta \implies |(g \circ f)(p) - (g \circ f)(p_0)| < \varepsilon$$

as required. $\square$

13.5. The map f is said to be **Lipschitz** iff there is a constant M such that

$$|f(p) - f(q)| \leq M|p - q|$$

for all $p, q \in X$. A Lipschitz function is uniformly continuous. (Proof: $\delta = \varepsilon/M$.)

Problem 13.6. Let $f(x) = x^p$. Show that f is Lipschitz on every closed interval $[a, b] \subseteq (0, \infty)$. For which values of p is f uniformly continuous on $(0, \infty)$? Hint: Use the Mean Value Theorem from calculus. (See Theorem 17.4 below.) Theorem 16.4 below may also help.

14 Open Sets and Closed Sets

In all the following definitions the term *set* means *subset of $\mathbb{R}^m$* .

Definition 14.1. A set U is **open** iff for every $p \in U$ there exists a $\delta > 0$ such that $B(p, \delta) \subseteq U$.

Problem 14.2. Prove that the ball $B(p, \delta)$ is open. Hint: You must choose an arbitrary $q \in B(p, \delta)$ and then find an $\eta > 0$ so that $B(q, \eta) \subseteq B(p, \delta)$. Use the triangle inequality.

Theorem 14.3. *The collection of all open sets in $\mathbb{R}^m$ satisfies the following conditions:*

- (i) *The set $\mathbb{R}^m$ and the empty set $\emptyset$ are both open.*
- (ii) *The intersection of a finite collection of open sets is open.*
- (ii) *The union of an arbitrary collection of open sets is open.*

Proof. The set $\mathbb{R}^m$ is open because $B(p, 1) \subseteq \mathbb{R}^m$ for $p \in \mathbb{R}^m$. The empty set is open because for every $p \in \emptyset$ satisfies the required condition – or any other condition – since ‘*false implies anything*’ is true. See 3.3). To prove (ii) assume U is open. Then for every point $p \in U$ there is a $\delta = \delta_p$ such that $B(p, \delta_p) \subseteq U$.⁹ It follows that

$$U = \bigcup_{p \in U} B(p, \delta_p), \quad (*)$$

i.e. that U is an union of balls. A union of unions is a union:

$$\bigcup_{i \in I} \bigcup_{j \in I_j} B_{ij} = \bigcup_{(i,j) \in K} B_{ij}, \quad K := \{(i, j) : i \in I, j \in I_j\}$$

so (ii) follows. To prove (iii) assume that $U_1, U_2, \dots, U_m$ are open and choose an arbitrary point $p \in \bigcap_{i=1}^m U_i$. Then $p \in U_i$ so there is a $\delta_i > 0$ with $B(p, \delta_i) \subseteq U_i$. Let $\delta = \min(\delta_1, \dots, \delta_m)$. Then

$$B(p, \delta) \subseteq \bigcap_{i=1}^m B(p, \delta_i) \subseteq \bigcap_{i=1}^m U_i$$

as required. □

⁹ This is actually an example of an application of the Axiom of Choice. See Remark 6.5.

Definition 14.4. A set $W \subseteq X$ is called **relatively open** in X iff for every $p \in W$ there exists a $\delta > 0$ such that $B_X(p, \delta) \subseteq W$. (See 10.4.)

Remark 14.5. A set $U \subseteq \mathbb{R}^m$ is open if and only if it is relatively open in $\mathbb{R}^m$. For this reason many theorems can be generalized by systematically replacing $\mathbb{R}^m$ by X , $B(p, \delta)$ by $B_X(p, \delta)$, and the word *open* by the phrase *relatively open in X* . In Math 551 you will learn to generalize further by replacing $\mathbb{R}^n$ by something called a *topological space*.

Corollary 14.6. A set W is relatively open in X if and only if $W = X \cap U$ for some open set $U \subset \mathbb{R}^n$.

Proof. Equation (*) in the last proof and the distributive law from 4.7. $\square$

Proposition 14.7. A map $f : X \rightarrow Y$ is continuous if and only if the inverse image $f^{-1}(V)$ of every relatively open subset V of Y is a relatively open subset of X .

Proof. Exercise. $\square$

Corollary 14.8. Assume that $f : X \rightarrow Y$ and $g : Y \rightarrow Z$ are continuous. Then $g \circ f : X \rightarrow Z$ is continuous.

Proof. Exercise. Hint: Prove that $(g \circ f)^{-1}(W) = f^{-1}(g^{-1}(W))$. $\square$

Definition 14.9. A set X is **closed** iff its complement $\mathbb{R}^n \setminus X$ is open.

Corollary 14.10. The collection of all closed sets in $\mathbb{R}^n$ satisfies the following conditions:

- (i) The set $\mathbb{R}^n$ and the empty set $\emptyset$ of both closed.
- (ii) The intersection of an arbitrary collection of closed sets is closed.
- (iii) The union of a finite collection of closed sets is closed.

Proof. Apply the De Morgan laws from 4.7 to Theorem 14.3. $\square$

Theorem 14.11. A set S is closed if and only if it is closed under limits of sequences, i.e. whenever $\lim_{n \rightarrow \infty} p_n = p$ and each $p_n \in S$ we have $p \in S$.

Proof. (See Buck Theorem 5 page 40.) To prove *only if* assume that S is closed, that $\lim_{n \rightarrow \infty} p_n = p$, and that each $p_n \in S$. If $p \notin S$ then $p \in \mathbb{R}^m \setminus S$. As this set is open there is a $\delta > 0$ such that $B(p, \delta) \subset \mathbb{R}^m \setminus S$. As the sequence converges to p there is an N such that $p_n \in B(p, \delta)$ for $n > N$ contradicting the hypothesis that $p_n \in S$. To prove *if* assume that S is not closed. Then $\mathbb{R}^m \setminus S$ is not open so there is a point $p \in \mathbb{R}^m \setminus S$ such that $B(p, \delta) \not\subset \mathbb{R}^m \setminus S$ for every $\delta > 0$. In particular for $\delta = 1/n$ there is a point $p_n \in B(p, 1/n)$ (i.e. $|p_n - p| < 1/n$) such that $p_n \notin \mathbb{R}^m \setminus S$, i.e. $p_n \in S$. Thus $\lim_{n \rightarrow \infty} p_n = p$ and $p \notin S$ as desired. $\square$

14.12. Let $S \subseteq \mathbb{R}^n$. For any point $p \in \mathbb{R}^n$ exactly one of the following alternatives holds:

- (i) $B(p, \delta) \subseteq S$ for some $\delta > 0$.
- (ii) $B(p, \delta) \subseteq \mathbb{R}^n \setminus S$ for some $\delta > 0$.
- (iii) $B(p, \delta) \cap S \neq \emptyset$ and $B(p, \delta) \cap (\mathbb{R}^n \setminus S) \neq \emptyset$ for all $\delta > 0$.

The **interior** of S is the set $\text{int}(S)$ of all points p where (i) holds, the **exterior** of S is the set $\text{ext}(S)$ of all points p where (ii) holds, and the **boundary** of a set S is the set $\text{bdry}(S)$ of all points p where (iii) holds. The ambient space $\mathbb{R}^n$ may be written as the pairwise disjoint union

$$\mathbb{R}^n = \text{int}(S) \cup \text{ext}(S) \cup \text{bdry}(S).$$

The notations

$$\overset{\circ}{S} := \text{int}(S), \quad \partial S := \text{bdry}(S)$$

are commonly used.

Example 14.13. For the half open interval $S = [a, b) \subseteq \mathbb{R}$ we have

$$\text{int}(S) = (a, b), \quad \text{ext}(S) = (-\infty, a) \cup (b, \infty), \quad \text{bdry}(S) = \{a, b\}.$$

Definition 14.14. A set $S \subseteq \mathbb{R}^n$ is **closed** iff its complement $\mathbb{R}^n \setminus S$ is open. The **closure** of the set S is the set

$$\text{cl}(S) := \bar{S} := S \cup \text{bdry}(S).$$

Proposition 14.15. The interior $\text{int}(S)$ of S is the largest open set contained in S and closure $\bar{S}$ of S is the smallest closed set containing S .

Proof. Exercise. Hint: You must show (1a) $\text{int}(S)$ is open, (1b) $\bar{S}$ is closed, (2a) $\text{int}(S) \subseteq S$, (2b) $S \subseteq \bar{S}$, (3a) if U is open and $U \subseteq S$ then $U \subseteq \text{int}(S)$, and (3b) if T is closed and $S \subseteq T$ then $\bar{S} \subseteq T$. $\square$

Problem 14.16. Prove that if $U \subseteq \mathbb{R}^m$ is an open set and $p \in U$, then there is a point q with rational coordinates and a positive rational number δ such that $p \in B(q, \delta) \subseteq U$.

Solution. As U is open there is an $r > 0$ such that $B(p, r) \subseteq U$. Let p_i denote the i th coordinate of p so that $p = (p_1, p_2, \dots, p_m)$. By Problem 9.10 there are rational numbers $q_1, q_2, \dots, q_m$ such that

$$p_i - \frac{r}{2m} < q_i < p_i + \frac{r}{2m}.$$

Let $q = (q_1, q_2, \dots, q_m)$. By the Triangle Inequality $|p - q| < r/2$. By Problem 9.10 there is rational number δ with $|p - q| < \delta < r/2$. Then $p \in B(q, \delta)$. Choose $x \in B(q, \delta)$. Then $|p - x| < |p - q| + |q - x| < r$ so $x \in B(p, r) \subseteq U$. This shows that $B(q, \delta) \subseteq U$ as required.

15 Connected Sets

Definition 15.1. A set S is **disconnected** iff there are disjoint open sets U and V such that $S \subseteq U \cup V$ and both $S \cap U$ and $S \cap V$ are nonempty. A set is **connected** iff it is not disconnected.

Theorem 15.2. A subset $S \subseteq \mathbb{R}$ of the real line is connected if and only if S is an interval, i.e. $[a, b] \subseteq S$ whenever $a, b \in S$.

Proof. We prove only *if*. Assume S is not an interval, i.e. that there exist $a, b \in S$ with $[a, b] \not\subseteq S$. Then there is a $c \in [a, b]$ with $c \notin S$. Let $U = (-\infty, c)$ and $V = (c, \infty)$. The point c lies in the open interval (a, b) as $a, b \in S$ so $a \in U$ and $b \in V$. Hence both $S \cap U$ and $S \cap V$ are nonempty and clearly $S \subseteq U \cup V$ (as $c \notin S$). Hence the open sets U and V separate S so S is disconnected as required.

We prove *if*. Assume that S is disconnected, i.e. that there exist open sets $U, V \subseteq \mathbb{R}$ with $S \subseteq U \cup V$, $S \cap U \neq \emptyset$, $S \cap V \neq \emptyset$, and $U \cap V = \emptyset$. We must show that S is not an interval. Choose $a \in S \cap U$ and $b \in S \cap V$. Then $a \neq b$ as $U \cap V = \emptyset$. Assume without loss of generality that $a < b$. (The case $b < a$ is the same.)

The set $[a, b] \cap U$ is nonempty (it contains a) and bounded above (b is an upper bound). Let $c = \sup([a, b] \cap U)$. Since $a \in U$ there is an $\varepsilon > 0$ with $(a - \varepsilon, a + \varepsilon) \subseteq U$. Making ε smaller we also have $a + \varepsilon < b$. Therefore $[a, a + \varepsilon] \subseteq [a, b] \cap U$ so $a + \varepsilon = \sup[a, a + \varepsilon] \leq \sup[a, b] \cap U = c$. Since $b \in V$ there is an (other) $\varepsilon > 0$ with $(b - \varepsilon, b + \varepsilon) \subseteq V$. Making ε smaller we also have $a < b - \varepsilon$. Therefore $(b - \varepsilon, b] \subseteq [a, b] \cap V$ so $[b - \varepsilon, b] \cap V = \emptyset$ so $b - \varepsilon$ is an upperbound for $[a, b] \cap U$, so $c \leq b - \varepsilon$. We have proved that $a < c < b$. If $c \in U$ there is an $\varepsilon > 0$ with $a < c - \varepsilon < c < c + \varepsilon < b$ and $(c - \varepsilon, c + \varepsilon) \subseteq U$ contradicting the fact that c is an upper bound of $[a, b] \cap U$. If $c \in V$ there is an $\varepsilon > 0$ with $a < c - \varepsilon < c < c + \varepsilon < b$ and $(c - \varepsilon, c + \varepsilon) \subseteq V$ so $c - \varepsilon$ is an upperbound for $[a, b] \cap U$ contradicting the fact that c is the least upper bound of $[a, b] \cap U$. Hence $c \notin U \cup V$ so (as $S \subseteq U \cup V$) $c \notin S$. Thus $a < c < b$, $a \in S$, $b \in S$, $c \notin S$, so S is not an interval. $\square$

Theorem 15.3. *The continuous image of a connected set is connected: If $f : X \rightarrow \mathbb{R}^m$ is continuous and X is connected, then $f(X)$ is connected.*

Proof. (Buck Theorem 15 on page 94.) $\square$

Corollary 15.4 (Intermediate Value Theorem). *Assume that S is connected and that $f : S \rightarrow \mathbb{R}$ is continuous. Suppose that $a, b \in f(S)$ and that $a < c < b$. Then $c \in f(S)$.*

Proof. (Buck Theorem 14 on page 93.) $\square$

Remark 15.5. The Intermediate Value Theorem from calculus is a special case. It says that if $f : [\alpha, \beta] \rightarrow \mathbb{R}$ is a real valued continuous function on the closed interval $[\alpha, \beta] \subseteq \mathbb{R}$, $\{a, b\} = \{f(\alpha), f(\beta)\}$, and $a \leq c \leq b$, then the equation $f(x) = c$ has a solution $x \in [\alpha, \beta]$.

Theorem 15.6. *A continuous function $f : I \rightarrow \mathbb{R}$ defined on an interval $I \subseteq \mathbb{R}$ is injective if and only if it is strictly monotonic. When these equivalent conditions hold, the image $J = f(I)$ is again an interval and the inverse function is continuous.*

Problem 15.7. Prove Theorem 15.6. (This theorem is proved in Buck Theorem 18 page 96 and Theorem 25 page 114, but Buck assumes that the intervals are closed and bounded. This assumption can be removed.)

Theorem 15.8. *Let $I \subseteq \mathbb{R}$ be an interval and $f : I \rightarrow \mathbb{R}$ be f is continuous. Then the set*

$$\text{graph}(f) := \{(x, y) \in I \times \mathbb{R} : y = f(x)\}$$

is connected.

Proof. Define $F : I \rightarrow \mathbb{R}^2$ by $F(x) = (x, f(x))$ so that $F(I) = \text{graph}(f)$. Clearly f is continuous if and only if F is continuous. We will assume that I is an open interval; the case where I contains one of its endpoints is similar. Assume that $F(I)$ is not connected. Then there are open sets $U, V \subseteq \mathbb{R}^2$ with $F(I) \subseteq U \cup V$, $U \cap V = \emptyset$, $F(I) \cap U \neq \emptyset$, $F(I) \cap V \neq \emptyset$. Then $F^{-1}(U), F^{-1}(V) \subseteq \mathbb{R}$ are open, $I \subseteq F^{-1}(U) \cup F^{-1}(V)$, and $F^{-1}(U) \cap F^{-1}(V) = F^{-1}(U \cap V) = \emptyset$. This contradicts the fact that I is an interval and therefore connected. $\square$

Example 15.9. The converse is false. Consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$f(x) = \begin{cases} \sin(1/x) & \text{if } x > 0, \\ 0 & \text{if } x \leq 0. \end{cases}$$

This function is not continuous as follows. Let $x_n = (2n\pi + \pi/2)^{-1}$. Then $f(x_n) = 1$, $\lim_{n \rightarrow \infty} x_n = 0$, but $\lim_{n \rightarrow \infty} f(x_n) = 1 \neq 0 = f(0)$. However, the graph of f is connected. To see this suppose U and V are open subsets of $\mathbb{R}^2$ and $\text{graph}(f) \subseteq U \cup V$ with $U \cap V = \emptyset$. Suppose that $(0, 0) \in U$. Then $(x, f(x)) \in U$ for $x \leq 0$ as f is continuous on $(-\infty, 0]$ and $(x, f(x)) \in U$ in U for $x > 0$ as f is continuous on $(0, \infty)$. But then $\text{graph}(f) \subset U$ so $\text{graph}(f) \cap V = \emptyset$.

16 Compact Sets

Definition 16.1. An **open cover** of a set S is a collection $(U_\alpha)_{\alpha \in A}$ of open sets such that $S \subseteq \bigcup_{\alpha \in A} U_\alpha$. The subset S is **compact** iff every open cover $(U_\alpha)_{\alpha \in A}$ of S has finite subcover, i.e. there are indices $\alpha_1, \alpha_2, \dots, \alpha_n \in A$ such that $S \subseteq U_{\alpha_1} \cup U_{\alpha_2} \cup \dots \cup U_{\alpha_n}$.

Theorem 16.2 (Heine Borel). *The following are equivalent conditions on a set $S \subseteq \mathbb{R}^m$:*

- (1) *For every sequence $(p_n)_n$ of points of S there is a subsequence $(p_{n_k})_k$ which converges to $p \in S$.*

(2) *The set S closed and bounded.*

(3) *The set S is compact.*

Proof. (This is Theorem 9.2 on page 41 of Morgan. See also Buck Theorem 25 page 65.) The equivalence of (1) and (2) is Corollary 12.12.

We prove (3) $\implies$ (2). Assume that (2) is false. Then either S is not closed or S is not bounded. In the former case by Theorem 14.11 there is a convergent sequence of points $p_n \in S$ whose limit $p := \lim_{n \rightarrow \infty} p_n$ is not in S . This implies $S \subseteq \mathbb{R}^m \setminus \{p\} = \bigcup_{k=1}^{\infty} U_k$ where $U_k := \{q \in \mathbb{R}^m : |q - p| > 1/k\}$. The sets U_k are open but we cannot have $S \subseteq U_1 \cup U_2 \cup \cdots \cup U_N$ as $U_1 \subseteq U_2 \subseteq \cdots \subseteq U_N$ and $p_n \notin U_N$ for $n > N$ so we have an open cover of S with no finite subcover so S is not compact. In the latter case the open balls $V_n := B(0, n) := \{q \in \mathbb{R}^m : |q| < n\}$ cover S as $\mathbb{R}^m = \bigcup_n V_n$ but no finite collection subcollection covers so again S is not compact.

We prove (1) $\implies$ (3). Choose an open cover $(U_\alpha)_{\alpha \in A}$ of S . We first construct a countable subcover. Consider the set

$$I := \{(q, \delta) \in \mathbb{Q}^m \times \mathbb{Q} : \delta > 0 \text{ and } B(q, \delta) \subseteq U_\alpha \text{ for some } \alpha \in A\}$$

of open balls with rational center and rational radius which are contained in some element of the open cover. For every point $p \in S$ there is an $\alpha \in A$ with $p \in U_\alpha$ (because the sets U_α cover S), and so by Problem 14.16 there is a point $q \in \mathbb{Q}^m$ so that $p \in B(q, \delta) \subseteq U_\alpha$. In other words there is a point $(q, \delta) \in I$ with $p \in B(q, \delta)$. We have proved that

$$S \subseteq \bigcup_{(q, \delta) \in I} B(q, \delta).$$

The set I is countable because I is a subset of the countable set $\mathbb{Q}^m \times \mathbb{Q}$. Let $V_1, V_2, V_3, \dots$ be an enumeration of the sets $(B(q, \delta))_{(q, \delta) \in I}$. By construction each V_i is a subset of some U_α . Hence if finitely many of the sets V_i cover S so do finitely many of the sets U_α from the original cover. But (1) implies that finitely many of the sets V_i cover S . If not, we have $S \not\subseteq V_1 \cup V_2 \cup \cdots \cup V_n$ for all n so for each n there is a point $p_n \in S \setminus (V_1 \cup V_2 \cup \cdots \cup V_n)$. By condition (1) the sequence $(p_n)_n$ contains a subsequence $(p_{n_k})_k$ converging to a point $p \in S$. The point p lies in some set V_i and this set must contain p_{n_k} for sufficiently large k which is a contradiction if $n_k > i$. $\square$

Corollary 16.3. *The closed interval $[a, b]$ is compact. (Buck Theorem 24 page 65.)*

Theorem 16.4. *Assume that X is compact and f is continuous. Then f is uniformly continuous.*

Proof. Choose $\varepsilon > 0$. Then, because f is continuous, for every $p \in X$ there is a $\delta = \delta(p) > 0$ such that

$$|q - p| < \delta(p) \implies |f(q) - f(p)| < \frac{\varepsilon}{2}.$$

Let $U_p := B(p, \delta(p)/2)$. The sets U_p cover X (since $p \in U_p$), i.e. $X \subseteq \bigcup_p U_p$. As X is compact, finitely many of these sets cover X , i.e.

$$X \subseteq U_{p_1} \cup U_{p_2} \cup \cdots \cup U_{p_n}. \quad (5)$$

Define

$$\delta := \frac{1}{2} \min\{\delta(p_1), \delta(p_2), \dots, \delta(p_n)\}.$$

Choose $p, q \in X$. Assume $|q - p| < \delta$. By (5) we have that $p \in U_{p_k}$ for some k . Hence

$$|p - p_k| < \delta(p_k)/2. \quad (6)$$

But $\delta \leq \delta(p_k)/2$ by its definition so

$$|q - p_k| \leq |q - p| + |p - p_k| \leq \delta + \frac{\delta(p_k)}{2} \leq \frac{\delta(p_k)}{2} + \frac{\delta(p_k)}{2} = \delta(p_k). \quad (7)$$

Hence, by the definition of $\delta(\cdot)$ and Equations (6) and (7) we have

$$|f(p) - f(q)| \leq |f(p) - f(p_k)| + |f(q) - f(p_k)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

as required. □

Remark 16.5. A proof using the Bolzano-Weierstrass theorem instead on the Heine-Borel theorem is given in Buck page 85. The proof given here is like the proof sketched in Exercises 11 and 12 in Buck page 89.

Theorem 16.6. *The continuous image of a compact set is compact: If $f : X \rightarrow \mathbb{R}^m$ is continuous and X is compact, then $f(X)$ is compact.*

Proof. See Buck Theorem 13 on page 93 or □

Corollary 16.7. *The continuous image of compact set is bounded.*

Proof. See Buck Theorem 10 on page 90 or $\square$

Theorem 16.8. *If $f : X \rightarrow \mathbb{R}$ is continuous and X is compact, then f assumes its maximum on X , i.e. there exists $p \in X$ such that $f(q) \leq f(p)$ for all $q \in X$. Similarly for the minimum.*

Proof. See Buck Theorem 11 page 91 or $\square$

Theorem 16.9. *Let $f : X \rightarrow Y$ be bijective and continuous, and assume X (and hence by Theorem 16.6 also Y) is compact. Then $f^{-1} : Y \rightarrow X$ is continuous.*

Proof. Choose a convergent sequence $(q_n)_n$ in E and a let

$$q := \lim_{n \rightarrow \infty} q_n$$

be its limit. We will show that

$$f^{-1}(q) = \lim_{n \rightarrow \infty} f^{-1}(q_n); \quad (\#)$$

the theorem will then follow by Theorem 13.2. By Bolzano Weierstrass and Heine Borel there is a convergent subsequence $(f^{-1}(q_{n_k}))_k$. Let

$$p := \lim_{k \rightarrow \infty} f^{-1}(q_{n_k})$$

be its limit. Now f is assumed to be continuous so

$$f(p) = f\left(\lim_{k \rightarrow \infty} f^{-1}(q_{n_k})\right) = \lim_{k \rightarrow \infty} f(f^{-1}(q_{n_k})) = \lim_{k \rightarrow \infty} q_{n_k} = q.$$

But $f(p) = q \implies p = f^{-1}(q)$ so $f^{-1}(q) = \lim_{k \rightarrow \infty} f^{-1}(q_{n_k})$. If $(\#)$ fails, then there is a neighborhood U of $f^{-1}(q)$ such that for every N there exists $n > N$ with $f^{-1}(q_n) \notin U$, i.e. there is a subsequence $f^{-1}(q_{m_j})$ with $f^{-1}(q_{m_j}) \notin U$. As before choose a further subsequence (again denoted $f^{-1}(q_{m_j})$) which converges and let

$$p' := \lim_{j \rightarrow \infty} f^{-1}(q_{m_j})$$

denote the limit. Then $p' \notin U$ (else we would have $f^{-1}(q_{m_j}) \in U$ for sufficiently large j) so $p' \neq p$. But as before

$$f(p') = f\left(\lim_{j \rightarrow \infty} f^{-1}(q_{m_j})\right) = \lim_{j \rightarrow \infty} f(f^{-1}(q_{m_j})) = \lim_{j \rightarrow \infty} q_{m_j} = q.$$

But now $f(p) = q = f(p')$ which contradicts the fact that f is injective. $\square$

Theorem 16.10. Let $S \subseteq \mathbb{R}^n$ be and $f : S \rightarrow \mathbb{R}^m$ be uniformly continuous. Then the function f can be continuously extended to the closure $\bar{S}$ of S . i.e. there is a continuous function $F : \bar{S} \rightarrow \mathbb{R}^m$ such that $F(p) = f(p)$ for $p \in S$.

Proof. (Buck Theorem 25 on page 109.) □

Example 16.11. The function $f : (0, 1] \rightarrow \mathbb{R}$ defined by $f(x) = \sin(1/x)$ cannot be extended to a continuous function on the closure $[0, 1]$ of $(0, 1]$.

17 Derivatives

Definition 17.1. The function $f : I \rightarrow \mathbb{R}$ is said to be **differentiable** at the point $x_0 \in I$ iff the limit

$$f'(x_0) := \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

exists; we say that f is differentiable on a set iff it is differentiable at each point x_0 in the set. The function f' is called the **derivative** of f .

Theorem 17.2. A differentiable function is continuous.

Theorem 17.3. If I is an open interval, $f : I \rightarrow \mathbb{R}$ is differentiable on I , and f assumes its maximum (or minimum) at $c \in I$ then $f'(c) = 0$.

Theorem 17.4 (Mean Value Theorem). Suppose that f is differentiable of (a, b) and continuous on $[a, b]$. Then there exists $c \in (a, b)$ such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Corollary 17.5. Assume that f is differentiable on I . Then the derivative f' vanishes identically on I if and only if f is constant on I .

Problem 17.6. *The Intermediate Value Theorem for Derivatives.* Assume that f is differentiable on an open interval I , that $a, b \in I$ with $a < b$, and that $f'(a) < w < f'(b)$. Show that there is a c with $a < c < b$ and $f'(c) = w$. Hint: Consider the function $g(x) = f(x) - wx$.

18 The Integral

18.1. A **partition** of the closed interval $[a, b]$ is an increasing finite sequence $P = (x_k)_{0 \leq k \leq n}$ with $x_0 = a$ and $x_n = b$. For any bounded function f defined on $[a, b]$ and any partition $P = (x_k)_{0 \leq k \leq n}$ of $[a, b]$ define the **upper sum** by $\overline{S}(f, P)$ and the **lower sum** $\underline{S}(f, P)$ by

$$\overline{S}(f, P) := \sum_{k=1}^n \overline{y}_k (x_k - x_{k-1}), \quad \underline{S}(f, P) := \sum_{k=1}^n \underline{y}_k (x_k - x_{k-1}),$$

where we have used the abbreviations

$$\overline{y}_k := \sup\{f(x) : x_{k-1} \leq x \leq x_k\}, \quad \underline{y}_k := \inf\{f(x) : x_{k-1} \leq x \leq x_k\},$$

Theorem 18.2. *Assume that $f : [a, b] \rightarrow \mathbb{R}$ is continuous. Then there is a unique number $\int_a^b f$ called the **definite integral** of f on the interval $[a, b]$ such that the inequality*

$$\underline{S}(f, P) \leq \int_a^b f \leq \overline{S}(f, P)$$

holds for every partition P of the interval $[a, b]$.

Proof. We say that a partition P of $[a, b]$ **refines** the partition Q of $[a, b]$ iff Q is a subsequence of P , i.e. iff $P = (x_k)_{0 \leq k \leq n}$, $Q = (x_{k_j})_{0 \leq j \leq m}$ where $0 = k_0 < k_1 < \dots < k_m = n$. The **mesh** of a partition $P = (x_k)_{0 \leq k \leq n}$ to be the maximum of the positive numbers $\Delta_k := x_k - x_{k-1}$, $k = 1, 2, \dots, n$. We prove the theorem in five steps.

Step 1. *For any partition P of $[a, b]$ we have*

$$\underline{y}(b-a) \leq \underline{S}(f, P) \leq \overline{S}(f, P) \leq \overline{y}(b-a)$$

where $\underline{y}$ is the infimum of $f(x)$ for $x \in [a, b]$ and $\overline{y}$ is the supremum. The middle inequality follows the inequality $\underline{y}_k \leq \overline{y}_k$ which in turn is an immediate consequence of the fact that the infimum of a bounded nonempty set is less than or equal to the supremum of that set. For the inequality on the left note that

$$\underline{y}(x_k - x_{k-1}) \leq \underline{y}_k(x_k - x_{k-1})$$

since the infimum on the left is over a larger set. From the “collapsing sum”

$$b - a = (x_1 - x_0) + (x_2 - x_1) + \cdots + (x_k - x_{k-1})$$

we obtain the inequality $\underline{y}(b-a) \leq \underline{S}(y, P)$ by summing on k . The inequality $\overline{S}(f, P) \leq \overline{y}(b-a)$ is proved similarly.

Step 2. *If the partition P of $[a, b]$ refines the partition Q of $[a, b]$, then*

$$\underline{S}(f, Q) \leq \underline{S}(f, P) \leq \overline{S}(f, P) \leq \overline{S}(f, Q).$$

Since $x_{k_{j-1}} < x_{k_{j-1}+1} < x_{k_{j-1}+2} < \cdots < x_{k_j}$ the partition P determines a partition P_j of the interval $[x_{k_{j-1}}, x_{k_j}]$. Applying Step 1 to this partition gives

$$\underline{y}_{k_j}(x_{k_j} - x_{k_{j-1}}) \leq \underline{S}(f, P_j) \leq \overline{S}(f, P_j) \leq \overline{y}(x_{k_j} - x_{k_{j-1}}).$$

Step 2 follows by summing on j .

Step 3. *For every $\varepsilon > 0$ there is a $\delta > 0$ such that for every partition P whose mesh is less than δ we have*

$$\overline{S}(f, P) < \underline{S}(f, P) + \varepsilon.$$

To see this choose $\varepsilon > 0$. By Theorem 16.4 there is a $\delta > 0$ such that $|f(x) - f(x')| < \varepsilon/(b-a)$ whenever $|x - x'| < \delta$. If the partition P has mesh less than δ it follows that $\overline{y}_k \leq \underline{y}_k + \varepsilon/(b-a)$ for $k = 1, 2, \dots, n$. The desired inequality follows by multiplying by $x_k - x_{k-1}$ and summing on k . (Use the collapsing sum from Step 1.)

Step 4. *Given any two partitions of the interval $[a, b]$ there is a third partition which refines both of them.* To construct this third partition we simply take the union of the underlying sets of the two partitions and list the elements of this union in increasing order.

Step 6. $\underline{S}(f, P') \leq \overline{S}(f, P'')$ for any two partitions P' and P'' of the interval $[a, b]$. This is because by Step 4 the partitions P' and P'' have a common refinement P so Step 2 we have

$$\underline{S}(f, P') \leq \underline{S}(f, P) \leq \overline{S}(f, P) \leq \overline{S}(f, P'').$$

Step 7. $\sup_P \underline{S}(f, P) \leq \inf_P \overline{S}(f, P)$. To see this take the supremum over P' in Step 4 to get $\sup_P \underline{S}(f, P) \leq \overline{S}(f, P'')$ for every P'' . Then take the infimum over P'' .

Step 8. $\inf_P \overline{S}(f, P) \leq \sup_P \underline{S}(f, P')$. To see this choose $\varepsilon > 0$ and let $\delta > 0$ be given as in Step 3. There certainly are partitions with mesh less than δ , for example the partition $P_n = (x_k)_k$ with $x_k = a + k(b-a)/n$ where n is so large that $(b-a)/n < \delta$. From Step 3 we conclude that

$$\inf_P \overline{S}(f, P) \leq \overline{S}(f, P_n) \leq \underline{S}(f, P_n) + \varepsilon \leq \sup_P \underline{S}(f, P) + \varepsilon.$$

Step 7 follows since this inequality is true for any $\varepsilon > 0$.

By Steps 6 and 7 we have that $\sup_P \underline{S}(f, P) = \inf_P \overline{S}(f, P)$. This common value satisfies the conclusion of the theorem and is the definite integral $\int_a^b f$. $\square$

Remark 18.3. A **Riemann sum** for a partition $P = (x_k)_k$ of the interval $[a, b]$ is a sum of form

$$S = \sum_{k=1}^n f(c_k)(x_k - x_{k-1})$$

where $c_k \in [x_{k-1}, x_k]$. Since $\underline{f}_k \leq f(c_k) \leq \overline{f}_k$ it follows that

$$\underline{S}(f, P) \leq S \leq \overline{S}(f, P)$$

for any Riemann sum for the partition P . It follows from the proof of Theorem 18.2 that the definite integral is the limit of the Riemann sums S as the mesh of the partition P tends to zero, i.e. for every $\varepsilon > 0$ there exists a $\delta > 0$ such that $|S - \int_a^b f| < \varepsilon$ whenever the mesh of P is less than δ . In Math 221 the definite integral is usually defined as the limit of Riemann sums in this sense.

Theorem 18.4. *The definite integral of a continuous function satisfies the following properties.*

- (1) (Normalization). $\int_a^b 1 = b - a$.
- (2) (Linearity). $\int_a^b (f + g) = \int_a^b f + \int_a^b g$
- (3) (Linearity). $\int_a^b cf = c \int_a^b f$ for $c \in \mathbb{R}$.

$$(4) \text{ (Additivity). } \int_a^b f + \int_b^c f = \int_a^c f.$$

$$(5) \text{ (Order). If } f(x) \leq g(x) \text{ for all } x \in [a, b], \text{ then } \int_a^b f \leq \int_a^b g.$$

$$(6) \text{ (Triangle Inequality). } \left| \int_a^b f \right| \leq \int_a^b |f|.$$

In (3) c is any real number, and in (4) $a, b, c \in I$ satisfy $a \leq b \leq c$.¹⁰

Proof. These properties hold because Riemann sums satisfy similar properties:

$$(1) \quad \sum_{k=1}^n \Delta_k = b - a.$$

$$(2) \quad \sum_{k=1}^n (f(c_k) + g(c_k)) \Delta_k = \sum_{k=1}^n f(c_k) \Delta_k + \sum_{k=1}^n g(c_k) \Delta_k$$

$$(3) \quad \sum_{k=1}^n c f(c_k) \Delta_k = c \sum_{k=1}^n f(c_k) \Delta_k$$

$$(4) \quad \sum_{k=1}^n f(c_k) \Delta_k + \sum_{k=n+1}^m f(c_k) \Delta_k = \sum_{k=1}^m f(c_k) \Delta_k$$

$$(5) \quad \sum_{k=1}^n f(c_k) \Delta_k \leq \sum_{k=1}^n g(c_k) \Delta_k \text{ if } f(c_k) \leq g(c_k)$$

$$(6) \quad \left| \sum_{k=1}^n f(c_k) \Delta_k \right| \leq \sum_{k=1}^n |f(c_k)| \Delta_k$$

In these formulas $(x_k)_{k=0}^n$ is a partition of $[a, b] = [x_0, x_n]$ (in (4) it is extended to a partition $(x_k)_{k=0}^m$ of $[a, c]$), c_k is in the k th interval (i.e. $x_{k-1} \leq c_k \leq x_k$), and we have used the abbreviation

$$\Delta_k := x_k - x_{k-1}.$$

¹⁰It is customary to define $\int_a^b f = -\int_b^a f$ if $b < a$. With this definition the additivity formula holds without the restriction that $a \leq b \leq c$.

These formulas and Remark 18.3 imply the theorem by taking the limit as the mesh of the partition goes to zero. $\square$

Theorem 18.5 (Fundamental Theorem of Calculus). *Assume that I is an open interval, that $f : I \rightarrow \mathbb{R}$ is continuous, that $a \in I$, and that $F(x)$ is defined by*

$$F(x) := \int_a^x f.$$

Then

(I) *F is differentiable on I and its derivative is f and hence*

(II) *$\int_a^b f = F(b) - F(a)$ for $a, b \in I$ with $b \geq a$.*

Proof. We must show that for every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$-\varepsilon < \frac{F(x+h) - F(x)}{h} - f(x) < \varepsilon$$

whenever $0 < |h| < \delta$. We will only use the properties in Theorem 18.4 to prove this. Assume $h > 0$ (the case $h < 0$ is similar). Then

$$F(x+h) - F(x) = \int_a^{x+h} f - \int_a^x f = \int_x^{x+h} f$$

by property (4). Choose $\varepsilon > 0$. By Theorem 16.4 f is uniformly continuous so there is a $\delta > 0$ such that $0 < h < \delta$ implies that

$$f(x) - \varepsilon < f(t) < f(x) + \varepsilon$$

for all t in the interval $[x, x+h]$. This implies that

$$(f(x) - \varepsilon)h < \int_x^{x+h} f < (f(x) + \varepsilon)h$$

by (1), (3), and (5). Dividing by h and subtracting $f(x)$ and gives

$$-\varepsilon < \left(\frac{1}{h} \int_x^{x+h} f \right) - f(x) < \varepsilon.$$

whenever $0 < h < \delta$, i.e.

$$-\varepsilon < \frac{F(x+h) - F(x)}{h} - f(x) < \varepsilon$$

as required. $\square$

Remark 18.6. Henceforth we use the more traditional notation $\int_a^b f(x) dx$ (rather than the notation $\int_a^b f$) for the integral. The reader is reminded that the variable x in this expression is a **dummy variable**, i.e.

$$\int_a^b f(x) dx = \int_a^b f(t) dt.$$

19 Taylor's Formula

Theorem 19.1 (Taylor's Formula – Lagrange Form). *Let I be an interval, $a \in I$, and $f : I \rightarrow \mathbb{R}^m$ be of class C^{n+1} . Then*

$$f(x) = P_n(a, x) + R_n(a, x)$$

for any $x \in I$ where

$$P_n(a, x) := \sum_{k=0}^n \frac{f^{(k)}(a)}{k!} (x-a)^k, \quad R_n(a, x) := \int_a^x \frac{f^{(n+1)}(t)}{n!} (x-t)^n dt.$$

The polynomial $P_n(a, x)$ is called the **Taylor polynomial** of degree n at a and $R_n(a, x)$ is called the n th **Taylor remainder**.

Proof. By induction on n . For $n = 0$ this is the Fundamental Theorem of Calculus. We assume the formula for n and integrate by parts, viewing x as a constant and t as a variable:

$$\begin{aligned} v &= -\frac{(x-t)^{n+1}}{(n+1)!}, & u &= f^{(n+1)}(t), \\ dv &= \frac{(x-t)^n}{n!} dt, & du &= f^{(n+2)}(t) dt, \\ R_n(a, x) &= \int_a^x u dv = uv \Big|_a^x - \int_a^x v du \\ &= \frac{f^{(n+1)}(a)}{(n+1)!} (x-a)^{n+1} + \int_a^x \frac{(x-t)^{n+1}}{(n+1)!} f^{(n+2)}(t) dt \\ &= \frac{f^{(n+1)}(a)}{(n+1)!} (x-a)^{n+1} + R_{n+1}(a, x). \end{aligned}$$

Adding $P_n(a, c)$ to both sides gives $f(x) = P_{n+1}(a, x) + R_{n+1}(a, x)$. $\square$

Corollary 19.2. If $|f^{(n+1)}(t)| \leq M$ for $t \in I$ then

$$|R_n(a, x)| \leq \frac{M|x - a|^{n+1}}{(n+1)!}$$

for $x \in I$.

Proof. Assume that $x > a$. (The case $x < a$ is similar.)

$$\begin{aligned} |R_n(a, x)| &= \left| \int_a^x \frac{f^{(n+1)}(t)}{n!} (x-t)^n dt \right| \\ &\leq \int_a^x \left| \frac{f^{(n+1)}(t)}{n!} (x-t)^n \right| dt \\ &\leq \int_a^x M \frac{(x-t)^n}{n!} dt = M \frac{(x-a)^{n+1}}{(n+1)!}. \quad \square \end{aligned}$$

Remark 19.3. In Math 222 it is shown that there is a number c between a and x such that

$$R_n(a, x) = \frac{f^{(n+1)}(c)}{(n+1)!} (x-a)^{n+1}.$$

This form of the remainder has the advantage that it is easy to remember: the remainder is the next term in the series with $f^{(n+1)}(a)$ replaced by $f^{(n+1)}(c)$. However, this version of the theorem only holds when f is real valued, i.e. when $m = 1$. For $m > 1$ there will be a different value of c for each component of f .

20 Series

20.1. A sequence determines a series and a series determines a sequence. More precisely, a sequence $(a_k)_k$ determines a **series** whose **partial sums** are

$$S_n = \sum_{k=1}^n a_k := a_1 + a_2 + \cdots + a_n,$$

and the **terms** of the series may be recovered from the sequence of partial sums via the formula

$$a_n = S_n - S_{n-1}, \quad a_1 = S_1.$$

Convergence of the series is synonymous with convergence of the sequence of partial sums:

$$\sum_{k=1}^{\infty} a_k := \lim_{n \rightarrow \infty} \sum_{k=1}^n a_k.$$

Since the limit of a difference is the difference of the limits and the limit of a constant is the constant we have the useful formula

$$\sum_{k=1}^{\infty} a_k - \sum_{k=1}^n a_k = \sum_{k=n+1}^{\infty} a_k.$$

Here the notation on the right has the obvious definition, namely

$$\sum_{k=n+1}^{\infty} a_k := \lim_{m \rightarrow \infty} \sum_{k=n+1}^m a_k.$$

20.2. The notation

$$\sum_{k=1}^{\infty} a_k = \infty$$

means that for every $M > 0$ there exists an integer $N > 0$ such that $\sum_{k=1}^n a_k > M$ for $n > N$. If $a_k \geq 0$ for all k then the sequence of partial sums is monotonic increasing so by Theorem 12.7 either the limit $\sum_{k=1}^{\infty} a_k$ exists (i.e. the sequence of partial sums is bounded) or $\sum_{k=1}^{\infty} a_k = \infty$ (i.e. the sequence of partial sums is unbounded).

Example 20.3. The n th partial sum

$$\sum_{k=0}^n x^k = 1 + x + x^2 + \cdots + x^n$$

of the **geometric series** is easy to compute:

$$(1 - x) \sum_{k=0}^n x^k = \sum_{k=0}^n (x^k - x^{k+1}) = 1 - x^{n+1}$$

(as the sum telescopes) so dividing by $(1 - x)$ gives

$$\sum_{k=0}^n x^k = \frac{1 - x^{n+1}}{1 - x}$$

Hence if $|x| < 1$ we have the formula

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}$$

for the infinite sum.

Theorem 20.4 (nth Term Test). *If the series $\sum_k a_k$ converges, then the n th term converges to zero.*

Proof. (See Buck Theorem 2 page 230.) Let $S_n = \sum_{k=0}^n a_k$ be the n th partial sum so $a_n = S_n - S_{n-1}$. If the series converges then $\lim_{n \rightarrow \infty} S_n = \lim_{n \rightarrow \infty} S_{n-1}$ so $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} S_n - \lim_{n \rightarrow \infty} S_{n-1} = 0$. $\square$

Example 20.5. (Harmonic series) Since

$$\sum_{k=1}^{\infty} \frac{1}{k} = 1 + \left(\frac{1}{2} + \cdots + \frac{1}{10} \right) + \left(\frac{1}{11} + \cdots + \frac{1}{100} \right) + \cdots$$

and each sum in parentheses is $\geq 9/10$ we see that $\sum_{k=1}^{\infty} \frac{1}{k} = \infty$. But $\lim_{n \rightarrow \infty} 1/n = 0$ so the converse to Theorem 20.4 is false.

Theorem 20.6 (Cauchy Convergence Criterion for Series). *A series $\sum_k a_k$ converges if and only if*

$$\lim_{m, n \rightarrow \infty} \sum_{k=m+1}^n a_k = 0,$$

i.e. for every $\varepsilon > 0$ there exists $N = N(\varepsilon)$ such that $|\sum_{k=m+1}^n a_k| < \varepsilon$ for $n > m > N$.

Proof. This follows immediately from the formulas

$$\sum_{k=m+1}^n a_k = S_n - S_m, \quad S_n := \sum_{k=1}^n a_k, \quad \sum_{k=m+1}^n a_k := a_{m+1} + \cdots + a_n.$$

and the Cauchy Convergence Criterion for Sequences (Theorem 12.16). $\square$

Definition 20.7. The series $\sum_k a_k$ is said to **converge absolutely** iff

$$\sum_{k=1}^{\infty} |a_k| < \infty.$$

A series which converges but does not converge absolutely is said to **converge conditionally**.

Theorem 20.8. *If a series converges absolutely, then it converges.*

Proof. This is an immediate consequence of the inequality

$$\left| \sum_{k=m+1}^n a_k \right| \leq \sum_{k=m+1}^n |a_k|$$

and Theorem 20.6. □

Remark 20.9. Of course, if $a_k \geq 0$ then $|a_k| = a_k$ so convergence and absolute convergence coincide in this case. When $a_k \geq 0$ the sequence of partial sums is monotonic increasing so either the series converges or diverges to infinity.

Corollary 20.10. *Assume that $a_0 \geq a_1 \geq a_2 \cdots \geq 0$. Then the series $\sum_k (-1)^k a_k$ converges if and only if $\lim_{n \rightarrow \infty} a_n = 0$. (A series like this whose terms alternate sign is called an **alternating series**.)*

Proof. “Only if” follows by the n th Term Test (Theorem 20.4). For “if” assume that $\lim_{n \rightarrow \infty} a_n = 0$ and let

$$S_n := \sum_{k=0}^n (-1)^k a_k$$

denote the n th partial sum. Then the sequence

$$S_{2m+1} = \sum_{k=0}^m (a_{2k} - a_{2k+1})$$

is monotonic non decreasing (since $(a_{2k} - a_{2k+1}) \geq 0$) so it either converges or tends to infinity. But the latter cannot happen as

$$S_{2m+1} = a_0 - \sum_{k=0}^{m-1} (a_{2k+1} - a_{2k+2}) - a_{2m+1} \leq a_0 - a_{2m+1}$$

and the sequence $(a_{2m+1})_m$ is bounded (as it converges to zero). Hence the sequence $(S_{2m+1})_m$ converges by Theorem 12.7. But the sequence $(S_{2m})_m$ converges to the same limit as $S_{2m+1} = S_{2m} + a_{2m+1}$ and we have assumed that $\lim_{n \rightarrow \infty} a_n = 0$. Hence the sequence $(S_n)_n$ converges as claimed. $\square$

Example 20.11. (Alternating harmonic series) We will see later (see Remark 24.3) that

$$-\ln 2 = \sum_{k=1}^{\infty} \frac{(-1)^k}{k}.$$

The convergence is conditional but (by Example 20.5) not absolute.

Theorem 20.12 (Comparison Test). *If $0 \leq a_k \leq b_k$ for sufficiently large k and the series $\sum_k b_k$ converges, then the series $\sum_k a_k$ does.*

Proof. (Buck Theorem 5 page 231.) The partial sums satisfy the inequality

$$0 \leq \sum_{k=m+1}^n a_k \leq \sum_{k=m+1}^n b_k.$$

Because the $\sum_k b_k$ is assumed to converge we have $\lim_{m,n \rightarrow \infty} \sum_{k=m+1}^n b_k = 0$. Hence $\lim_{m,n \rightarrow \infty} \sum_{k=m+1}^n a_k = 0$ so the series $\sum_k a_k$ converges by Theorem 12.16 $\square$

Theorem 20.13 (Integral Test). *Assume that $a_k = f(k)$ where $f : [1, \infty) \rightarrow [0, \infty)$ is monotonic decreasing. Then the improper integral $\int_1^{\infty} f(x) dx < \infty$ converges if and only if the sum $\sum_{k=1}^{\infty} a_k < \infty$ does.*

Proof. (Buck Theorem 10 page 233.) We have $a_{k+1} = f(k+1) \leq f(x) \leq f(k) = a_k$ for $k \leq x \leq k+1$ so

$$\sum_{k=2}^{n+1} a_k \leq \int_1^n f(x) dx \leq \sum_{k=1}^n a_k.$$

This inequality shows that the infinite sum is finite if and only if the integral is finite. $\square$

Theorem 20.14 (Root Test). *Let*

$$R = \limsup_{k \rightarrow \infty} |a_k|^{1/k}$$

Then the series $\sum_k a_k$ converges absolutely if $R < 1$ and diverges (does not converge) if $R > 1$.

Proof. (Buck Theorem 9 page 232.) Let $S_n = \{|a_k|^{1/k} : k = n, n+1, \dots\}$, $s_n = \sup S_n$ so $S_{n+1} \subseteq S_n$ and hence $s_{n+1} \leq s_n$. (We are using the convention that $\sup S = \infty$ if S is not bounded above.) Either $s_n = \infty$ for all n (this falls under the case $R > 1$) or else $(s_n)_n$ is a monotonic sequence decreasing (i.e. non-increasing) converging to R . If $R < 1$ then $R < r < 1$ where $r = (1 + R)/2$ so there is an N such that $s_n < r$ for $n > N$ and hence $|a_k|^{1/k} \leq s_n < r$ for $k \geq n > N$. From this we deduce that $|a_k| \leq r^k$ and hence

$$\sum_{k=N+1}^{\infty} |a_k| \leq \sum_{k=N+1}^{\infty} r^k$$

so the series converges by the Comparison Test (Theorem 20.12, see also) and the fact that the geometric series converges for $r < 1$. Conversely if $R > 1$ then $s_n \geq R > 1$ for all n so for every n there exists a $k \geq n$ with $|a_k| = (|a_k|^{1/k})^k \geq 1$. This means that it is *not* the case that $\lim_{n \rightarrow \infty} a_n = 0$ so the series diverges by the n th term test (Theorem 20.4). $\square$

Remark 20.15. Theorem 20.14 gives no information when $R = 1$. If $a_k = 1/k$, then $R = 1$ but $\sum_k a_k = \infty$. If $a_k = 1/k^2$, then $R = 1$ and $\sum_k a_k < \infty$.

Remark 20.16. The proofs the convergence tests tell us how to estimate the error, i.e. the difference between a partial sum and the infinite sum. For example, by the Integral Test, the series $\sum_k 1/k^2$ converges and

$$\left| \sum_{k=1}^{\infty} \frac{1}{k^2} - \sum_{k=1}^n \frac{1}{k^2} \right| = \left| \sum_{k=n+1}^{\infty} \frac{1}{k^2} \right| \leq \int_n^{\infty} \frac{dx}{x^2} = \frac{1}{n}.$$

Similarly if $R < 1$ in the Root Test and $R < r < 1$ there is an N such that $|a_k| < r^k$ for $k > N$ and hence

$$\left| \sum_{k=1}^{\infty} a_k - \sum_{k=1}^n a_k \right| = \sum_{k=n+1}^{\infty} |a_k| \leq \sum_{k=n+1}^{\infty} r^k = \frac{r^{n+1}}{1-r} \leq \frac{r^N}{1-r}$$

for $n > N$.

Problem 20.17. Show that for $p > 1$ the series $\sum_{k=1}^{\infty} k^{-p}$ converges and that the estimate

$$\left| \sum_{k=1}^{\infty} k^{-p} - \sum_{k=1}^n k^{-p} \right| \leq \frac{(n+1)^{1-p}}{p-1}$$

holds for the difference (error) between the n th partial sum and the limit.

Definition 20.18. A series $\sum_{k=1}^{\infty} b_k$ is said to be a **rearrangement** of the series $\sum_{k=1}^{\infty} a_k$ iff there is a permutation¹¹ $\sigma : \mathbb{Z}^+ \rightarrow \mathbb{Z}^+$ such that $b_k = a_{\sigma(k)}$.

Theorem 20.19. (1) *Any rearrangement of an absolutely convergent series converges absolutely to the same limit.*

(2) *Assume that the series $\sum_{k=1}^n a_k$ converges conditionally and $L \in \hat{\mathbb{R}}$ (see 11.4). Then there is a rearrangement $\sum_{k=1}^{\infty} b_k$ of the series $\sum_{k=1}^n a_k$ such that*

$$\sum_{k=1}^{\infty} b_k = L.$$

Proof. Part (1) is Theorem 13 page 239 of Buck. Part (2) is proved on pages 238-9 of Buck in the special case where $a_n = (-1)^n/n$ and $L = 10$; the general argument is much the same but uses Theorem 20.4. $\square$

21 Uniform Convergence

Definition 21.1. A sequence $(f_n)_n$ of functions with common domain X said to **converge pointwise** to the function $f : X \rightarrow \mathbb{R}$ iff

$$\lim_{n \rightarrow \infty} f_n(p) = f(p)$$

for all $p \in X$, i.e. iff

$$\forall p \in X \forall \varepsilon > 0 \exists N \forall n [n > N \implies |f_n(p) - f(p)| < \varepsilon].$$

The sequence is said to **converge uniformly** to the function f iff

$$\lim_{n \rightarrow \infty} \sup_{p \in U} |f_n(p) - f(p)| = 0,$$

i.e. iff

$$\forall \varepsilon > 0 \exists N \forall p \in U \forall n [n > N \implies |f_n(p) - f(p)| < \varepsilon].$$

For a sequence $(u_k : X \rightarrow \mathbb{R}^m)_k$ of functions the series $\sum_k u_k$ of functions is said to converge pointwise or uniformly iff the sequence $f_n = \sum_{k=0}^n u_k$ of partial sum does.

¹¹ A **permutation** is bijective map from a set itself. This terminology is most often used for finite sets, but here it is used for a infinite set.

Example 21.2. Define $f_n : [0, 1] \rightarrow \mathbb{R}$ by $f_n(x) = x^n$. Then the sequence $(f_n)_n$ converges pointwise but not uniformly to the function

$$f(x) = \begin{cases} 0 & \text{for } 0 \leq x < 1 \\ 1 & \text{for } x = 1. \end{cases}$$

Theorem 21.3. *If $f_n : X \rightarrow Y$ is continuous for each n and the sequence $(f_n)_n$ converges uniformly to $f : X \rightarrow Y$, then the limit f is also continuous.*

Proof. (Buck Theorem 3 page 266.) Choose x_0 and $\varepsilon > 0$. By uniform convergence there exists an n such that $|f(x) - f_n(x)| < \varepsilon/3$ for all x and all $n > N$. Let $n = N + 1$. As f_n is continuous at x_0 there exists $\delta > 0$ such that $|f_n(x) - f_n(x_0)| < \varepsilon/3$ whenever $|x - x_0| < \delta$. Then

$$|f(x) - f(x_0)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(x_0)| + |f_n(x_0) - f(x_0)| < \varepsilon$$

whenever $|x - x_0| < \delta$. □

Theorem 21.4 (Weierstrass Comparison Test). *Assume that the functions $u_k : X \rightarrow \mathbb{R}^m$ satisfy $|u_k(p)| \leq M_k$ where $\sum_k M_k < \infty$. Then the series $\sum_k u_k$ converges uniformly.*

Proof. (Buck Theorem 2 page 266.) This is an immediate consequence of the inequality

$$\left| \sum_{k=n+1}^{\infty} u_k(p) \right| \leq \sum_{k=n+1}^{\infty} |u_k(p)| \leq \sum_{k=n+1}^{\infty} M_k$$

as follows. Since the series on the right converges we have that for every $\varepsilon > 0$ there is an $N = N(\varepsilon)$ such that the right hand side is $< \varepsilon$ if $n > N$ and hence the left hand side is $< \varepsilon$ for all p . □

Remark 21.5. The proof doesn't require that the inequality $|u_k(p)| \leq M_k$ hold for *all* k but only for all sufficiently large k . This inequality is often expressed by saying that the series $\sum_k u_k$ is **dominated** by the series $\sum_k M_k$.

Theorem 21.6. *Assume that the sequence $(f_n : [a, b] \rightarrow \mathbb{R})_n$ converges uniformly to a function f and that each f_n is continuous. Then the limit of the integrals is the integral of the limit, i.e.*

$$\lim_{n \rightarrow \infty} \int_a^b f_n(x) dx = \int_a^b f(x) dx.$$

Proof. Choose $\varepsilon > 0$. Then there is an N such that $|f(x) - f_n(x)| < \varepsilon/(b-a)$ for all $x \in [a, b]$ and all $n > N$. For $n > N$ we have, by the various properties listed in Theorem 18.4, that

$$\begin{aligned} \left| \int_a^b f(x) dx - \int_a^b f_n(x) dx \right| &= \left| \int_a^b (f(x) - f_n(x)) dx \right| \\ &\leq \int_a^b |f(x) - f_n(x)| dx \\ &\leq \frac{\varepsilon(b-a)}{b-a} = \varepsilon \end{aligned}$$

as required. $\square$

Corollary 21.7. *Let $(f_n)_n$ be a sequence of functions defined on an open interval I . Assume that*

- (i) *Each f_n is differentiable;*
- (ii) *Each derivative f'_n is continuous;*
- (iii) *The sequence $(f_n)_n$ converges uniformly on I ;*
- (iv) *The sequence $(f'_n)_n$ converges uniformly on I .*

Then the limit f is differentiable and the limit of the derivatives is the derivative of the limit, i.e.

$$\lim_{n \rightarrow \infty} f'_n(x) = f'(x)$$

for $x \in I$.

Proof. Define $g(x) := \lim_{n \rightarrow \infty} f'_n(x)$. By Theorem 21.3 g is continuous. By the Fundamental Theorem of Calculus part (II) we have

$$f_n(x) - f_n(a) = \int_a^x f'_n(t) dt$$

so Theorem 21.6 we have

$$f(x) - f(a) = \lim_{n \rightarrow \infty} \int_a^x f'_n(t) dt = \int_a^x \lim_{n \rightarrow \infty} f'_n(t) dt = \int_a^x g(t) dt.$$

Hence $f' = g$ by the Fundamental Theorem of Calculus part (I). $\square$

22 Power Series - I

22.1. A series of form $\sum_{k=0}^{\infty} c_k(x-a)^k$ is called a **power series** centered at a . The **radius of convergence** of the power series is the number R defined by

$$\frac{1}{R} := \limsup_{k \rightarrow \infty} |c_k|^{1/k}.$$

(If the lim sup is infinite, then $R := 0$ and if the lim sup is zero, then $R := \infty$.)

Problem 22.2. (A formula for the radius of convergence). Assume that the coefficients c_k are nonzero. Show that

$$\limsup_{k \rightarrow \infty} |c_k|^{1/k} = \lim_{k \rightarrow \infty} \frac{|c_{k+1}|}{|c_k|}$$

if the limit on the right exists.

Theorem 22.3. Let $\sum_{k=0}^{\infty} c_k(x-a)^k$ be a power series and R be its radius of convergence. Then the series converges if $|x-a| < R$ and diverges if $|x-a| > R$. More precisely,

- (i) If $0 < r < R$ the series converges uniformly on the interval $[a-r, a+r]$.
- (ii) If $|x-a| > R$ then the n th term of the series is unbounded and hence does not converge to zero (so the series does not converge by Theorem 20.4).

Proof. (Buck Theorem 14 page 240.) Let $a_k = c_k(x-a)^k$. Then by the Root Test (Theorem 20.14) the series converges if $\limsup_{k \rightarrow \infty} |a_k|^{1/k}$ is less than one and diverges if it is greater than one. But

$$\limsup_{k \rightarrow \infty} |a_k|^{1/k} = \limsup_{k \rightarrow \infty} |c_k|^{1/k} |x-a| = \frac{|x-a|}{R}$$

so the series converges if $|x-a| < R$ and diverges if $|x-a| > R$. To prove that the convergence is uniform on the interval $[a-r, a+r]$ we need to repeat the proof of the Root Test. Let $\rho = (r+R)/2$ so $0 < r < \rho < R$. As $\limsup_k |c_k|^{1/k} = 1/R < 1/\rho$ it follows that $|c_k| < (1/\rho)^k$ for sufficiently large k (say $k > N$) so

$$|c_k(x-a)^k| < \left(\frac{r}{\rho}\right)^k.$$

Since $r/\rho < 1$ this says that the power series is dominated by a convergent geometric series if $|x-a| < r$. Hence the power series converges uniformly by Theorem 21.4. (See also Remark 21.5.) $\square$

Example 22.4. If $a = 0$ and $c_k = 1$, then $R = 1$ and

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k$$

where the convergence is uniform on each interval $[-r, r]$ with $0 \leq r < 1$. The convergence is not uniform on the interval $(-r, 1)$ or on the interval $(-1, r)$ and the series does not converge at $x = \pm 1$.

Example 22.5. If $a = 0$ and $c_k = 1/k!$, then $R = \infty$ and

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

where the convergence is uniform on each interval $[-r, r]$ with $r < \infty$.

Remark 22.6. A power series *never* converges uniformly on an unbounded interval unless it is a polynomial. To see this we argue by contradiction. Let

$$P_n(x) = \sum_{k=0}^n c_k(x-a)^k$$

denote the n th partial sum of the power series $\sum_k c_k(x-a)^k$ and assume that $(P_n)_n$ converges uniformly to a function f on some interval I . Let $\varepsilon = 1$. Then there exists N such that $|P_n(x) - f(x)| < 1$ whenever $n > N$ and $x \in I$. If the power series $\sum_k c_k(x-a)^k$ has infinitely many nonzero terms, there exist $m > n > N$ with $c_m \neq 0$ and $c_n \neq 0$ and hence $P_n - P_m$ is a nonconstant polynomial. But then $|P_n(x) - P_m(x)| \leq |P_n(x) - f(x)| + |f(x) - P_m(x)| \leq 2$ for all $x \in I$. Hence the interval I must be unbounded as a nonconstant polynomial becomes infinite as $x \rightarrow \pm\infty$.

Theorem 22.7. Let $\sum_{k=0}^{\infty} c_k(x-a)^k$ be a power series and R be its radius of convergence. Denote the sum by

$$f(x) := \sum_{k=0}^{\infty} c_k(x-a)^k$$

for $|x-a| < R$. Then f is differentiable on the interval $(a-R, a+R)$, its derivative is given by term-wise differentiation, i.e.

$$f'(x) = \sum_{k=1}^{\infty} k c_k(x-a)^{k-1},$$

and the radius of convergence of this last power series is also R .

Corollary 22.8 (Taylor Series). *Continue the hypotheses of Theorem 22.7. Then f is infinitely differentiable¹² on the interval $(a - R, a + R)$, the n th derivative $f^{(n)}$ of f is*

$$f^{(n)}(x) = \sum_{k=n}^{\infty} k(k-1)\cdots(k-n+1)c_k(x-a)^{k-n} = \sum_{k=n}^{\infty} \frac{k!c_k}{(k-n)!} (x-a)^{k-n},$$

so that $c_n = f^{(n)}(a)/n!$, i.e.

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k.$$

Corollary 22.9. *The n th derivative of the sum of the geometric series is*

$$\sum_{k=n}^{\infty} \frac{k!}{(k-n)!} x^{k-n} = \frac{n!}{(1-x)^n}$$

for $-1 < x < 1$.

23 Analytic Functions

Definition 23.1. A function f is called **analytic** iff for every a in its domain there is a power series with

$$f(x) = \sum_{k=0}^{\infty} c_k(x-a)^k$$

for all x in some interval about a . From Corollary 22.8 it follows that an analytic function is infinitely differentiable and that it equals its Taylor series at each point a in its domain, i.e. $c_k = f^{(k)}(a)/k!$ and hence

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!} (x-a)^k.$$

Problem 23.2. Show that function $f : \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$f(x) = \begin{cases} e^{-1/x} & \text{if } x > 0 \\ 0 & \text{if } x \leq 0 \end{cases}$$

¹² A function is called **infinitely differentiable** iff it has derivatives of all orders.

is infinitely differentiable but not analytic. (All its derivatives vanish at zero so it cannot equal its Taylor series.) Hint: Show inductively that the k th derivative of f has form

$$f^{(k)}(x) = \frac{P(x)}{Q(x)} e^{-1/x}$$

for $x > 0$ where $P(x)$ and $Q(x)$ are polynomials. Then use the definition of the derivative to show that $f^{(k+1)}(0) = 0$. Then

$$\frac{P(x)}{Q(x)} e^{-1/x} = \frac{\tilde{P}(y)}{\tilde{Q}(y)} e^{-y}, \quad y = \frac{1}{x}$$

where $\tilde{P}(y)$ and $\tilde{Q}(y)$ are also polynomials. The variable x approaches 0 as y approaches ∞ . Use l'Hôpital's rule.

Problem 23.3. Give two proofs that the function $f(x) = x^{-1}$ defined for $x > 0$ is analytic. First: Write $f(x) = P_n(x, a) + R_n(x, a)$ where

$$P_n(x, a) := \sum_{k=0}^n \frac{f^{(k)}(a)(x-a)^k}{k!}$$

and show that $\lim_{n \rightarrow \infty} R_n(x, a) = 0$ for x sufficiently near a . Second: Use the identity

$$\frac{1}{x} = \frac{1}{a(1+y)}$$

where $y = (x-a)/a$ and the formula for the sum of a geometric series.

Theorem 23.4. *If a function f is represented by a power series in an open interval I about of some point in its domain, it is analytic in that interval, i.e. if $a \in I$ and*

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)(x-a)^k}{k!} \tag{†}$$

for $x \in I$, then

$$f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(b)(x-b)^k}{k!}$$

for $x, b \in I$.

Proof. This is normally proved in a course on complex variables (Math 623 at UW). Here is a proof that doesn't use complex numbers. Choose $b \in I$ and write Taylor's Formula (Theorem 19.1) centered at b

$$f(x) = P_n(b, x) + R_n(b, x)$$

$$P_n(b, x) := \sum_{k=0}^n \frac{f^{(k)}(b)}{k!} (x-b)^k, \quad R_n(b, x) := \int_b^x \frac{f^{(n+1)}(t)}{n!} (x-t)^n dt.$$

We must show that $\lim_{n \rightarrow \infty} R_n(b, x) = 0$ for x in an open interval about b . Let R be the radius of convergence of $(\dagger)$ so that

$$R^{-1} = \limsup_{k \rightarrow \infty} |c_k|^{1/k}, \quad c_k := \frac{f^{(k)}(a)}{k!}.$$

By Theorem 22.3 the series $(\dagger)$ diverges for $|x - a| > R$ so we may as well assume that $I = (a - R, a + R)$. Since $b \in I$ we have $|b - a| < R$. Let $r = (R + |b - a|)/2$ and $\rho = (R + r)/2$ so that $|b - a| < r < \rho < R$. As in the proof of that theorem we have the inequality

$$|c_k| |x - a|^k \leq \left(\frac{\rho}{r}\right)^k$$

for sufficiently large k and $x \in (a - r, a + r)$. Since this last inequality holds for sufficiently large k there is an M such that

$$|c_k| |x - a|^k \leq M \left(\frac{r}{\rho}\right)^k$$

for all k and all $x \in (a - r, a + r)$. Take the absolute value of the series for the n th derivative:

$$\begin{aligned} |f^{(n)}(x)| &= \left| \sum_{k=n}^{\infty} \frac{k! c_k}{(k-n)!} (x-a)^{k-n} \right| \\ &\leq \sum_{k=n}^{\infty} \left| \frac{k! c_k}{(k-n)!} (x-a)^{k-n} \right| \\ &\leq \sum_{k=n}^{\infty} M \frac{k!}{(k-n)!} \left(\frac{r}{\rho}\right)^{k-n} \\ &= M n! \left(1 - \frac{r}{\rho}\right)^{-n} \end{aligned}$$

by Corollary 22.9. Now choose $\delta > 0$ so small that $(b - \delta, b + \delta) \subseteq (a - r, a + r)$ and

$$\mu := \left(1 - \frac{\rho}{r}\right)^{-1} \delta < 1.$$

Then for $x \in (b - \delta, b + \delta)$ and t between b and x we have $|x - t| < \delta$ so

$$\begin{aligned} |R_n(b, x)| &= \left| \int_b^x \frac{f^{(n+1)}(t)}{n!} (x - t)^n dt \right| \\ &\leq |b - x| M \left(1 - \frac{r}{\rho}\right)^{-n-1} \frac{\delta^n}{n!} \\ &\leq M(n+1)! \left(1 - \frac{r}{\rho}\right)^{-n-1} \frac{\delta^{n+1}}{n!} \\ &= M(n+1)\mu^{n+1}. \end{aligned}$$

But $\mu < 1$ so $\lim_{n \rightarrow \infty} |R_n(b, x)| = M \lim_{n \rightarrow \infty} n\mu^n = 0$ as required. $\square$

24 Power Series - II

Theorem 24.1. *Assume that the power series*

$$f(x) = \sum_{k=0}^{\infty} c_k (x - a)^k$$

has radius of convergence R and converges pointwise on the closed interval $[a, a + R]$. Then it converges uniformly on that interval. Similarly for the closed interval $[a - R, a]$.

Proof. (See Buck page 279.) We assume that $R = 1$ and $a = 0$ to simplify the notation: the proof in general is similar but messier. Thus define

$$f(x) = \sum_{k=0}^{\infty} c_k x^k, \quad 0 \leq x < 1; \quad f(1) = \sum_{k=0}^{\infty} c_k.$$

Define $B_n := \sum_{k=n}^{\infty} c_k$. Then $c_k = B_k - B_{k+1}$. Since the series $f(1)$ converges, it follows that $\lim_{n \rightarrow \infty} B_n = 0$. Choose $\varepsilon > 0$. There exists N such that

$|B_n| < \varepsilon/2$ for $n > N$. For $x \in [0, 1)$ we have

$$\begin{aligned}\sum_{k=n}^{\infty} c_k x^k &= \sum_{k=n}^{\infty} (B_k - B_{k+1}) x^k \\ &= B_n x^n + \sum_{k=n+1}^{\infty} B_{k+1} (x^{k+1} - x^k) \\ &= B_n x^n + (x - 1) \sum_{k=n}^{\infty} B_{k+1} x^k\end{aligned}$$

so for $n > N$

$$\begin{aligned}\left| \sum_{k=n}^{\infty} c_k x^k \right| &\leq |B_n| x^n + (1 - x) \sum_{k=n}^{\infty} |B_{k+1}| x^k \\ &\leq \frac{\varepsilon x^n}{2} + \frac{(1 - x)\varepsilon x^n}{2(1 - x)} \sum_{j=0}^{\infty} x^j \\ &= \varepsilon\end{aligned}$$

When $x = 1$ this inequality holds because $|B_n| < \varepsilon/2 < \varepsilon$. $\square$

Corollary 24.2. *Suppose the power series*

$$f(x) = \sum_{k=0}^{\infty} c_k (x - a)^k$$

converges pointwise on the interval $[a, a + R]$. Then the function f is continuous on the interval $[a, a + R]$. Similarly for the closed interval $[a - R, a]$.

Proof. By Theorem 24.1 the series converges uniformly so the function f is continuous by Theorem 21.3. $\square$

Remark 24.3. The formula

$$\frac{1}{1 - t} = \sum_{k=0}^{\infty} t^k$$

was proved in 20.3. By Theorem 21.6 we may integrate and get

$$\ln(1 - x) = - \int_0^x \frac{dt}{1 - t} = - \sum_{k=0}^{\infty} \frac{x^{k+1}}{k+1} = - \sum_{k=1}^{\infty} \frac{x^k}{k}$$

which holds for $-1 < x < 1$. Let $f(x)$ denote the right hand side. When $x = -1$ the right hand side is the Alternating Harmonics Series from Example 20.11 and it converges. The series converges for $x = -1$ by Theorem 20.10. Hence f is continuous on $[-1, 0)$ by Theorem 24.1. But $\ln(1 - x)$ is also continuous on $[-1, 0)$. Since $f(x) = \ln(1 - x)$ on $(-1, 1)$ this must remain true on $[-1, 1)$ so taking $x = -1$ gives

$$\ln(2) = - \sum_{k=1}^{\infty} \frac{(-1)^k}{k}.$$

This may be stated more dramatically as

$$\lim_{n \rightarrow \infty} \lim_{x \rightarrow 1} \sum_{k=1}^n \frac{(-x)^k}{k} = \lim_{x \rightarrow 1} \lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{(-x)^k}{k}$$

in contrast to

$$\lim_{n \rightarrow \infty} \lim_{x \rightarrow 1} x^n = 1 \neq 0 = \lim_{x \rightarrow 1} \lim_{n \rightarrow \infty} x^n.$$

25 The Heat Equation

Problem 25.1. Assume that the sequence $(b_n)_n$ is dominated by the sequence $(n^{-p})_n$ i.e there is an N and C such that

$$|b_n| \leq Cn^{-p}$$

for $n > N$.

(1) Show that, if $p > 1$, the series

$$f(x) := \sum_{n=1}^{\infty} b_n \sin nx \tag{25.1-1}$$

converges uniformly.

(2) Show that, if $p > 1$, then

$$b_n = \frac{2}{\pi} \int_0^{\pi} f(x) \sin(nx) dx. \tag{25.1-2}$$

(3) Show that, if $p > 2$, then f is differentiable and that

$$f'(x) := \sum_{n=1}^{\infty} n b_n \cos nx.$$

Hint: See Problem 20.17 above. You may use any of the theorems stated above but state which theorems you are using and verify that the hypotheses of the theorems are satisfied.

Problem 25.2. Continue the notation of Problem 25.1. Show that the series

$$u(t, x) := \sum_{n=1}^{\infty} e^{-n^2 t} b_n \sin nx \quad (25.2-1)$$

converges uniformly on $[0, \infty) \times [0, \pi]$ if $p > 1$ and that the limit satisfies the partial differential equation¹³

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} \quad (25.2-2)$$

on the open set $(0, \infty) \times (0, \pi)$. Show also that u is continuous on the closed set $[0, \infty) \times [0, \pi]$, that it satisfies the initial condition

$$u(0, x) = f(x) \quad (25.2-3)$$

and the boundary condition

$$u(t, 0) = u(t, \pi) = 0. \quad (25.2-4)$$

(You may use any of the theorems stated above but state which theorems you are using and verify that the hypotheses of the theorems are satisfied. Hint: $e^{-n^2 t}$ is very small if $t > 0$ and n large.) It looks like this exercise proves that the solution of the partial differential equation (25.2-2) subject to the initial condition (25.2-3) and the boundary condition (25.2-4) is given by (25.2-1) where the coefficients are defined by (25.1-2). Is there anything missing for a rigorous proof?

¹³ This PDE is called the **Heat Equation**.

26 Fourier Series - I

27 Fourier Series - II

Theorem 27.1. *Assume that the function $f : \mathbb{R} \rightarrow \mathbb{C}$ is 2π periodic and Lipshitz. Then the Fourier series for f converges uniformly to f , i.e.*

$$f(x) = \lim_{n \rightarrow \infty} \sum_{k=-n}^n c_k e^{ikx}, \quad c_k := \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx$$

and the convergence is uniform.

Proof. See Rudin page 189. □

A Uniqueness of the Real Numbers

Definition A.1. A **field** is a set R equipped two binary operations

$$R \times R \rightarrow R : (x, y) \mapsto x + y, \quad R \times R \rightarrow R : (x, y) \mapsto x \cdot y,$$

two distinct special elements 0 and 1, and two unary operations

$$R \rightarrow R : a \mapsto -a, \quad R \setminus \{0\} \rightarrow R \setminus \{0\} : a \mapsto a^{-1},$$

satisfying the following (additive and multiplicative) commutative, associative, identity, and inverse laws:

	Addition	Multiplication
commutative	$a + b = b + a$	$a \cdot b = b \cdot a$
associative	$(a + b) + c = a + (b + c)$	$(a \cdot b) \cdot c = a \cdot (b \cdot c)$
identity	$a + 0 = a$	$a \cdot 1 = a$
inverse	$a + (-a) = 0$	$a \cdot a^{-1} = 1$

and the distributive law:

$$(a + b) \cdot c = (a \cdot c) + (b \cdot c).$$

The operations of *subtraction* and *division* are then defined by

$$a - b := a + (-b) \quad \frac{a}{b} := a \cdot b^{-1} = a \cdot \frac{1}{b}.$$

The standard abbreviations $ab := a \cdot b$, $a - b := a + (-b)$, and $a/b = a \cdot b^{-1}$ are used. The above axioms are those which appear in 9.1.

A.2. The rational numbers $\mathbb{Q}$, the real numbers $\mathbb{R}$, and the complex numbers $\mathbb{C}$ are the most important examples of fields, but there are many others, e.g. the field $\mathcal{R}$ of rational functions from problem 9.9 and the field $\mathbb{Q}(\sqrt{2})$ from problem 9.11. In Math 441 (or 541) you will even meet finite fields. (The simplest example of a finite field is the set $\{0, 1, 2, \dots, p-1\}$ where p is a prime, and addition and multiplication are done 'modulo p '.)

A.3. The axioms in the column headed *Addition* above are the axioms for an abelian group in additive notation, and the axioms in the column headed *Multiplication* above are the axioms for an abelian group in multiplicative. If you replace 0 by 1, $a+b$ by $a \cdot b$, and $-a$ by a^{-1} in the former column you get the axioms in the latter column. Because both addition and multiplication satisfy the axioms for an abelian group there are further analogies.

Addition	Multiplication
(0) $a + b = 0 \implies b = -a$	(0) $a \cdot b = 1 \implies b = a^{-1}$
(i) $-(-a) = a$	(i) $(a^{-1})^{-1} = a$
(ii) $-(a + b) = -a - b$	(ii) $(a \cdot b)^{-1} = a^{-1} \cdot b^{-1}$
(iii) $-(a - b) = b - a$	(iii) $\left(\frac{a}{b}\right)^{-1} = \frac{b}{a}$
(iv) $(a - b) + (c - d) = (a + c) - (b + d)$	(iv) $\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}$
(v) $a - b = (a + c) - (b + c)$	(v) $\frac{a}{b} = \frac{ac}{bc}$
(vi) $(a - b) - (c - d) = (a - b) + (c - d)$	(vi) $\frac{a/b}{c/d} = \frac{a}{b} \cdot \frac{c}{d}$

Line (0) explains the phrase 'necessarily unique' used in 9.1. The last line explains why *we invert and multiply to divide fractions*.

A.4. Using the distributive law one can also prove the following familiar identities.

$$\text{(vii)} \quad a \cdot 0 = 0$$

$$\text{(viii)} \quad -a = (-1)a$$

$$\text{(ix)} \quad a(-b) = -ab$$

$$\text{(x)} \quad (-a)(-b) = ab$$

$$\text{(xi)} \quad \frac{a}{b} + \frac{c}{d} = \frac{ad + cb}{bd}$$

$$\text{(xii)} \quad (a + b)(c + d) = ab + ad + bc + bd$$

$$\text{(xiii)} \quad (a + b)^2 = a^2 + 2ab + b^2$$

$$\text{(xiv)} \quad (a + b)(a - b) = a^2 - b^2$$

It follows from these laws that if a product is zero, then one of its factors must be zero. (Proof: if $ab = 0$ and $a \neq 0$ then $b = 1 \cdot b(a^{-1}a)b = a^{-1}(ab) = a^{-1} \cdot 0 = 0$.)

Definition A.5. A **homomorphism** from a field R to a field R' is a map $\iota : R \rightarrow R'$ such that

$$\iota(a + b) = \iota(a) + \iota(b), \quad \iota(a \cdot b) = \iota(a) \cdot \iota(b)$$

for $a, b \in R$. (In each equation the operation on the left is the one for R and the operation on the right is the one for R' .) It follows easily that $\iota(0) = 0$, $\iota(1) = 1$, $\iota(-a) = -\iota(a)$, and $\iota(a^{-1}) = \iota(a)^{-1}$. An **isomorphism** is a bijective homomorphism.

Remark A.6. The inverse of an isomorphism is an isomorphism. For example, if $a', b', c' := a' + b' \in R$ and $a = \iota^{-1}(a')$, $b = \iota^{-1}(b')$, $c = \iota^{-1}(c')$, then $\iota(a + b) = a' + b'$ (as ι is a homomorphism) and $\iota(c) = c' = a' + b'$, so $\iota(c) = \iota(a + b)$ so $c = a + b$ (as ι is injective).

Proposition A.7. *A homomorphism of fields is injective.*

Proof. Assume that $\iota(a) = \iota(b)$; we must show that $a = b$. If $a \neq b$, then $c := a - b \neq 0$ so $1 = \iota(1) = \iota(c^{-1}c) = \iota(c)^{-1}\iota(c) = \iota(c)^{-1}\iota(a - b) = \iota(c)^{-1}\iota(0) = \iota(c)^{-1}0 = 0$ contradicting $1 \neq 0$. $\square$

Definition A.8. An **ordered field** is a field equipped with an order relation satisfying the conditions in 9.2. An **order preserving** homomorphism from an ordered field R to an ordered field R' is a homomorphism $\iota : R \rightarrow R'$ such that

$$a < b \implies \iota(a) < \iota(b).$$

Example A.9. The inclusions $\mathbb{Q} \rightarrow \mathbb{Q}(\sqrt{2})$, $\mathbb{Q} \rightarrow \mathbb{R}$, $\mathbb{Q} \rightarrow \mathbb{R}$ are all order preserving homomorphisms. The map

$$\mathbb{Q}(\sqrt{2}) \mapsto \mathbb{Q}(\sqrt{2}) : a + b\sqrt{2} \mapsto a - b\sqrt{2}$$

is a field homomorphism but is not order preserving.

Lemma A.10. *For any ordered field R , there is a unique order preserving homomorphism $\iota : \mathbb{Q} \rightarrow R$ from the set $\mathbb{Q}$ of rational numbers into R .*

Proof. The homomorphism ι must send $0 \in \mathbb{Q}$ to $0 \in R$, $1 \in \mathbb{Q}$ to $1 \in R$, $n \in \mathbb{Z}^+$ to

$$\iota(n) := \underbrace{\iota(1) + \iota(1) + \cdots + \iota(1)}_n \in R$$

a negative integer n to $\iota(n) = -\iota(-n)$, and therefore a ratio m/n of integers to $\iota(m/n) = \iota(m)/\iota(n)$. Since $\iota(1) > 0$ and R obeys the same algebraic and order laws as $\mathbb{Q}$, it is immediate that ι is an order preserving homomorphism. $\square$

Definition A.11. An ordered field R is **complete** iff every subset S of R which is bounded above has a least upper bound $\sup(S)$. (The definitions are *mutatis mutandis* the same as in 9.5.)

Theorem A.12. *For any complete ordered field R , there is a unique order preserving isomorphism $\iota : \mathbb{R} \rightarrow R$ from the set $\mathbb{R}$ of real numbers onto R .*

Proof. For $a \in \mathbb{R}$ we have $a = \sup\{q \in \mathbb{Q} : q < a\}$ so we must have

$$\iota(a) = \sup\{\iota(q) : q \in \mathbb{Q}, q < a\}.$$

We must show that ι is an order preserving homomorphism and that it is surjective. We omit the details. $\square$

B Additional Problems

Problem B.1. Fix a positive number $a \in \mathbb{R}$. The purpose of this problem is to define the exponential a^x for $x \in \mathbb{R}$. Define $a^0 := 1$ and for n a positive integer define

$$a^n := \underbrace{a \cdot a \cdots a}_n, \quad a^{-n} := 1/a^n.$$

Then

- (1) Prove that for every nonzero integer n there is a unique solution $b > 0$ to the equation $b^n = a$. Define $a^{1/n}$ to be this unique solution, i.e. $a^{1/n} = b \iff b^n = a$.
- (2) For a rational number q define a^q by $a^q = (a^m)^{1/n}$ where $q = m/n$. Prove that this definition is independent of the choice of the integers m and n such that $q = m/n$.
- (3) Prove that there is a unique continuous function

$$\mathbb{R} \rightarrow (0, \infty) : x \mapsto a^x$$

such that $a^x = a^q$ when x is a rational number q .

In your proof make clear which theorems from these notes you are appealing to. Also make your proof self contained so that a person who doesn't have access to the statement of the problem can follow it. (You needn't provide proofs for the theorems you use, but do provide references to them.) In your proof of (3) you may use the inequality

$$|a^p - a^q| \leq M|p - q|$$

which is true when $a > 1$, N is a positive integer, $1 \leq p, q \leq N + 1$, and

$$M := \left(\sum_{k=1}^N \frac{1}{k} \right) a^N.$$

You need not prove this inequality but use calculus to show where it comes from. Hint: What is the definition of $\ln x$, e^x , and a^x used in calculus? Consider the Integral Test 20.13. The natural logarithm function $\ln x$ is usually defined as an integral. How do you bound an integral by a sum? How does the Mean Value Theorem from calculus (see 17.4)) give inequalities like this?

Index

- n space, 9
- n th Term Test, 52
- n tuple, 9

- accumulation point, 24, 26
- Algebraic Axioms, 18
- Alternating harmonic series, 54
- alternating series, 53
- analytic, 61
- Archimedean Property, 20
- Axiom of Choice, 14

- bijjective, 14
- Bolzano-Weierstrass, 30
- boundary, 36
- bounded, 24
- bounded above, 19
- bounded below, 19

- cardinality, 16
- Cartesian product, 9
- Cauchy, 32
- Cauchy Convergence Criterion, 32
- Cauchy Convergence Criterion for Series, 52
- closed, 35, 36
- closure, 36
- cluster point, 24
- compact, 39
- Comparison Test, 54
- complement, 9
- complete, 71
- Completeness Axiom, 19
- composition, 13
- connected, 37
- continuous, 32, 33
- contrapositive, 7
- converge, 26
- converge absolutely, 53
- converge conditionally, 53
- converge pointwise, 56
- converge uniformly, 56
- convergent, 26
- converges, 26
- converse, 7
- countable, 16

- De Morgan's Laws, 10
- decreasing, 28
- Dedekind cuts, 20
- definite integral, 44
- denumerable, 16
- derivative, 25, 43
- differentiable, 43
- direct product, 9
- disconnected, 37
- disjoint, 8
- distance, 22
- diverge, 26
- domain, 11
- dominated, 57
- dot product, 23
- dummy variable, 49

- empty set, 8
- equal, 8, 11
- equality of sets, 11
- extended real numbers, 25
- exterior, 36

- Fibonacci numbers, 15
- field, 68

finite, 16	lower sum, 44
function, 11	map, 11
Fundamental Theorem of Calculus, 48	mathematical induction, 15
geometric series, 51	Mean Value Theorem, 43
graph, 11	mesh, 44
greatest lower bound, 19	monotonic, 28
Harmonic series, 52	natural numbers, 5
Heat Equation, 67	neighborhood, 24, 25
Heine Borel, 39	nondecreasing, 28
homomorphism, 70	nondenumerable, 16
identity map, 13	nonincreasing, 28
image, 11	norm, 22
increasing, 28	one-one, 14
index set, 9	one-one onto, 14
indexed family of sets, 9	onto, 14
inductive definitions, 15	open, 34
infimum, 19	open ball, 23
infinite, 16	open cover, 39
infinitely differentiable, 61	Order Axioms, 18
injective, 14	order preserving, 70
inner product, 23	ordered field, 70
Integral Test, 54	partial sums, 50
interior, 36	partition, 44
Intermediate Value Theorem, 38	permutation, 56
intersection, 8, 9	power series, 59
interval notation, 5	preimage, 12
inverse, 13	proof, 7
inverse image, 11	proof by contradiction, 7
isomorphism, 70	punctured neighborhood, 24
least upper bound, 19	radius of convergence, 59
left inverse, 13	range, 12
limit point, 24	real numbers, 18
Lipschitz, 33	rearrangement, 56
logical operations, 6	refines, 44
lower bound, 19	

- relatively open, 35
- Riemann sum, 46
- right inverse, 13
- Root Test, 54
- Schwarz inequality, 23
- sequence, 26
- series, 50
- source, 12
- strictly, 28
- subsequence, 29
- subset, 8
- supremum, 19
- surjective, 14
- target, 12
- Taylor polynomial, 49
- Taylor remainder, 49
- Taylor Series, 61
- Taylor's Formula – Lagrange Form, 49
- terms, 50
- trace, 31
- truth tables, 10
- uncountable, 16
- uniformly continuous, 33
- union, 8, 9
- upper bound, 19
- upper sum, 44
- weakly monotonic, 28
- Weierstrass Comparison Test, 57