Real Analysis Notes

Franklin Mendivil

These are my lecture notes for Math 3533 and 4533 (Real Analysis I and II) as I have delivered these courses the last few times I have taught them. I would recommend that you also either get the recommended text or find some real analysis text(s) from the library or some other source. These notes provide the main definitions and theorems along with their proofs, but they don't provide very many examples. These notes are intended to provide an outline of the material covered in an introductory analysis course. During the lectures I provide additional material, comments and examples. In particular, notice that there are very few figures in these notes.

0.1 To the student:

Since I give you these lecture notes, you might ask *why should I go to class?* or *why should I pay attention in class?*. Well, analysis is not something that comes easy to most people. I would **highly** recommend that you read ahead a little bit and come to class already having read the material. Then you can use the lecture time to fill in any gaps in your understanding and to make sure that you get the "big picture". Since you don't have to spend your class time copying everything from the board, you can actually spend more of the class time thinking about the material!

Warning: There will be things discussed in class that are NOT in these notes, so you should copy those down. I would suggest that you use the reverse sides of the pages to write further notes, draw the diagrams, add my comments about "why" something is true, add your own thoughts and insights about the material, etc.

Following along with the proofs in the class, trying to see the "big picture" will help you (the student) to start to see patterns in why one approach is used in one proof but another approach is used in another.

Proving theorems is important because it provides the element of rigor necessary in a mathematics class. However, this is not the only (or most important) reason for proofs. We prove things not only to make sure that they are indeed correct, but also to make sure that we understand *why* they work as they do. Most often, the best way to approach a proof of some statement is first to develop an intuitive, rough idea as to why the statement is true. This is often done by drawing some pictures or diagrams or using analogies with similar (previous) situations. However, this is just the first step. We must next turn our vague intuition into a concrete proof, using the language of mathematics, including all the definitions and theorems at our disposal.

Constructing proofs is quite different from solving problems in calculus. Many calculus problems are purely computational – once you know the techniques or pattern, you replicate this pattern (with minor variations) for all similar problems. However, proof courses don't usually offer these types of computational questions. Often each question in these notes will seem to be an individual question totally different from all the other questions. There is usually no complete set of *question types* from which to draw examples to follow.

So how do you approach a given problem? First, make sure that you understand all the terms in the question. Read the definitions. Look at any examples in the notes (or from your class notes) and see what examples have been given (for example, if the question is about sequences, look at various examples of sequences). What you should look for is to see if you understand how the formal definition fits with the examples and what are the range of possible behaviors.

Next see if you can come up with some example that violates the statement that you are trying to prove (for example, if you are trying to prove that all bounded sequences have an increasing or decreasing subsequence, try to find a bounded sequence which has neither an increasing nor a decreasing subsequence). If you can see why you cannot come up with such a counter-example, this might give you insight into why it is true.

If the problem asks you to prove a statement for all $n \in I\!N$, see if you understand why it is true for small n. If the problem asks you to prove something true for all continuous f, try polynomials first. Basically, try easier versions of the problem first.

Once you have accumulated enough insight into the problem that you think you might know WHY it is true (at least, in some vague, intuitive way), then you should try to put your intuition into formal language.

Last but not least, **PLEASE COME AND SEE ME FOR HELP!!!!**

Contents

Chapter 1

Set Theory Preliminaries

Basic Set Theory

Let A, B be subsets of some set X (that is, let $A, B \subset X$). Then we define

- $A \cup B = \{x \in X : x \in A \text{ or } x \in B\}$
- $A \cap B = \{x \in X : x \in A \text{ and } x \in B\}$
- $A \setminus B = \{x \in A : x \notin B\}$.

In case X is understood, we sometimes write A^c for $X \setminus A$.

Distributivity of $\cap$ and $\cup$

If $A, B, C \subset X$ then

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

We prove the first one of these as an example. First we make some comments on how to prove equality of sets.

What do we mean by the statement that $A = B$, for two sets A, B? What we mean is that these two sets have the same elements. That is, if $x \in A$ then $x \in B$ and conversely. So, the statement that $A = B$ is the same as the two statements $A \subset B$ and $B \subset A$. So, to prove equality of two sets it suffices to prove the latter two statements. This is our strategy in the next proof.

Proposition 1.1 *Let $A, B, C \subset X$. Then $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.*

Proof: We prove that $A \cup (B \cap C) \subset (A \cup B) \cap (A \cup C)$ and $A \cup (B \cap C) \supset (A \cup B) \cap (A \cup C)$.

Take $x \in A \cup (B \cap C)$. Then $x \in A$ or $x \in B \cap C$. First we consider the case where $x \in A$. In this case, $x \in A$ so $x \in A \cup B$ and $x \in A \cup C$ so $x \in (A \cup B) \cap (A \cup C)$. On the other hand, if $x \in B \cap C$ then $x \in B$ and $x \in C$ so that $x \in A \cup B$ and $x \in A \cup C$ so it must be true that $x \in (A \cup B) \cap (A \cup C)$. Thus, in any case we have shown that if $x \in A \cup (B \cap C)$ then $x \in (A \cup B) \cap (A \cup C)$. Therefore, $A \cup (B \cap C) \subset (A \cup B) \cap (A \cup C)$.

Now we need to show the reverse inclusion. To this end, let $x \in (A \cup B) \cap (A \cup C)$. Then $x \in A \cup B$ or $x \in A \cup C$. There are four cases:

1. $x \in A$ and $x \in A$
2. $x \in A$ and $x \in C$
3. $x \in B$ and $x \in A$
4. $x \in B$ and $x \in C$

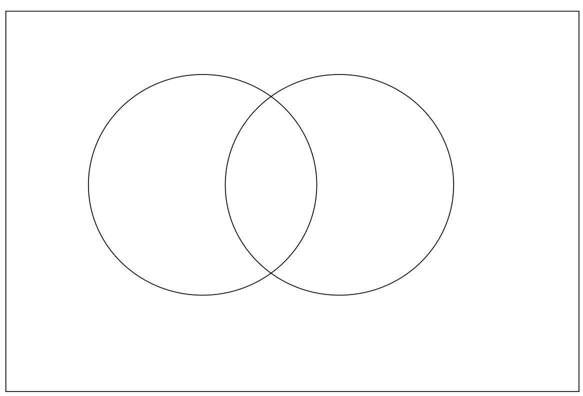
Notice that in cases 1) - 3), we know that $x \in A$ so that $x \in A \cup (B \cap C)$. In case 4), we know that $x \in B \cap C$ which implies that $x \in A \cup (B \cap C)$. Therefore, in all cases we see that if $x \in (A \cup B) \cap (A \cup C)$ then $x \in A \cup (B \cap C)$. Thus, $(A \cup B) \cap (A \cup C) \subset A \cup (B \cap C)$.

Since we have proven both $A \cup (B \cap C) \subset (A \cup B) \cap (A \cup C)$ and $(A \cup B) \cap (A \cup C) \subset A \cup (B \cap C)$, we know that these two sets are equal. ∎

de Morgan's Laws

Let $A, B \subset X$. Then

- $X \setminus (A \cup B) = (X \setminus A) \cap (X \setminus B)$.
- $X \setminus (A \cap B) = (X \setminus A) \cup (X \setminus B)$.



Cartesian Products

Definition 1.1 *Let A and B be sets. The* Cartesian Product of A and B, *denoted by $A \times B$, is the set of ordered pairs*

$$A \times B = \{(a, b) : a \in A, b \in B\}.$$

Example If $A = \{1, 2, 3\}$ and $B = \{red, green, blue\}$ then

$$A \times B = \{(1, red), (1, green), (1, blue), (2, red), (2, green), (2, blue), (3, red), (3, green), (3, blue)\}.$$

Insert a nice example here with $A = [1, 2] \cup (3, 4)$ and $B = \{1\} \cup (2, 3]$

Notice that we haven't defined what we mean by an ordered pair (in set theoretical terms, that is). As an exercise, try to think of a way to define an ordered pair using only set theoretical constructions.

Relations and functions

Now we use Cartiesian Products to define what we mean by relations and functions.

Definition 1.2 *A* relation R on $X \times Y$ *is any subset of $X \times Y$. If $Y = X$, then we say that we have a* relation on X.

We write xRy if $(x, y) \in R$.

Definition 1.3 *Let R be a relation on $X \times Y$. The* domain of R *is the subset of X given by*

$$Dom(R) = \{x : (x, y) \in R\}$$

and the range of R *is the subset of Y given by*

$$Range(R) = \{y : (x, y) \in R\}.$$

Example Let R be the usual order on $\{0, 1, 2, 3, \ldots\}$. Then

$$Dom(R) = \{0, 1, 2, 3, \ldots\}$$

while

$$Range(R) = \{1, 2, 3, \ldots\}.$$

Insert nice picture of this relation on $\{0, 1, 2, \ldots\} \times \{0, 1, 2, \ldots, \}$

There are three important classes of relations:

1. A relation $\sim$ on X is called an *equivalence relation* if it satisfies:

   - for all $x \in X$, $x \sim x$ (i.e., $(x,x) \in\sim$).
   - if $x \sim y$ then $y \sim x$ (i.e., if $(x,y) \in\sim$ then $(y,x) \in\sim$).
   - if x, y, z are all distinct and $x \sim y$ and $y \sim z$ then $x \sim z$ (i.e., if $(x,y) \in\sim$ and $(y,z) \in\sim$ then $(x,z) \in\sim$).

   These are called *reflexive, symmetric* and *transitive* properties. For each $x \in X$, let $[x] = \{z : z \sim x\}$ be the equivalence class which contains x. By the three properties above, we can see that either $[x] = [y]$ or $[x] \cap [y] = \emptyset$ and X is the union of the equivalence classes. Thus the equivalence classes form a partition of X. In fact, given an arbitrary partition of X, we can obtain an equivalence relation by defining $x \sim y$ if both x and y are in the same element of the partition. Thus, equivalence relations and partitions are, in some sense, equivalent.

2. A relation $<$ on X is called a *partial order* if

   - $x < y$ and $y < z$ implies $x < z$
   - for all $x \in X$ it is not the case that $x < x$.

   If, in addition, we know that

   - for all $x, y \in X$ either $x < y$ or $x = y$ or $x > y$,

   then the relation is called a *total order*.

3. A relation f on $X \times Y$ is called a *function from X to Y*, denoted by $f : X \to Y$, if it satisfies:

   - $Dom(f) = X$
   - for each $x \in X$, there is a unique $y \in Y$ with $(x,y) \in f$.

   We usually write $f(x) = y$ rather than $(x,y) \in f$ or xfy. The second condition is the so-called "vertical line test".

There are many different examples of these relations.

Example For a simple example of an equivalence relation, let $X = \mathbb{Z}$ and define $n \sim m$ if $n - m$ is a multiple of 7. Then the 7 equivalence classes are $C_i = \{7n + i : n \in \mathbb{Z}\}$ for $i = 0, 1, \ldots, 6$.

Example The usual order on $\mathbb{R}$ is a total order.

Example Let $X = \{1,2,3\}$ and $Pow(X) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}$ be the set of all subsets of X. We define a partial order $<$ on $Pow(X)$ by letting

$A < B$ if $A \subset B$. This order is not a total order since $\{1, 2\}$ is neither larger than nor smaller than $\{1, 3\}$. Clearly we can do the same with an arbitrary set X.

Definition 1.4 *Let $f : X \to Y$ be a function.*

- *We say that f is* surjective *or* onto *if $Range(f) = Y$.*
- *We say that f is* injective *or* one-to-one *if $f(x_1) = f(x_2)$ implies that $x_1 = x_2$.*
- *We say that f is* bijective *if it is both injective and surjective.*

Example The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^3$ is a bijection.

Example The function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2$ is neither injective nor surjective.

Example The function $f : \mathbb{R} \to [0, \infty)$ given by $f(x) = x^2$ is surjective but not injective.

Example The function $f : [0, \infty) \to [0, \infty)$ given by $f(x) = x^2$ is a bijection.

Theorem 1.2 *Let $f : X \to Y$ be a function. Then f is a bijection iff there is a function $g : Y \to X$ so that $f(g(y)) = y$ for all $y \in Y$ and $g(f(x)) = x$ for all $x \in X$.*

Proof: Suppose that $f : X \to Y$ is a bijection. We must prove that there is a function $g : Y \to X$ with $f(g(y)) = y$ and $g(f(x)) = x$. To do this, we construct g as the relation on $Y \times X$

$$g = \{(y, x) : (x, y) \in f\}.$$

We want to show that g is a function from Y to X and that $g(f(x)) = x$ and $f(g(y)) = y$.

First, $Dom(g) = Range(f) = Y$ by the definition of g and the fact that f is a surjection. Furthermore, for each $y \in Y$ there is only one $x \in X$ with $(y, x) \in g$ since f is one-to-one. Thus, g is a function.

By definition we know that $g(f(x)) = x$ and $f(g(y)) = y$ Similarly, $f(g(y)) = y$. Thus, we have proved the existence of such a function.

Now to prove the converse we assume that there is a function $g : Y \to X$ so that $g(f(x)) = x$ and $f(g(y)) = y$. We must prove that f is a bijection.

Suppose that $f(x_1) = f(x_2)$. Then $x_1 = g(f(x_1)) = g(f(x_2)) = x_2$. Thus f is injective.

Furthermore, for all $y \in Y$ we see that $f(g(y)) = y$ so letting $x = g(y)$ we have that $f(x) = y$. Thus $Range(f) = Y$ so f is a surjection. Since it is injective and surjective, f is a bijection. ∎

The function g in the previous theorem is called the *inverse of* f and is denoted by f^{-1}.

Images and Preimages

Definition 1.5 *Let* $f : X \to Y$ *and* $A \subset X$ *and* $B \subset Y$.

- *The* image of A under f *(or by* f*) is the set* $f(A) = \{f(x) : x \in A\}$.
- *The* preimage of B under f *(or by* f*) is the set* $f^{-1}(B) = \{x \in X : f(x) \in B\}$.

Notice that f need not be invertible for the preimage to be defined.

Example Let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x^2 + 5$, $A = (-3, 2]$, $B = [-2, -1]$ and $C = [5, 6]$. Then $f(A) = [5, 14]$, $f^{-1}(B) = \emptyset$ and $f^{-1}(C) = [-1, 1]$. This function does not have an inverse since it is not injective (or surjective, for that matter).

If the inverse exists, then the preimage of B under f is the same as the image of B under f^{-1}. That is to say the notation makes sense.

Proposition 1.3 *Let* $A, B \subset X$, $E, F \subset Y$ *and* $f : X \to Y$. *Then*

1. $f(A \cup B) = f(A) \cup f(B)$.
2. $f(A \cap B) \subset f(A) \cap f(B)$.
3. $f^{-1}(A \cup B) = f^{-1}(A) \cup f^{-1}(B)$.
4. $f^{-1}(A \cap B) = f^{-1}(A) \cap f^{-1}(B)$.

It is instructive to prove these properties and to construct an example to show why property 2 is not an equality.

Cardinality: Finite, infinite, countable, and uncountable

First we start out with some notation. Throughout these notes we use $\mathbb{N}$ to denote the set $\{1, 2, 3, \ldots\}$ and $\mathbb{Z}$ to denote the set $\{0, \pm 1, \pm 2, \pm 3, \ldots\}$.

Before we define what we mean by finite and infinite sets, we will prove a very useful version of a counting principle – the Pigeon Hole Principle.

Theorem 1.4 *(Pigeon Hole Principle) Let* $n, m \in \mathbb{N}$ *with* $n < m$. *Then no function* $f : \{1, 2, \ldots, m\} \to \{1, 2, \ldots, n\}$ *can be an injection. Furthermore, no function* $g : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, m\}$ *can be a surjection.*

Proof: Since there is an injection $f : X \to Y$ iff there is a surjection $g : Y \to X$ (see Theorem 1.10), the two statements are equivalent. Thus we need only prove one of them, and we choose to prove the first.

If $f : \{1, 2, \ldots, m\} \to \{1, 2, \ldots, n\}$ is an injection, then so is the restriction of f to $\{1, 2, \ldots, n+1\}$. Thus, it is sufficient to prove that for all $n \in I\!N$ there can be no injection $f : \{1, 2, \ldots, n, n+1\} \to \{1, 2, \ldots, n\}$. We do this by induction.

BASIS: $n = 1$. This is clearly true as $f : \{1, 2\} \to \{1\}$ must have $f(1) = f(2) = 1$, which is not an injection.

INDUCTION STEP: Suppose that for $n \in I\!N$ we have that no injection $g : \{1, 2, \ldots, n+1\} \to \{1, 2, \ldots, n\}$ can exist. Let $f : \{1, 2, \ldots, n+1, n+2\} \to \{1, 2, \ldots, n, n+1\}$ be a function. If f is an injection, then the restriction of f to $\{1, 2, \ldots, n+1\}$ is also an injection (we call this restriction f_R)

$$f_R : \{1, 2, \ldots, n+1\} \to \{1, 2, \ldots, n+1\} \setminus \{f(n+2)\} = S.$$

Define $h : \{1, 2, \ldots, n\} \to S$ by

$$h(i) = \begin{cases} i, & \text{if } i < f(n+2) \\ i+1, & \text{if } i > f(n+2) \end{cases} .$$

Then h is a bijection and

$$h^{-1} \circ f_R : \{1, 2, \ldots, n+1\} \to \{1, 2, \ldots, n\}$$

is an injection, contradicting our assumption. Thus, no such injection f can exist. ∎

We now proceed with the basic definitions.

Definition 1.6 *Let E be a set.*

1. *E is said to be* finite *if there exists some $N \in I\!N$ and a surjection $f : \{1, 2, \ldots, N\} \to E$. If E is not finite, it is* infinite.
2. *E is said to be* countable *if there is a surjection $f : I\!N \to E$. If E is not countable, it is* uncountable.

Notice that to prove that a set E is uncountable, we have to prove that there cannot be a surjection $f : I\!N \to E$.

Example The set $I\!N$ is countable. This is easy to see since $f : I\!N \to I\!N$ given by $f(n) = n$ is a bijection so is onto.

Example The set $E = \{2, 4, 6, \ldots\}$ is countable. This is easy to see since the function $f : I\!N \to E$ defined by $f(n) = 2n$ is clearly a surjection.

Example The set $\mathbb{Z}$ is countable. To see this, we need to define a surjectin $f : \mathbb{N} \to \mathbb{Z}$. We indicate how to do this by

$$\begin{array}{cccccccccccc} 0, & -1, & 1, & -2, & 2, & -3, & 3, & -4, & 4, & -5, & 5, & \cdots \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \cdots \\ 1, & 2, & 3, & 4, & 5, & 6, & 7, & 8, & 9, & 10, & 11, & \cdots \end{array}$$

so we define $f(1) = 0$, $f(2) = -1$, etc. We could find an explicit formula for f, but it is not necessary since it is clear that the function so defined is surjective (for those interested $f(n) = (-1)^{1+n}\lfloor n/2 \rfloor$ where $\lfloor x \rfloor$ is the greatest integer smaller than x).

Example Any bounded subset of N is finite. To show this, suppose that $n < M \in \mathbb{N}$ for all $n \in E$. Choose some $s \in E$ (what if $E = \emptyset$?). Define the function $f : \{1, 2, \ldots, M\} \to E$ by

$$f(n) = \begin{cases} n, & \text{if } n \in E \\ s, & \text{if } n \notin E \end{cases} .$$

Then f is clearly a surjection.

The next result is sometimes used as the definition of what it means to be an infinite set Notice that the proof depends on two of the problems at the end of this chapter. We will not use this result.

Theorem 1.5 *The set E is infinite iff there is a proper subset $A \subset E$ (that is, $A \neq E$) and an injection $E \to A$.*

Proof: Suppose that E is an infinite set. Then by problem 45 there is a proper subset $A \subset E$ and a bijection $f : A \to S$.

Conversely, suppose that there is a proper subset $A \subset E$ and an injection $f : E \to A$. Let $x_1 \in E \setminus A$ and $x_2 = f(x_1)$. Then $x_2 \neq x_1$ since $x_2 \in f(E)$ but $x_1 \notin f(E)$. Continuing, let $x_3 = f(x_2)$. Then $x_3 \neq x_2$ since f is an injection and $x_3 \neq x_1$ since $x_3 \in f(E)$. By induction we define $x_{n+1} = f(x_n)$ to obtain the distinct elements $x_1, x_2, \ldots, x_n \in E$. Now, define $g : \mathbb{N} \to E$ by $g(n) = x_n$. Then g is an injection, so by problem 31, E must be infinite. ∎

Theorem 1.6 *The set $\mathbb{N}$ is infinite.*

Proof: Let $N \in \mathbb{N}$ and $f : \{1, 2, \ldots, N\} \to \mathbb{N}$ be a function. We show that f is not a surjection.

If $f(i) < N{+}1$ for all $i = 1, 2, \ldots, N$, then clearly $N{+}1$ is not in the range of f so f is not surjective. On the other hand, if there is some j so that $f(j) \geq N{+}1$, then by the Pigeonhole Principle there must be some $n \in \{1, 2, 3, \ldots, N\}$ so that $f(i) \neq n$ for any $i \in \{1, 2, 3, \ldots, N\}$. This again shows that f is not surjective. ∎

Theorem 1.7 *Let $\{A_i : i = 1, 2, \ldots,\}$ be a collection of countable sets. Then*

$$A = \bigcup_i A_i$$

is also countable.

Proof: We let $A_i = \{a_{i,j} : j = 1, 2, \ldots\}$ (which we can do since each A_i is countable). Now we arrange the elements $a_{i,j}$ in an array

$$\begin{array}{ccccc} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} & \cdots \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & \cdots \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} & \cdots \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{array}$$

We define $f : I\!N \to A$ by $f(1) = a_{1,1}$, $f(2) = a_{2,1}$, $f(3) = a_{1,2}$, $f(4) = a_{3,1}$, $f(5) = a_{2,2}$, $f(6) = a_{1,3}$, etc. In general $f(n) = a_{i,j}$ where $N(N+1)/2 < n \leq (N+1)(N+2)/2$, $i + j = N + 2$ and $j = n - N(N+1)/2$. ∎

As a simple example, for $n = 5$ we see that $3 = (2)(2+1)/2 < 5 \leq (3)(4)/2 = 6$ so $N = 2$ and thus $i + j = 2 + 2 = 4$ and $j = 5 - 3 = 2$ so $i = 2$.

As an exercise, find an explicit formula for i and j given n.

The next result gives us one example of an uncountable set (thus showing that the definition is not an empty definition). Using this theorem, and the problems we can construct uncountably many examples of uncountable sets. For example, since $(0, 1)$ is uncountable, so are $[-1, 1]$ and $(0, 1) \cup \{2\} \cup [100, 500)$. Furthermore, any interval (a, b) with $a < b$ is uncountable since we have the bijection $f : (0, 1) \to (a, b)$ given by $f(x) = (b - a)x + a$.

Theorem 1.8 *(Cantor) The interval $(0, 1)$ is uncountable.*

Proof: Suppose that $(0, 1)$ is countable and let $f : I\!N \to (0, 1)$ be a surjection. Let $f_n = f(n)$ and consider the list:

$$\begin{array}{ccl} f_1 & = & .a_{1,1}a_{1,2}a_{1,3}a_{1,4}\cdots \\ f_2 & = & .a_{2,1}a_{2,2}a_{2,3}a_{2,4}\cdots \\ f_3 & = & .a_{3,1}a_{3,2}a_{3,3}a_{3,4}\cdots \\ f_4 & = & .a_{4,1}a_{4,2}a_{4,3}a_{4,4}\cdots \\ \vdots & \vdots & \quad\vdots\ \ \vdots\ \ \vdots\ \ \vdots\ \ \ddots \end{array}$$

where these are the decimal representations of f_i. We define a number b by specifying all the decimal digits of b. We do this by defining

$$b_i = \begin{cases} 1, & \text{if } a_{i,i} \neq 1 \\ 2, & \text{if } a_{i,i} = 1. \end{cases}$$

We can see that $b \in (0, 1)$ since all its decimal digits are either 1 or 2.

We claim that $b \neq f_n$ for any n. Suppose, on the contrary, that $b = f_n$ for some n. Then all the decimal digits of b and f_n must agree, that is, $b_i = a_{n,i}$ for all i. However, $b_n \neq a_{n,n}$ by construction. Thus, b is not in the range of f so f cannot be surjective. Thus no such surjection can exist. Therefore, $(0, 1)$ is uncountable. ∎

Definition 1.7 *Let A and B be sets. We say that the* cardinality of A is less than or equal to the cardinality of B*, written as $card(A) \leq card(B)$, if there is an injection $f : A \to B$. Furthermore, we say that $card(A) = card(B)$ if there is a bijection $f : A \to B$.*

The next theorem indicates that the notation makes sense as an order relation. We don't give the proof.

Theorem 1.9 *(Schroeder-Bernstein Theorem) Let A and B be sets. Then $card(A) \leq card(B)$ and $card(B) \leq card(A)$ together imply that $card(A) = card(B)$.*

We could also define $card(A) \leq card(B)$ using surjections, as the next theorem indicates.

Theorem 1.10 *Let A and B be sets. Then $card(A) \leq card(B)$ iff there is a surjection $g : B \to A$.*

Proof: We assume that $A \neq \emptyset$ (else all functions $g : B \to A$ are surjections and all functions $f : A \to B$ are injections, so there is nothing to prove).

Suppose first that $card(A) \leq card(B)$, that is, there is an injection $f : A \to B$. We need to prove the existence of a surjection $g : B \to A$. Let $a_0 \in A$ be some fixed element. For each $b \in Range(f)$, there is a unique $a_b \in A$ with $f(a_b) = b$ since f is injective. Define $g(b) = a_b$ for such a b. Now, for all $b \notin Range(f)$, we simply define $g(b) = a_0$. This defines a surjective function $g : B \to A$.

Insert a nice picture of the situation here

Conversely, suppose $g : B \to A$ is a surjection. Then for each $a \in A$ there is **some** $b_a \in B$ so that $g(b_a) = a$ (there could certainly be more than one such b_a – we pick one). Now define $f(a) = b_a$. Then this defines a function $f : A \to B$. If $b_1 = f(a_1) = f(a_2) = b_2$, then $a_1 = g(b_1) = g(b_2) = a_2$ (by construction) so f is injective. ∎

Definition 1.8 *Let A be a set. The* power set of A *is the set*

$$Pow(A) = \{S : S \subset A\}.$$

Example If $A = \{a, b, 1\}$ then

$$Pow(A) = \{\emptyset, \{a\}, \{b\}, \{1\}, \{a, b\}, \{a, 1\}, \{b, 1\}, \{a, b, 1\}\}.$$

Theorem 1.11 *Let S be a set. Then $card(S) < card(Pow(S))$.*

Proof: The function $g : S \to Pow(S)$ given by $g(s) = \{s\}$ is clearly injective, so $card(S) \leq card(Pow(S))$.

Suppose that $card(S) = card(Pow(S))$. Then there is a bijection $f : S \to Pow(S)$. Let

$$T = \{x \in S : x \notin f(x)\}.$$

We claim that $T \notin Range(f)$. Suppose that it is, that is that there is some $y \in S$ with $f(y) = T$. If $y \in T$ then, by definition of T, $y \notin f(y) = T$. Thus we must have that $y \notin T = f(y)$. However, then by definition of T we must have that $y \in T$. This contradiction shows that T cannot be in the range of f and thus that f is not onto. ■

Problems

1. Let X be a set and $Pow(X) = \{A : A \subset X\}$ be the *power set of* X (that is, the set of all subsets of X). Define the relation $\smile$ on $Pow(X)$ by $A \smile B$ if $A \cap B \neq \emptyset$. Is $\smile$ and equivalence relation? Is $\smile$ a function? Is $\smile$ a partial order?

2. Find an example of a collection of distinct non-empty sets $\{A_n\}$ so that for ANY $F \neq I\!N$ and $F \subset I\!N$ we have $\bigcap_{n \in F} A_n \neq \emptyset$ but $\bigcap_n A_n = \emptyset$.

3. Find $f^{-1}(B)$ and $f(A)$ for each of the examples below.

   (a) $f : I\!R \to I\!R$ given by $f(x) = x^2 - x$, $A = [0, 3]$ and $B = [1, 2]$.

   (b) $f : I\!N \to I\!N$ given by $f(x) = n^2 - n + 1$, $A = \{1, 2, 3, \ldots, 20\} = B$.

4. Give a simple description of each of the following sets.

   - $\displaystyle\bigcap_{n=1}^{i} nfty(-1/n, 1 + 1/n).$
   - $\displaystyle\bigcup_{x \in [0,1]} (4x(x-1)(x+1), -4x(x-1)(x+1))$
   - $\displaystyle\bigcap_{x \in [0,1]} [4x(x-1)(x+1), -4x(x-1)(x+1)]$

5. - Let $f : I\!R \to I\!R$ be given by $f(x) = 4x(x+1)(x-1)$. Describe the sets $f(\ (-1, 1)\ )$ and $f^{-1}(\ [1, 2)\ )$.

- Let $f : I\!N \to \mathbb{Z}$ be given by $f(n) = (-1)^n n^2$. Describe the sets $f^{-1}(\ \{z \in \mathbb{Z} : 10 \le z \le 100\}\ )$ and $f^{-1}(\ \{z \in \mathbb{Z} : z \ge 1000\}\ )$.

6. Give an example of a function $f : I\!R \to I\!R$ such that

$$f^{-1}(\ (0,1)\ ) = \bigcup_{n\in\mathbb{Z}} (2n, 2n+1) = \cdots\cup(-4,-3)\cup(-2,-1)\cup(0,1)\cup(2,3)\cup(4,5)\cup\cdots$$

   Give an example of a bijection $g : [0,4] \to [0,2]$ so that $g^{-1}(\ (0,1)\ ) = (0,1] \cup (2,3)$ and $g^{-1}(\ (1,2)\ ) = (1,2] \cup (3,4)$.

7. Let $X = \{1,2,3\}$. Find an example of a function $f : X \to X$ and a set $A \subset X$ so that $f^{-1}(A) \ne \emptyset$ but $f^{-1}\left(f^{-1}(A)\right) = \emptyset$.

8. Let $f : X \to Y$ be a function. Show that if f is not injective then you can find a set Z and two functions $g, h : Z \to X$ with $f(g(z)) = f(h(z))$ for all $z \in Z$ but $g(z) \ne h(z)$ for at least one $z \in Z$.

9. Let X, Y be sets and $f : X \to Y$. Prove that the following conditions are equivalent:

   (a) f is injective.
   (b) $f(A \setminus B) = f(A) \setminus f(B)$ for all $B \subset A \subset X$
   (c) $f^{-1}(f(S)) = S$ for all $S \subset X$.

10. Define the function $f : [0,1] \to (0,1]$ by

$$f(x) = \begin{cases} 0, & \text{if } x = 1 \text{ or } x = 0 \\ 1, & \text{if } x = 1/2 \\ (2\lfloor i/2\rfloor + 1)/2^{n-1}, & \text{if } x = (2i+1)/2^n \\ x, & \text{otherwise.} \end{cases}$$

   (Here $\lfloor x\rfloor$ is defined to be the largest integer smaller than or equal to x). With this definition, for example, $f(1/4) = f(3/4) = 1/2$ and $f(1/8) = f(3/8) = 1/4$, $f(5/8) = f(7/8) = 3/4$.

   Show that f is onto but is not one-to-one on any interval.

11. Let $f : X \to Y$. Show that

   (a) f is surjective iff there exists a function $g : Y \to X$ so that $f(g(y)) = y$ for all $y \in Y$.
   (b) f is injective iff there exists a function $g : Range(X) \to X$ so that $x = g(f(x))$ for all $x \in X$.
   (c) f is surjective iff for all sets Z and all $g, h : Y \to Z$ we have that $g \circ f = h \circ f$ implies that $g = h$.
   (d) f is injective iff for all sets Z and all $g, h : Z \to X$ we have that $f \circ g = f \circ h$ implies that $g = h$.

12. Let X be a set and $A_n \subset X$ for each $n \in I\!N$. Define the two sets $\bar{A}$ and $\underline{A}$ by

$$\bar{A} = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

and

$$\underline{A} = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k.$$

   (a) Prove that

$$\bigcap_{n=1}^{\infty} A_n \subset \underline{A} \subset \bar{A} \subset \bigcup_{n=1}^{\infty} A_n.$$

   (b) Prove that

$$\bar{A} = \{x \in X : x \in A_n \text{ for all but finitely many } n\}.$$

   (c) Prove that

$$\underline{A} = \{x \in X : x \in A_n \text{ for infinitely many } n\}.$$

13. This problem deals with composition of functions.

   (a) Show that there exist sets X, Y, Z and functions $f : Y \to Z$ and $g, h : X \to Y$ so that $f \circ g = f \circ h$ but $g \neq h$. What property is necessary from f for $f \circ g = f \circ h$ to imply that $g = h$?

   (b) Show that there exist sets X, Y, Z and functions $g, h : Y \to Z$ and $f : X \to Y$ so that $g \circ f = h \circ f$ but $g \neq h$. What property is necessary from f for $g \circ f = h \circ f$ to imply that $g = h$?

14. In this problem we consider functions induced on the powerset of a set.

   (a) Let $f : X \to Y$. Define the function $F : Pow(X) \to Pow(Y)$ by $F(S) = f(S)$. What conditions on f ensure that F is injective? What conditions on f ensure that F is surjective? Show that if f is a bijection then so is F.

   (b) Let $f : X \to Y$. Define the function $G : Pow(Y) \to Pow(X)$ by $G(S) = f^{-1}(S)$. What conditions on f ensure that G is injective? What conditions on f ensure that G is surjective? Show that if f is a bijection then so is G.

15. Find a bijection $f : \{1/n : n \in I\!N\} \to \{1/n : n \in I\!N\} \cup \{0\}$.

16. Find a bijection $f : (0, 1) \to [0, 1]$.

17. This question deals with Cartesian products of sets.

(a) Suppose that A is a non-empty set and ϕ is the emptyset. Is $\phi \times A$ non-empty? Prove or disprove.

(b) Suppose that A, B, C are non-empty sets. Is $(A \times B) \times C$ the same set as $A \times (B \times C)$? Prove or disprove.

18. Suppose that $A, B \subset X$ and $C \subset Y$. Prove that

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

19. Let A, B, C be sets.

(a) Show that there is a bijection $f : A \cup B \to B \cup A$ (yes it isn't a trick question).

(b) Show that there is a bijection $f : A \times B \to B \times A$.

(c) Show that there is a bijection $f : A \times (B \cup C) \to (A \times B) \cup (A \times C)$.

(d) Find a set X so that for each set A there is a bijection $f : X \cup A \to A$.

(e) Find a set X so that for each set A there is a bijection $f : X \times A \to A$.

20. Let $T = \{0, 1\}$ and X be some set. Define the set Y by

$$Y = \{f : X \to T\}$$

(so that Y is the set of all possible functions from X to T).

(a) If $X = \{a, b, c, d\}$, list all the elements of Y.

(b) Prove that there is never a surjection $\Phi : X \to Y$.

21. For two sets X, Y, we define X^Y to be the set

$$X^Y = \{f : Y \to X\},$$

that is, X^Y is the set of all possible functions from Y to X.

(a) List all the elements of X^Y for $X = \{0, 1\}$ and $Y = \{0, 1, 2\}$.

(b) Assume that $Y \neq \emptyset$. Is $card(Y) \leq card(X^Y)$ for all X, Y? Find a condition on X so that $card(Y) \leq card(X^Y)$. Prove that $card(X) \leq card(X^Y)$ by finding an injection $\phi : X \to X^Y$.

(c) Assume that $A \cap B = \emptyset$. Prove that there is a bijection between $X^{A \cup B}$ and $X^A \times X^B$.

(d) Prove that there is a bijection between $(X^Y)^Z$ and $X^{Y \times Z}$ for all sets X, Y, Z.

As a comment, we can think of $\mathbb{R}^2$ as $\mathbb{R}^{\{1,2\}}$ and $\mathbb{R}^3$ as $\mathbb{R}^{\{1,2,3\}}$ so $\mathbb{R}^\infty$ could be thought of as $\mathbb{R}^{\mathbb{N}}$.

22. Assume that $Y \neq \emptyset$. Prove that X^Y is uncountable iff either X is uncountable or X has at least two elements and Y is infinite.

23. Let X be a finite set and $f : X \to X$ be a function. Let $x_1 \in X$ be fixed and $x_{n+1} = f(x_n)$. Suppose that $x_{101} = x_1$. Show that $x_2 = x_{102}$ and $x_3 = x_{103}$. Show that $x_{n+100} = x_n$ for any $n \in I\!\!N$.

24. Let E be a finite set. Prove that $f : E \to E$ is injective iff it is surjective. Give an example to show that this isn't true for infinite sets.

25. Show that any subset of $I\!\!N$ is countable.

26. Show that any finite set is countable.

27. Show that if A is infinite and $A \subset B$ then B is infinite as well.

28. Show that if A is finite and $B \subset A$ then B is finite as well.

29. Show that if A is uncountable and $A \subset B$ then B is uncountable as well.

30. Show that if A is countable and $B \subset A$ then B is countable as well.

31. Show that A is infinite iff there exists an injection $f : I\!\!N \to A$.

32. Suppose that S is a countably infinite set (that is, it is countable and it is infinite). Prove that there is a bijection $f : I\!\!N \to S$.

33. Prove that any finite subset of $I\!\!R$ is bounded.

34. Show that the finite union of finite sets is finite.

35. Show that the finite product of finite sets is finite.

36. Show that the finite product of countable sets is countable.

37. Prove that the set of all functions $f : I\!\!N \to \{0, 1\}$ is uncountable.

38. Let Λ_n be a set for each $n \in I\!\!N$ with the property that there are at least two elements in each Λ_n. Consider the set of all functions

$$T = \{f : I\!\!N \to \bigcup_n \Lambda_n : f(n) \in \Lambda_n\}.$$

Prove that T is uncountable.

39. Let $T = \{0, 1\}$ and X be some nonempty set. Define the set Y by

$$Y = \{f : X \to T\}$$

(so that Y is the set of all possible functions from X into T).

(a) If $X = \{a, b, c, d\}$, list all the elements of Y.
(b) Prove that there is no surjection $f : X \to Y$.

40. Prove that if A is finite and $g : A \to B$ is a surjection then B is finite as well.

41. Prove that if A is countable and $g : A \to B$ is a surjection then B is countable as well.

42. Prove that if A is infinite and $g : A \to B$ is an injection then B is infinite as well.

43. Prove that if A is uncountable and $g : A \to B$ is an injection then B is uncountable as well.

44. Find an example of a sequence of sets A_n so that A_n is countably infinite for all n and $A_{n+1} \subset A_n$ with $A_n \neq A_{n+1}$ and $\cap_n A_n$ is infinite.

45. Let E be an infinite set. Prove that there exists a proper subset $S \subset E$ (that is, $S \neq E$) and a bijection $f : E \to S$. (**Hint:** use the fact that there is a injection $g : I\!N \to E$ and let $S = E \setminus \{f(1)\}$).

46. Let A be a finite set (an "alphabet"). Let W be the set of all "words" from the "alphabet" A. That is,

$$W = \{f : \{1, 2, \ldots, N\} \to A : N \in I\!N\}.$$

    Prove that W is countable.

47. Let B be an uncountable set and $A \subset B$. If A is countable, prove that $B \setminus A$ is uncountable.

48. Let X be a set and $f : X \to X$ be a function. Define $A_0 = X$ and $A_{n+1} = f(A_n)$.

    (a) Prove that $B \subseteq C$ implies that $f(B) \subseteq f(C)$.
    (b) Prove that $A_{n+1} \subseteq A_n$ for all n.
    (c) Let $A = \bigcap_n A_n$. Prove that $f(A) \subseteq A$.
    (d) We now show that it is not necessarily true that $A \subset f(A)$ by giving an example. Let $X = \{(i, j) : i, j \in I\!N, j \geq i\} \cup \{x, a\}$ (x, a are just two "extra" elements). Define $f : X \to X$ by

$$f(i, j) = \begin{cases} x, & \text{if } i = j \\ (i + 1, j) & \text{if } i \neq j \end{cases}$$

    and $f(x) = a = f(a)$. Show that $A_n = \{(i, j) : i \geq n + 1, j \geq i\} \cup \{x, a\}$ and so $\{x, a\} = A = \cap_n A_n$ but $x \notin f(A)$.

49. Give an example of a function $f : [0, 1] \to I\!R$ for which $f([0, 1])$ is a countably infinite collection of disjoint intervals.

Chapter 2

Ordered Fields and the Real Numbers

Ordered Fields

We start with the axioms of a field.

Definition 2.1 *A* field *is a set $I\!F$, containing at least two elements, together with two binary mappings of $I\!F \times I\!F \to I\!F$ called multiplication and addition and usually denoted by $(a, b) \to a + b$ and $(a, b) \to a \cdot b$ (or $(a, b) \to ab$) so that*

1. *For all $a, b, c \in I\!F$, $(a + b) + c = a + (b + c)$.*
2. *For all $a, b \in I\!F$, $a + b = b + a$.*
3. *For all $a, b \in I\!F$, there is a $c \in I\!F$ with $a + c = b$.*
4. *For all $a, b, c \in I\!F$, $a(bc) = (ab)c$.*
5. *For all $a, b \in I\!F$, $ab = ba$.*
6. *For all $a, b \in I\!F$ with $a + b \neq b$, there is a $c \in I\!F$ so that $ac = b$.*
7. *For all $a, b, c \in I\!F$, $a(b + c) = (ab) + (ac) = ab + ac$.*

We use the standard convention that multiplication has higher precedence than addition so that $ab + c = (ab) + c$.

Example The rationals, **Q**, the real numbers $I\!R$ and the complex numbers **C** are all examples of fields.

Example The set $\{0, 1\}$ with $0+0 = 0 = 1+1$, $0+1 = 1$ and $1 \cdot 0 = 0 \cdot 0 = 0$ and $1 \cdot 1 = 1$ is a field.

Proposition 2.1 *Let $\mathbb{F}$ be a field. Then there are two elements $1, 0 \in \mathbb{F}$ so that $0 + a = a$ for all $a \in \mathbb{F}$ and $1 \cdot a = a$ for all $a \in \mathbb{F}$. These elements are the unique ones with these properties. Furthermore, $0 \cdot a = 0$ for all $a \in \mathbb{F}$ and $1 \neq 0$.*

Proof: Let $x \in \mathbb{F}$ (we know that there is such an element since $\mathbb{F}$ is not empty). Then there is an element $0 \in \mathbb{F}$ so that $0 + x = x$. Now, for any element $a \in \mathbb{F}$ there is an element b so that $a = b + x$. Then we have

$$a + 0 = (b + x) + 0 = b + (x + 0) = b + x = a.$$

Thus, $a + 0 = a$ for all $a \in \mathbb{F}$.

Suppose that for some elements $a, b \in \mathbb{F}$ we have that $a + b = a$. Then there is an element $c \in \mathbb{F}$ so that $a + c = 0$ which implies that

$$b = b + 0 = b + (a + c) = (b + a) + c = a + c = 0.$$

Thus if $b + a = a$ then $b = 0$. Therefore, 0 is the unique element with $a + 0 = a$ for any $a \in \mathbb{F}$.

Now choose any $y \neq 0$ in $\mathbb{F}$. Then for all $a \in \mathbb{F}$ we have that $y + a \neq a$ so there is an element $1 \in \mathbb{F}$ with $1 \cdot y = y$. For all $a \in \mathbb{F}$ there is some $b \in \mathbb{F}$ so that $by = a$. Thus,

$$a \cdot 1 = (by) \cdot 1 = b(y \cdot 1) = by = a.$$

To prove that 1 is unique, suppose that $ab = a$ with $a \neq 0$. Then there is some $c \in \mathbb{F}$ with $ac = 1$ and

$$1 = ac = (ab)c = b(ac) = b(1) = b.$$

Finally, if $a \in \mathbb{F}$ we see that

$$a + a \cdot 0 = a \cdot 1 + a \cdot 0 = a(1 + 0) = a \cdot 1 = a$$

so that $a \cdot 0 = 0$. If $1 = 0$ then $a = a \cdot 1 = a(0) = 0$ for all $a \in \mathbb{F}$ which contradicts the fact that $\mathbb{F}$ has at least two elements. ∎

Corollary 2.2 *For each $a, b \in \mathbb{F}$ there is a unique $c \in \mathbb{F}$ with $a + c = b$; we denote this element by $b - a$, or when $b = 0$ by $-a$. If $a \neq 0$ there is a unique element $1/a = a^{-1} \in \mathbb{F}$ with $a \cdot a^{-1} = 1$.*

Proposition 2.3 *For all $a \in \mathbb{F}$ we have $-(-a) = a$.*

Proof: $a + (-a) = 0$ so $-(-a) = a$. ∎

Proposition 2.4 *For all $a \in \mathbb{F}$ we have $-1 \cdot a = -a$.*

Proof: $a + ((-1) \cdot a) = a(1 + -1) = a \cdot 0 = 0$ so $-a = (-1) \cdot a$. ▮

Proposition 2.5 *We have that* $(-1)(-1) = 1$.

Proof: We see that $0 = -1 \cdot 0 = -1 \cdot (1 + -1) = (-1) \cdot 1 + (-1) \cdot (-1) = -1 + (-1)(-1)$. Thus, $1 = -(-1) = (-1)(-1)$. ▮

Definition 2.2 *A field* $I\!\!F$ *is an* ordered field *if there is a total order* $<$ *on* $I\!\!F$ *so that*

1. *if* $a < b$ *and* $c \in I\!\!F$ *then* $a + c < b + c$
2. *if* $a < b$ *and* $c > 0$ *then* $ac < bc$

Notice that $a < b \Leftrightarrow b - a > 0$. Thus, specifying which elements are greater than 0 is sufficient to define the order. That is, if we know the set $P = \{x \in I\!\!F : x > 0\}$ then we can say that $a < b$ iff $b - a \in P$.

Now, if $c < 0$ then $0 = c + (-c) < 0 + (-c) = -c$ and conversely if $-c < 0$ then $c > 0$. Furthermore, if $c < 0$ and $a < b$ then $-c > 0$ so

$$a(-c) < b(-c) \Rightarrow -ac < -bc \Rightarrow 0 < ac - bc \Rightarrow bc - ac = -(ac - bc) < 0 \Rightarrow bc < ac.$$

If $a = 0$ then $a^2 = 0$. On the other hand, if $a > 0$ then $a^2 = a \cdot a > a \cdot 0 = 0$ while if $a < 0$ then $a^2 = a \cdot a > a \cdot 0 = 0$. Thus, in all cases $a^2 \geq 0$ with $a^2 = 0$ only when $a = 0$. This means that $1 = 1^2 > 0$ since $1 \neq 0$. However, then $2 = 1 + 1 > 1 > 0$ so, by induction, $n > 0$ for all $n \in I\!\!N$.

Proposition 2.6 *If* $0 < a < b$ *then* $0 < 1/b < 1/a$.

Proof: First we prove that if $0 < a$ then $0 < 1/a$ as well. Clearly $1/a \neq 0$ since if $1/a = 0$ then $1 = a \cdot 1/a = a \cdot 0 = 0$. Suppose that $1/a < 0$, then $1 = 1/a \cdot a < 0 \cdot a = 0$, which is a contradiction. Thus if $0 < a$ then $0 < 1/a$ as well. Thus we know that $0 < 1/b$ and $0 < 1/a$ since $0 < a < b$. Now

$$1 = 1/a \cdot a < 1/a \cdot b \Rightarrow 1/b = 1/b \cdot 1 < 1/a \cdot b \cdot 1/b = 1/a$$

as desired. ▮

Proposition 2.7 *If* $0 < x < y$ *then* $x^2 < y^2$.

Proof: We see that $0 < x$ and $x < y$ together imply that $x^2 < xy$ and $y > 0$ and $x < y$ together imply that $xy < y^2$. Thus, $x^2 < xy < y^2$ so $x^2 < y^2$. ▮

We can copy the proof of the previous proposition and use induction to prove that if $n \in \mathbb{N}$ and $0 < x < y$ then $x^n < y^n$. To do this, we notice that if $0 < x < y$ and $0 < x^n < y^n$ then $x^{n+1} < xy^n$ and $xy^n < y^{n+1}$ (since $x < y$ and $y^n > 0$). Thus, $x^{n+1} < y^{n+1}$. It is easy to see that $0 < x^{n+1}$ as well, so we get $0 < x^{n+1} < y^{n+1}$.

Proposition 2.8 *For every $x > 0$ and $n \in \mathbb{N}$ there is at most one $y > 0$ with $y^n = x$.*

Proof: Suppose that y_1, y_2 with $y_1^n = x$ and $y_2^n = x$. Now either $y_1 < y_2$ or $y_1 = y_2$ or $y_1 > y_2$.

If $y_1 < y_2$, then $x = y_1^n < y_2^n = x$, which is a contradiction. Thus, $y_1 \geq y_2$.

However, if $y_2 < y_1$ we again have that $x = y_2^n < y_1^n = x$. Therefore, the only possibility is that $y_1 = y_2$. ∎

It is much harder to prove that there is a solution to $y^n = x$.

We have yet to give an example of an ordered field. The primary example of an ordered field is the field $\mathbb{R}$, which we have yet to discuss (or define). However, the field of rational numbers $\mathbf{Q}$ is also an ordered field, as can be easily (if somewhat tediously) verified. We next give an example of an ordered field that is not one of the usual ones.

Example Let $\mathbb{F}$ be the set of rational functions, that is

$$\mathbb{F} = \{p(x)/q(x) : q(x) \neq 0\}.$$

We use the usual rules of arithmetic to add and multiply these expressions (common denominators and all that jazz). We need to put an order on $\mathbb{F}$. To do this, we first say that a polynomial $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ is greater than 0 (the zero polynomial) if $a_k > 0$ where a_k is the first nonzero coefficient of $p(x)$. Thus if $a_0 > 0$ then $p(x) > 0$. Now for a rational function $r(x) = p(x)/q(x)$ we say that $r(x) > 0$ if either both $p > 0$ and $q > 0$ or if both $p < 0$ and $q < 0$.

Definition 2.3 *Let X be a set and $<$ be a partial order on X.*

1. *We say that a subset $S \subset X$ has an* upper bound *(or is bounded from above) if there is some element $M \in X$ so that $s \leq M$ for all $s \in S$.*

2. *We say that S has a* lower bound *(or is bounded from below) if there is an element $m \in X$ so that $m \leq s$ for all $s \in S$.*

3. *We say that $M \in X$ is a* least upper bound *for the set S if M is an upper bound for S and whenever $N \in X$ is another upper bound for S we have that $M \leq N$.*

4. *We say that $m \in X$ is a* greatest lower bound *for the set S if m is an lower bound for S and whenever $n \in X$ is another lower bound for S we have that $n \leq m$.*

5. *We say that $M \in S$ is a* greatest element for *S if there is no element $s \in S$ for which $M < s$.*

6. *We say that $m \in S$ is a* least element for *S if there is no element $s \in S$ for which $s < m$.*

Example With the usual order on $I\!\!N$ and $S = \{10, 20, 30, \ldots, 100\}$ we see that 9 is a lower bound for S while 10 is the greatest lower bound for S. Similarly 2002 is an upper bound for S while 100 is the least upper bound.

Example For $S = (0, 1)$ (with the usual order), we see that 2 is an upper bound for S while $\sup S = 1$. Similarly -100 is a lower bound for S and $\inf S = 0$. In this case, S has neither a greatest nor a least element.

Definition 2.4 *An ordered field $I\!\!F$ is* complete *if for all $S \subset I\!\!F$ whenever S has an upper bound it also has a least upper bound.*

Example The field of rational numbers $\mathbf{Q}$ is not complete since the set

$$S = \{x \in \mathbf{Q} : x^2 < 2\}$$

clearly has an upper bound ($r < 2$ for all $r \in S$) but there is no rational number which functions as a least upper bound (why??).

The real numbers are THE complete ordered field. That is, if $I\!\!F$ is a complete ordered field then there is an isomorphism $\phi : I\!\!F \to I\!\!R$ (an isomorphism preserves all the properties of the field including the arithmetic and order properties).

We denote the real numbers by the symbol $I\!\!R$ (as we have done informally until this point).

Proposition 2.9 *Let $a, x, y \in I\!\!R$. Then*

1. *$x < y + \epsilon$ for all $\epsilon > 0$ iff $x \leq y$.*

2. *$x > y - \epsilon$ for all $\epsilon > 0$ iff $x \geq y$.*

3. *$-\epsilon < a < \epsilon$ for all $\epsilon > 0$ iff $a = 0$.*

Proof: (we only prove the first statement – the others are similar). Suppose that $x < y + \epsilon$ for all $\epsilon > 0$. Suppose further that $x > y$. Let $\epsilon_0 = x - y > 0$. Then we know that $x < y + \epsilon_0 = x$ which is a contradiction. Thus we must have that $x \leq y$.

Conversely, suppose that $x \leq y$ and let $\epsilon > 0$ be given. Now either $x = y$ or $x < y$. If $x < y$ then we know that $y < y + \epsilon$ so $x < y < y + \epsilon$ and thus $x < y + \epsilon$. On the other hand, if $x = y$ we again have that $y < y + \epsilon$ so $x = y < y + \epsilon$. ∎

Definition 2.5 *The* absolute value function $|\ | : I\!R \to I\!R$ *is the function defined by*

$$|x| = \begin{cases} x, & \text{if } x \geq 0 \\ -x, & \text{if } x < 0 \end{cases} .$$

We leave the proof of the following (simple) proposition to the readers.

Proposition 2.10 *The absolute value function satisfies:*

1. $|x| \geq 0$ *for all* $x \in I\!R$.
2. $|xy| = |x||y|$ *for all* $x, y \in I\!R$.
3. $|x| \leq M$ *for* $M > 0$ *iff* $-M \leq x \leq M$

Theorem 2.11 *(The triangle inequality) For all* $x, y \in I\!R$ *we have*

1. $|x + y| \leq |x| + |y|$
2. $|x| - |y| \leq |x - y|$
3. $||x| - |y|| \leq |x - y|$

Proof: Since $|x| \leq |x|$ and $-|x| \leq x \leq |x|$ we have $-(|x| + |y|) \leq x + y \leq |x| + |y|$ so $|x + y| \leq |x| + |y|$.

Continuing we see that

$$|x| - |y| = |x - y + y| - |y| \leq |x - y| + |y| - |y| = |x - y|.$$

Finally, to prove that $||x| - |y|| \leq |x - y|$ we show that

$$-|x - y| \leq |x| - |y| \leq |x - y|.$$

We already have the second part of this chain of inequalities, so for the first notice that $|y| - |x| \leq |y - x| = |x - y|$ so $-|x - y| \leq |x| - |y|$. ∎

Theorem 2.12 *(Approximation of suprema) Let* $E \subset I\!R$ *be such that* $\sup E$ *exists. Then for any* $\epsilon > 0$ *there is a point* $e \in E$ *so that*

$$\sup E - \epsilon < e \leq \sup E.$$

Proof: Suppose that there is no such element $e \in E$. Now since all $e \in E$ satisfy $e \leq \sup E$ (since $\sup E$ is an upper bound for E), this means that there is no element $e \in E$ with $e > \sup E - \epsilon$ so $e \leq \sup E - \epsilon$ for all $e \in E$. However, this means that $\sup E - \epsilon$ is an upper bound for E which is smaller than $\sup E$. This contradicts the definition of $\sup E$, so there must exist an element $e \in E$ with the desired properties. $\blacksquare$

Insert nice picture here

Clearly there is a similar result for $\inf E$. That is, if $\inf E$ exists then for all $\epsilon > 0$ there is a point $e \in E$ with $\inf E \leq e < \inf E + \epsilon$.

Theorem 2.13 *(Archimedean Property) Let $\epsilon > 0$. Then for any $M \in \mathbb{R}$ there is an $n \in \mathbb{N}$ so that $n\epsilon > M$.*

Proof: Consider the set $S = \{n\epsilon : n \in \mathbb{N}\} \subset \mathbb{R}$. If the theorem is false, then $n\epsilon < M$ for all $n \in \mathbb{N}$ so M is an upper bound for S. Thus, $\sup S$ exists. However, by the approximation property for $\sup S$ there is an element $s \in S$ with $\sup S - \epsilon < s \leq \sup S$ which means that there is an $n \in \mathbb{N}$ so that $\sup S - \epsilon < n\epsilon \leq \sup S$ which implies that $\epsilon(n+2) = \epsilon n + 2\epsilon > \sup S$. Now clearly $(n+2)\epsilon \in S$ since $n + 2 \in \mathbb{N}$. Thus we contradict the fact that $\sup S$ is an upper bound for S. This contradiction shows that $\sup S$ cannot exist so S cannot be bounded or there must exist some $n \in \mathbb{N}$ with the desired property. $\blacksquare$

Theorem 2.14 *(Density of the Rationals) If $a, b \in \mathbb{R}$ with $a < b$ then there is a rational q with $a < q < b$.*

Proof: If $a < 0 < b$ then clearly $q = 0$ works. Thus suppose that either $0 \leq a < b$ or $a < b \leq 0$.

Suppose that $0 \leq a < b$. Then we know that $b - a > 0$ so $1/(b-a) > 0$ as well. Thus by the Archimedean Property there is an integer $m \in \mathbb{N}$ with $0 < 1/(b-a) < m$ which implies that $0 < 1/m < b - a$. Define the set

$$E = \{k \in \mathbb{N} : a < \frac{k}{m}\}.$$

By the Archimedean Property we know that $E \neq \emptyset$. Let n be the smallest element in E (which exists since $E \subset \mathbb{R}$ is bounded below by 0). Now $n > 0$ since $0 \notin E$. Thus

$$\frac{n-1}{m} \leq a < \frac{n}{m}.$$

If $n/m \geq b$ then $(n-1)/m \leq a < b \geq n/m$ which implies that $b - a \leq 1/m$, which is a contradiction of the choice of m. Thus we know that $a < n/m < b$, so $q = n/m$ is the desired rational number.

Now if $a < b \leq 0$, then $0 \leq -b < -a$ so there is a rational q with $-b < q < -a$ which implies that $a < -q < b$. $\blacksquare$

We now state some basic properties of $\inf E$ and $\sup E$ for a set $E \subset \mathbb{R}$. We leave the simple proofs to the reader.

Proposition 2.15 *Suppose $A \subset B \subset \mathbb{R}$. If $\sup B$ exists then $\sup A$ exists as well and $\sup A \leq \sup B$. If $\inf B$ exists then $\inf A$ exists as well and $\inf B \leq \inf A$.*

Proposition 2.16 *Let $E \subset \mathbb{R}$ be a nonempty set.*

1. *$\sup E$ exists iff $\inf\{-e : e \in E\} = \inf -E$ exists. In this case, $\sup E = -\inf(-E)$.*

2. *$\inf E$ exists iff $\sup\{-e : e \in E\} = \sup -E$ exists. In this case, $\inf E = -\sup(-E)$.*

Insert one last nice picture for this chapter

Problems

1. Show that $\mathbb{Z}_n$ (the integers with arithmetic modulo $n \in \mathbb{N}$) is a field iff n is a prime number.

2. Let $\mathbb{F} = \{0, 1, a\}$ be a field with three distinct elements. Compute the addition and multiplication table for $\mathbb{F}$.

3. Let $\mathbb{F} = \{0, 1, \alpha, \beta\}$ be a field with four distinct elements. Compute the addition and multiplication tables for $\mathbb{F}$. Notice that $1 + 1 = 0$.

4. Let $\mathbb{F} = \{x \in \mathbb{R} : x > 0\}$ and define the operations $\oplus$ and $\otimes$ on $\mathbb{F}$ by $\alpha \oplus \beta = \alpha\beta$ (regular multiplication in $\mathbb{R}$) and $\alpha \otimes \beta = \alpha^{\ln(\beta)}$. Show that $(\mathbb{F}, \oplus, \otimes)$ is a field.

5. Is it possible for a finite field to be an ordered field? If it is, give an example. If it is not possible, provide a proof that it is not possible.

6. Is it possible to endow the complex numbers $\mathbf{C}$ with a total order to make it an ordered field? If it is, give such an order relation. If it is not, prove that it is not possible.

7. Show that $(1+\epsilon)^n \geq 1+n\epsilon$ for any $\epsilon > 0$ and $n \in I\!N$.

8. Suppose that $a, b, x, y \in I\!R$ with $|x-a| < \epsilon$ and $|y-b| < \epsilon$ for some $\epsilon > 0$. Show that $|xy-ab| < (|a|+|b|)\epsilon + \epsilon^2$.

9. In this problem, we will prove the existence and uniqueness of the number $\sqrt{a}$ for $a > 0$. Define $\alpha = \sup\{r : r^2 = a\}$.

   (a) Show that for any $\epsilon > 0$ we have that $(1+\epsilon)^2 < 1+3\epsilon$. Similarly, show that $(1-\epsilon)^2 > 1-2\epsilon$. Use this to show that $(a(1+\epsilon))^2 < a^2+3\epsilon a^2$ and $(a(1-\epsilon))^2 > a^2-2\epsilon a^2$.

   (b) Show that if $\alpha^2 < a$ then there is some $\epsilon > 0$ with $(\alpha(1+\epsilon))^2 < a$. Use this to prove that $\alpha^2 \geq a$.

   (c) Show that if $\alpha^2 > a$ then there is some $\epsilon > 0$ with $(\alpha(1-\epsilon))^2 > a$. Use this to prove that $\alpha^2 \leq a$.

   (d) Prove that there is at most one number $r \in I\!R$ with $r > 0$ and $r^2 = a$.

10. Suppose that $0 < x < y$. Prove that $0 < \sqrt{x} < \sqrt{y}$ (by $\sqrt{x}$ we mean the number $r > 0$ for which $r^2 = x$; see the previous problem).

11. Let $0 \leq a \leq b$. Prove that $a \leq \sqrt{ab} \leq (a+b)/2 \leq b$ (you need to assume that if $0 \leq x, y$ then $x < y$ iff $\sqrt{x} < \sqrt{y}$).

12. Let $0 < a < b$.

   (a) Prove that $a < \sqrt{ab} < (a+b)/2 < b$.

   (b) Define $a_0 = a$, $b_0 = b$ and for each $n \in I\!N$

   $$a_n = \sqrt{a_{n-1}b_{n-1}}, \quad b_n = \frac{a_{n-1}+b_{n-1}}{2}.$$

   Show that $a_0 \leq a_1 \leq a_2 \leq a_3 \leq \cdots$ and $b_0 \geq b_1 \geq b_2 \geq b_3 \geq \cdots$. Furthermore, show that $a_n \leq b_m$ for any $n, m \in I\!N$.

13. Suppose that $0 < a < 1$ and let $b = 1-\sqrt{1-a}$. Prove that $0 < b < a$.

14. Suppose that $0 < a/2 < x$ and $|x-a| < \epsilon$ for some $\epsilon > 0$. Show that

$$\left|\frac{1}{x}-\frac{1}{a}\right| \leq \frac{2\epsilon}{a^2}.$$

15. First, show that

$$x^n - a^n = (x-a)(x^{n-1}+ax^{n-2}+a^2x^{n-3}+\cdots+a^{n-3}x^2+a^{n-2}x+a^{n-1})$$

for any $n \in \mathbb{N}$ and $x, a \in \mathbb{R}$. Now fix $n \in \mathbb{N}$ and let $0 < a/2x$. Show that there is some $C > 0$ so that

$$|\sqrt[n]{x} - \sqrt[n]{a}| \leq \frac{|x-a|}{C}.$$

16. Let X and Y be subsets of $\mathbb{R}$ and suppose that $f : X \to Y$ is an order preserving function (that is, $x < y$ implies that $f(x) < f(y)$).
    (a) Prove that f is injective.
    (b) Suppose that f is a bijection. Prove that f^{-1} is also order preserving.
    (c) We say that a subset $X \subset \mathbb{R}$ is bounded if it is bounded from below and from above. Find an example of a bounded set $X \subset \mathbb{R}$ and an unbounded set Y and an order preserving bijection $f : X \to Y$.

17. Let $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_n\}$ be finite sets of real numbers. Prove that

$$\max A + \max B \leq \max\{a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n\}$$

and that

$$\min A + \min B \geq \min\{a_1 + b_1, a_2 + b_2, \ldots, a_n + b_n\}.$$

18. Suppose that $x \in E$ is an upper bound for E. Prove that x is the supremum of E. Similarly, suppose that $y \in E$ is a lower bound for E. Prove that y is the infimum of E.

19. Prove that if $S \subset \mathbb{Z}$ is bounded then $\inf S \in S$ and $\sup S \in S$.

20. Let $A, B \subset \mathbb{R}$ with $\sup A$ and $\sup B$ both existing. If we define $A + B = \{a + b : a \in A, b \in B\}$, prove that $\sup(A + B)$ exists. Is $\sup(A + B) < \sup A + \sup B$ or is $\sup A + \sup B < \sup(A + B)$ or are they the same? Prove your answer.

21. Find an example of a bounded set $A \subset \mathbb{R}$ and an unbounded set $B \subset \mathbb{R}$ and an increasing surjection $f : A \to B$.

22. Let $f : \mathbb{R} \to \mathbb{R}$ be an increasing function and $\emptyset \neq E \subset \mathbb{R}$ be a bounded set. Show that $\sup f(E) \leq f(\sup E)$.

23. Suppose that $x, a, y, b \in \mathbb{R}$ with $|x - a| < \epsilon$ and $|y - b| < \epsilon$ for some $\epsilon > 0$. Prove that $|xy - ab| < (|a| + |b|)\epsilon + \epsilon^2$.

24. Suppose that $A \subset B \subset \mathbb{R}$. Prove that $\inf B \leq \inf A$ if $\inf B$ exists (that is, if $\inf B$ exists then $\inf A$ also exists and the inequality holds). Similarly, prove that $\sup B \geq \sup A$.

25. Suppose that $E \subset \mathbb{R}$ is a nonempty set. Define $-E = \{-e : e \in E\}$. Prove that $\sup E = -\inf -E$ and that $\inf E = -\sup -E$.

26. Suppose $A, B \subset I\!R$ and $f : A \to B$ is increasing (that is, if $x < y$ then $f(x) < f(y)$). Prove that f is injective. If f is a bijection, prove that f^{-1} is also increasing.

27. Let
$$x_n = 1 + 1/2! + 1/3! + 1/4! + \cdots + 1/n!$$
and show that $x_n \leq 3 - 1/n!$ for all $n \in I\!N$.

28. Show that $(1 + 1/n)^n \geq 2$ for all $n \geq 1$ by induction. To do this, show that it is true for $n = 1$ and then, for the induction step, show that $(1 + \frac{1}{n+1})^{n+1} \geq (1 + 1/n)^n$.

29. Prove that it is impossible to have an uncountable collection of disjoint subintervals in $I\!R$.

30. Define $f : [0, 1] \to [0, 1]$ by
$$f(x) = \begin{cases} 0, & \text{if } x = 0 \text{ or } x = 1 \\ 1, & \text{if } x = 1/2 \\ (2\lfloor i/2 \rfloor + 1)/2^{n-1}, & \text{if } x = (2i + 1)/2^n \\ x, & \text{if otherwise.} \end{cases}$$
(where $\lfloor y \rfloor$ means the largest integer smaller than or equal to $y \in I\!R$). With this definition, for example, $f(1/4) = f(3/4) = 1/2$ and $f(1/8) = f(3/8) = 1/4$ and $f(5/8) = f(7/8) = 3/4$.

Show that f is surjective but is not injective on any interval (that is, for any $(a, b) \subset [0, 1]$ there are two distinct points $x_1, x_2 \in (a, b)$ with $f(x_1) = f(x_2)$).

Chapter 3

Sequences in $\mathbb{R}$

Definition 3.1 *Let X be a set. A* sequence in X *is a function $a : \mathbb{N} \to X$.*

We usually denote $a(n)$ by a_n and often denote the sequence by either $\{a_n\}$ or $\{a_n\}_{n=1}^{\infty}$ or (a_n).

Note that the values $a(n) = a_n$ might all be the same (that is, the range of the function a might be a single point).

Definition 3.2 *A sequence of real numbers (x_n) is said to* converge *to a real number L (the* limit*) if for every $\epsilon > 0$ there is an $N \in \mathbb{N}$ so that whenever $n \geq N$ we have $|x_n - L| < \epsilon$.*

We will denote this in many different ways:

- $x_n \to L$
- $x_n \to L$ as $n \to \infty$
- $\lim_{n\to\infty} x_n = L$
- The limit of x_n is L.

Example Let $x_n = 1/n^2$. Then $x_n \to 0$.
Proof: Let $\epsilon > 0$ be given. By the Archimedean Property, there is an $N \in \mathbb{N}$ so that $0 < 1/\epsilon < N$ which implies that $0 < 1/N < \epsilon$. We assume that $N > 1$ since if it isn't we can simply increase it by one. Thus $1/N^2 < 1/N < \epsilon$. Then for any $n \geq N$ we have that $1/n^2 \leq 1/N^2 < 1/N < \epsilon$. ∎

Similarly we can prove that for each fixed $k \in \mathbb{N}$ we have that $1/n^k \to 0$.

Proposition 3.1 *A sequence has at most one limit.*

Proof: Suppose that $x_n \to x$ and $x_n \to y$. Let $\epsilon > 0$ be given. Then there exist numbers $N_1, N_2 \in I\!N$ so that for all $n \geq N_1$ we have that $|x_n - x| < \epsilon/2$ and for all $n \geq N_2$ we have that $|x_n - y| < \epsilon/2$. Let $N = \max\{N_1, N_2\}$. Then for all $n \geq N$ we know

$$|x - y| = |x - x_n + x_n - y| \leq |x - x_n| + |y - x_n| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Since $|x - y| < \epsilon$ for any $\epsilon > 0$, it must be the case that $|x - y| = 0$ so $x = y$. ∎

Theorem 3.2 *Every convergent sequence is bounded.*

Proof: Suppose that $x_n \to x$. Then there is an $N \in I\!N$ so that for all $n \geq N$ we have $|x_n - x| < 1$ which implies that for all $n \geq N$ we know $|x_n| < |x| + 1$. Let

$$M = \max\{|x_1|, |x_2|, \ldots, |x_N|, |x| + 1\}.$$

Then for $n \geq N$ we have $|x_n| < |x| + 1 \leq M$ and for all $n = 1, 2, \ldots, N$ we have $|x_n| \leq M$ by definition of M. Thus for all n we know $|x_n| \leq M$ so the sequence is bounded. ∎

Notice that the converse is not true – that is, just because a sequence is bounded does not mean that it converges. The following example illustrates this:

Example Let $x_n = (-1)^n$. Then $|x_n| = 1$ so (x_n) is bounded by x_n does not converge.

Theorem 3.3 *(Squeeze Theorem) Suppose (x_n), (y_n) and (z_n) are real sequences.*

1. *If $x_n \to x$ and $y_n \to x$ and $x_n \leq z_n \leq y_n$ for all $n \geq N_0$ then $z_n \to x$ as well.*
2. *If $x_n \to 0$ and (y_n) is bounded then $x_n y_n \to 0$ as well.*

Proof: 1) Let $\epsilon > 0$. Then there are numbers $N_1, N_2 \in I\!N$ so that $n \geq N_1$ implies that $|x_n - x| < \epsilon$ and $n \geq N_2$ implies that $|y_n - x| < \epsilon$. Let $N = \max(N_0, N_1, N_2)$. Then for all $n \geq N$ we have $|x_n - x| < \epsilon$ and $|y_n - x| < \epsilon$ so $x_n > x - \epsilon$ and $y_n < x + \epsilon$ so

$$x - \epsilon < x_n \leq z_n \leq y_n < x + \epsilon \quad \Rightarrow \quad |z_n - x| < \epsilon$$

so $z_n \to x$ as well.

2) Suppose that $M > 0$ with $|y_n| \leq M$ for all n. Let $N \in I\!N$ be so that if $n \geq N$ then $|x_n| = |x_n - x| < \epsilon/M$. Then for $n \geq N$ we know that

$$|x_n y_n - 0| = |x_n y_n| \leq M|x_n| < M(\epsilon/M) = \epsilon$$

so $x_n y_n \to 0$. ∎

Theorem 3.4 *Let $E \subset \mathbb{R}$ with $E \neq \emptyset$. If E has a supremum (respectively an infimum) then there is a sequence $x_n \in E$ with $x_n \to \sup E$ (respectively, $x_n \to \inf E$) as $n \to \infty$.*

Proof: Suppose that $\sup E$ exists. For each $n \in \mathbb{N}$ let $x_n \in E$ be such that

$$\sup E - 1/n < x_n \leq \sup E.$$

Then by the Squeeze Theorem we know that $x_n \to \sup E$ (since $\sup E - 1/n \to \sup E$ and $\sup E \to \sup E$ as $n \to \infty$).

The proof for $\inf E$ is similar. ∎

The next theorem investigates the relationship between limits and the arithmetic operations on sequences.

Theorem 3.5 *Suppose that (x_n) and (y_n) are convergent real sequences and $\alpha \in \mathbb{R}$. Then*

1. $\lim_{n\to\infty} x_n + y_n = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n$
2. $\lim_{n\to\infty} \alpha x_n = \alpha \lim_{n\to\infty} x_n$
3. $\lim_{n\to\infty} x_n y_n = (\lim_{n\to\infty} x_n)(\lim_{n\to\infty} y_n)$
4. *If* $\lim_{n\to\infty} y_n \neq 0$ *then*

$$\lim_{n\to\infty} x_n/y_n = \frac{\lim_{n\to\infty} x_n}{\lim_{n\to\infty} y_n}.$$

Proof: Suppose that $x_n \to x$ and $y_n \to y$.

1) Let $\epsilon > 0$ be given. Choose $N_1, N_2 \in \mathbb{N}$ so that $n \geq N_1$ implies that $|x_n - x| < \epsilon/2$ and $n \geq N_2$ implies that $|y_n - y| < \epsilon/2$. Then if $n \geq N = \max(N_1, N_2)$ we know that

$$|x_n + y_n - (x + y)| \leq |x_n - x| + |y_n - y| < \epsilon/2 + \epsilon/2 = \epsilon$$

and thus $x_n + y_n \to x + y$ as desired.

2) Using the Squeeze Theorem (part 2) we see that $\alpha(x_n - x) \to 0$ since $z_n = \alpha$ is bounded and $x_n - x \to 0$. Thus, $\alpha x_n - \alpha x \to 0$ so $\alpha x_n = \alpha x_n - \alpha x + \alpha x \to 0 + \alpha x = \alpha x$ by part 1 of this theorem.

3) Since $x_n \to x$, the sequence (x_n) is bounded. Furthermore $x_n - x \to 0$ and $y_n - y \to 0$ so by the Squeeze Theorem (part 2) we know that $x_n(y_n - y) \to 0$ and $y(x_n - x) \to 0$. Now, notice that

$$x_n y_n - xy = x_n y_n - x_n y + x_n y - xy = x_n(y_n - y) + y(x_n - x)$$

so $x_n y_n - xy \to 0$ or $x_n y_n \to xy$.

4) First we prove that $1/y_n \to 1/y$. To see this, let $\epsilon > 0$ be given and let $N_1 \in I\!N$ be such that for $n \geq N_1$ we have $|y_n - y| < |y|/2$ so that $|y_n| > |y|/2$ for $n \geq N_1$. Now choose $N \geq N_1$ so that for all $n \geq N$ we have

$$|y_n - y| < \frac{|y|^2 \epsilon}{2}.$$

Then for all $n \geq N$ we have

$$\left| \frac{1}{y_n} - \frac{1}{y} \right| = \left| \frac{y_n - y}{y_n y} \right| < \frac{|y_n - y|}{|y|^2/2} < \left(\frac{2}{|y|^2} \right) \left(\frac{|y|^2}{2} \right) \epsilon = \epsilon.$$

But then by part 3 of this theorem we have $x_n/y_n = x_n \cdot 1/y_n \to x \cdot 1/y = x/y$. ▮

Finally we have one more standard theorem about limits.

Theorem 3.6 *(Comparison Theorem) If $x_n \to x$ and $y_n \to y$ and $x_n \leq y_n$ for $n \geq N_0$ then $x \leq y$.*

Proof: Suppose that $x > y$ and let $\epsilon = x - y$. Then there is some $N \in I\!N$ so that if $n \geq N$ we have $|x_n - x| < \epsilon/2$ and $|y_n - y| < \epsilon/2$. But then

$$y_n < y + \epsilon/2 = x - \epsilon/2 < x_n$$

which is a contradiction. Thus $x \leq y$. ▮

Monotone Sequences

Definition 3.3 *Let (x_n) be a real sequence.*

1. *(x_n) is an* increasing *(respectively,* strictly increasing*) sequence if $x_n \leq x_{n+1}$ (respectively $x_n < x_{n+1}$).*
2. *(x_n) is a* decreasing *(respectively,* strictly decreasing*) sequence if $x_n \geq x_{n+1}$ (respectively $x_n > x_{n+1}$).*
3. *(x_n) is a* monotone *sequence if it is either increasing or decreasing.*

Notation: If $x_n \to x$ and (x_n) is increasing we denote this by $x_n \nearrow x$ (similarly for $x_n \searrow x$).

Theorem 3.7 *(Monotone Sequences Converge)*

1. *If (x_n) is increasing and bounded above then $x_n \nearrow x$ for some x.*
2. *If (x_n) is decreasing and bounded below then $x_n \searrow x$ for some x.*

Proof: Suppose that (x_n) is increasing and bounded from above by M, that is $x_n \le M$ for all $n \in I\!N$. Then the set $E = \{x_1, x_2, x_3, \ldots\}$ is bounded above by M so $x = \sup E$ exists. Let $\epsilon > 0$. Then by the approximation property for $\sup E$ there is an $x_N \in E$ with $x - \epsilon < x_N \le x$. Then for any $n \ge N$ we know that

$$x - \epsilon < x_N \le x_n \le x \quad \Rightarrow \quad |x_n - x| < \epsilon$$

(since (x_n) is an increasing sequence). Thus $x_n \to x$.

The proof for (x_n) decreasing is similar. ∎

Example Let $x_1 = 1$ and $x_{n+1} = \sqrt{2x_n}$. We show that (x_n) is monotone increasing and bounded above so has a limit.

To do this notice that $x_2 = \sqrt{2}$ and $x_3 = \sqrt{2\sqrt{2}} > \sqrt{2}$ so $x_1 < x_2 < x_3 < 2$. We will prove by induction that $x_n < x_{n+1} < 2$. We already know that $x_1 < x_2 < 2$. Thus suppose that $x_n < x_{n+1} < 2$. Then

$$\begin{array}{rcl} 2x_n < & 2x_{n+1} & < 4 \\ \sqrt{2x_n} < & \sqrt{2x_{n+1}} & < 2 \\ x_{n+1} < & x_{n+2} & < 2 \end{array}$$

as desired.

Thus (x_n) is increasing and bounded above by 2, so must have a limit.

Suppose we say $x_n \to x$ and we wish to compute x. Now, $x_{n+1} \to x$ as well and $2x_n \to 2x$ and we will prove that $\sqrt{2x_n} \to \sqrt{2x}$. Thus we have the equation

$$x \leftarrow x_{n+1} = \sqrt{2x_n} \to \sqrt{2x} \qquad \Rightarrow \qquad x^2 = 2x \quad \Rightarrow \quad x(x-2) = 0.$$

Since $x_n > 1$ we know that $x \ge 1 > 0$ so $x \ne 0$ and thus $x = 2$.

Nested Cells Property

Theorem 3.8 *(Nested Cells) Let $I_n \subset I\!R$ be a closed bounded interval for each $n \in I\!N$ and suppose that $I_1 \supset I_2 \supset I_3 \supset \cdots$ (that is, $I_{n+1} \subset I_n$ for all n). Then there exists an $x \in \bigcap_n I_n$.*

Proof: Let $I_n = [a_n, b_n]$. Then by the nested property we know that $a_1 \le a_2 \le a_3 \le \cdots$ and $b_1 \ge b_2 \ge b_3 \ge \cdots$ (that is, (a_n) is increasing and (b_n) is decreasing).

We claim that $a_n \le b_m$ for all $n, m \in I\!N$. Now either $n \ge m$ or $m \ge n$, so without loss of generality suppose that $m \ge n$. Then $I_m \subset I_n$ so $[a_m, b_m] \subset [a_n, b_n]$ which implies that $a_n \le a_m \le b_m \le b_n$ so $a_n \le b_m$ as claimed.

Thus (a_n) is an increasing sequence that is bounded above by b_m (for any m) and (b_n) is a decreasing sequence that is bounded below by a_m (for any

m). Thus $a_n \nearrow a$ with $a \leq b_m$ for all m and $b_n \searrow b$ with $b \geq a_m$ for all m. Furthermore, $a \leq b$ (since $a_n \leq b_n$). Now let $x \in [a, b]$. Then

$$a_n \leq a \leq x \leq b \leq b_n$$

for all n so $x \in [a_n, b_n] = I_n$ for all n so $x \in \bigcap_n I_n$. ▮

Definition 3.4 *By a* subsequence *of a sequence* (x_n) *we shall mean a sequence of the form* (x_{n_k}) *where each* $n_k \in \mathbb{N}$ *and* $n_1 < n_2 < n_3 < \cdots$.

As an example, for the sequence (x_n) we have the subsequences (x_{2n}) and (x_{n^2}) and (x_{2^n}), among many others. We illustrate these as:

(x_n)	:	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	$\cdots$	x_{16}	$\cdots$
(x_{2n})	:		x_2		x_4		x_6		x_8		$\cdots$	x_{16}	$\cdots$
(x_{n^2})	:				x_4					x_9	$\cdots$	x_{16}	$\cdots$
(x_{2^n})	:		x_2		x_4				x_8		$\cdots$	x_{16}	$\cdots$

The next theorem says that for any bounded sequence *part* of the sequence converges. It is a very important theorem.

Theorem 3.9 *(Bolzano-Weierstrass Theorem) Every bounded real sequence has a convergent subsequence.*

Proof: Let (x_n) be a real sequence and $M > 0$ be such that $|x_n| \leq M$ for all n.

First suppose that the set $S = \{x_n : n \in \mathbb{N}\}$ is finite. Then there is an infinite set $K \subset \mathbb{N}$ and some $x \in \mathbb{R}$ with $x_n = x$ for all $n \in K$. We can define a subsequence (x_{n_k}) of (x_n) by letting $x_{n_k} = x$ for all $n_k \in K$. Clearly $x_{n_k} \to x$ (since this is a constant sequence).

Thus suppose that S is infinite. Let $I_1 = [-M, M]$ and split I_1 into two equal intervals A_1 and B_1 (here $A_1 = [-M, 0]$ and $B = [0, M]$, say). Now either A_1 or B_1 must contain infinitely many elements of S. Let I_2 be one with infinitely many elements of S. Repeating, we split I_2 into two equal closed intervals A_2 and B_2. Again at least one of these two intervals must contain infinitely many elements of S – let I_3 be one of A_2 or B_2 with infinitely many elements of S. Continuing we obtain a collection of closed intervals I_n with

- $I_{n+1} \subset I_n$
- $diam(I_n) = 2M/2^{n-1} = 4M/2^n$.

Thus by the nested cells property there is some $x \in \bigcap_n I_n$. We now construct a subsequence of (x_n) which converges to x.

First choose some $x_{n_1} \in I_1 \cap S$. Since the set $I_2 \cap S$ is infinite, there is some $x_{n_2} \in I_2 \cap S$ with $n_2 > n_1$. Continuing, suppose that we have chosen $x_{n_i} \in I_i \cap S$ with $n_1 < n_2 < \cdots < n_k$. We can choose $x_{n_{k+1}} \in I_{k+1} \cap S$ with $n_{k+1} > n_k$ since $I_{k+1} \cap S$ is infinite. This is the sequence (x_{n_k}) that we desire. Next we show that $x_{n_k} \to x$.

Let $\epsilon > 0$ be given. Then there is an $N \in \mathbb{N}$ so that $4M/2^N < \epsilon$ so for any $k > N$ we have $x_{n_k}, x \in I_k$ so

$$|x_{n_k} - x| < diam(I_k) = 4M/2^k < 4M/2^k < \epsilon$$

so $x_{n_k} \to x$ as $k \to \infty$. ▮

A more explicit construction of the subsequence (x_{n_k}) in the case S is finite might be illuminating, so we do this now. Again let $K = \{n \in \mathbb{N} : x_n = x\}$. Choose $n_1 \in K$ arbitrarily. Then since K is infinite, there must be some $n_2 \in K$ with $n_2 > n_1$. Continuing this way, suppose that we have chosen $n_1 < n_2 < n_3 < \cdots < n_k$ all in K. Then since K is infinite there must be some $n_{k+1} \in K$ with $n_{k+1} > n_k$. This explicitly constructs the subsequence (x_{n_k}) with $x_{n_k} \to x$ as $k \to \infty$.

Cauchy Sequences

Definition 3.5 *A sequence (x_n) is a* Cauchy *sequence if for any $\epsilon > 0$ there is an $N \in \mathbb{N}$ so that for all $n, m \geq N$ we have $|x_n - x_m| < \epsilon$.*

Proposition 3.10 *If (x_n) is a convergent sequence, it is a Cauchy sequence.*

Proof: Suppose that $x_n \to x$ and let $\epsilon > 0$ be given. Then there is an $N \in \mathbb{N}$ so that for any $n \geq N$ we have $|x_n - x| < \epsilon/2$. But then for $n, m \geq N$ we have

$$|x_n - x_m| = |x_n - x + x - x_m| \leq |x_n - x| + |x_m - x| < \epsilon/2 + \epsilon/2 = \epsilon$$

so (x_n) is Cauchy. ▮

Theorem 3.11 *If (x_n) is a Cauchy sequence and there is a subsequence (x_{n_k}) with $x_{n_k} \to x$ then $x_n \to x$ as well.*

Proof: Let $\epsilon > 0$ be given. Then since (x_n) is Cauchy there is an $N_1 \in \mathbb{N}$ so that for all $n, m \geq N_1$ we have $|x_n - x_m| < \epsilon/2$. Similarly since $x_{n_k} \to x$ there is an $N_2 \in \mathbb{N}$ so that for all $k > N_2$ we have $|x_{n_k} - x| < \epsilon/2$. Fix $k \geq N_2$ such that $n_k \geq N_1$. Then for any $n \geq N_1$ we have

$$|x_n - x| \leq |x_n - x_{n_k}| + |x_{n_k} - x| < \epsilon/2 + \epsilon/2 = \epsilon$$

so $x_n \to x$ as desired. ▮

Theorem 3.12 *(Cauchy) A real sequence converges iff it is a Cauchy sequence.*

Proof: We already proved that if $x_n \to x$ then (x_n) is a Cauchy sequence. Thus suppose that (x_n) is a Cauchy sequence. Choose $N \in \mathbb{N}$ so that $n, m \geq N$ implies that $|x_n - x_m| < 1$. This implies that for $n \geq N$ we have $|x_n| < |x_N| + 1$ which in turn implies that (x_n) is bounded by $\max\{|x_1|, |x_2|, \ldots, |x_N| + 1\}$.

Now since (x_n) is a bounded sequence by the Bolzano-Weierstrass Theorem there is a subsequence (x_{n_k}) which converges. However then we know that (x_n) must converge as well, since it is a Cauchy sequence. ∎

We comment that just because $|x_n - x_{n+1}| \to 0$ does not mean that the sequence is Cauchy, as the next example shows.

Example Let $x_n = \log(n)$. Then

$$|x_{n+1} - x_n| = \log(n+1) - \log(n) = \log\left(\frac{n+1}{n}\right) \to \log(1) = 0$$

but $\log(n) \to \infty$ so (x_n) does not converge so cannot be Cauchy.

In some contexts just because a sequence is a Cauchy sequence does not mean that it converges. In more general spaces (such as a general metric space), it is possible to have Cauchy sequences which do not converge. For example, suppose we consider only rational numbers. Then the sequence (x_n) defined by $x_0 = 1$ and $x_{n+1} = (x_n^2 + 2)/(2x_n)$ is a Cauchy sequence of rational numbers which does not converge to a rational number (it happens to converge to $\sqrt{2}$). So there are "holes" in the rational numbers as evidenced by the fact that some Cauchy sequences of rational numbers don't converge to rational numbers.

Extended Real Numbers

For some purposes it is convenient to extend the real numbers by adding $\pm\infty$. We call $\mathbb{R} \cup \{\pm\infty\}$ the *extended real numbers*. We think of ∞ as the "right endpoint" of $\mathbb{R}$ and $-\infty$ as the "left endpoint" of $\mathbb{R}$ (realizing, of course, that they aren't really endpoints). We can extend most of the arithmetic to the extended real numbers. For example, we define $a + \infty = \infty$ for all $a \in \mathbb{R}$, $a \cdot \infty = \infty$ for all $a > 0$ and $a \cdot \infty = -\infty$ for all $a < 0$.

WE SPECIFICALLY DO NOT DEFINE $\infty - \infty$, $0 \cdot \infty$ and ∞/∞.

Definition 3.6 *Let (x_n) be a sequence of extended real numbers.*

1. *We say $x_n \to \infty$ if for every $M > 0$ there is an $N \in \mathbb{N}$ so for all $n \geq N$ we have $x_n > M$.*

2. *We say that $x_n \to -\infty$ if for every $M < 0$ there is an $N \in \mathbb{N}$ so for all $n \geq N$ we have $x_n < M$.*

The rules for infinite limits are pretty much the same as the other limit rules as long as we remember that expressions of the form $0 \cdot \infty$ and $\infty - \infty$ and ∞/∞ are undefined.

Example If $x_n \to \infty$ and (y_n) is bounded then $x_n + y_n \to \infty$ as well.
Proof: Since (y_n) is bounded there is a $B > 0$ so that $|y_n| \leq B$ for all n. Since $x_n \to \infty$ there is a $N \in \mathbb{N}$ so that for all $n \geq N$ we have $x_n > M + B$. But then

$$x_n + y_n > M + B + y_n \geq M + B - B = M.$$

Problems

1. Use the definition to prove that $1/\sqrt{n} \to 0$.
2. Use the definition to show that if $a_n \to 0$ then $a_n^2 \to 0$ as well.
3. Let (a_n) and (b_n) be bounded sequences and define $c_n = a_n + b_n$. Show that (c_n) is also a bounded sequence. Show that

$$\sup\{c_n : n \in \mathbb{N}\} \leq \sup\{a_n : n \in \mathbb{N}\} + \sup\{b_n : n \in \mathbb{N}\}$$

   but that the inequality can be strict (that is, find an example where the left hand side is smaller than the right hand side).
4. Let $a > 0$ and $x_1 = 1$ and $x_{n+1} = \sqrt{2x_n + a}$ for $n > 1$. Show that (x_n) is an increasing sequence bounded above by $1 + \sqrt{1+a}$. To do this, use for induction hypothesis that $1 \leq x_{n-1} \leq x_n \leq 1 + \sqrt{1+a}$ and prove $x_n \leq x_{n+1} \leq 1 + \sqrt{1+a}$. Also, it might be useful to notice that

$$x_n^2 - 2x_n - a = (x_n - 1 - \sqrt{1+a})(x_n - 1 + \sqrt{1+a}).$$

5. Let $x_n \to x$ and define the function $\pi : \mathbb{N} \to \mathbb{N}$ by

$$\pi(i) = \begin{cases} i, & \text{if } i < 8 \\ 3 \times 2^n - 1 - i, & \text{if } 2^n \leq i < 2^{n+1} \end{cases}.$$

   Define $y_n = x_{\pi(n)}$ (so we have used π to rearrange the terms of (x_n) to get (y_n)) and show that $y_n \to x$ as well.
6. Let $x_n \to x$ and let $\pi : \mathbb{N} \to \mathbb{N}$ be any bijection. Define $y_n = x_{\pi(n)}$ and show taht $y_n \to x$ as well.
7. For each statement below, either prove that the statement is true or give a counter-example.
   - If $\lim_n a_n + b_n$ exists and $\lim_n a_n$ exists then $\lim_n b_n$ also exists.
   - If $\lim_n a_n + b_n$ exists then $\lim_n a_n$ and $\lim_n b_n$ exist.

- If $\lim_n a_n b_n$ and $\lim_n a_n$ both exist then $\lim_n b_n$ exists.
- If $\lim_n a_n/b_n$ exists and $\lim_n a_n$ both exist then $\lim_n b_n$ exists.

8. Let $x_1 = a > 1$ and $x_{n+1} = x_n^2/(2x_n - a)$. Show that (x_n) is a decreasing sequence which is bounded below by a. What does (x_n) converge to?

9. Use the definition to prove that $1/2^n \to 0$. To do this, first show that $n \leq 2^n$ for all n by induction.

10. Suppose that $a > 1$. Use the definition (and the results of the previous problem and problem 28 from Chapter 2) to show that $1/a^n \to 0$. To do this, find and $m \in I\!N$ so that $0 < 1 + 1/m \leq a$ so $2 < (1 + 1/m)^m \leq a^m$.

11. Let x_n be the sequence
$$x_n = (-1)^n(1 + \frac{1}{n}).$$
Suppose that $x \in I\!R$ is a number so that for all $\epsilon > 0$ there are infinitely many x_n's so that $|x_n - x| < \epsilon$. Prove that x must be ± 1.

12. Suppose that $x_n \to x$ and I is a closed interval with $x_n \in I$ for all n. Prove that $x \in I$ as well.

13. Suppose that $a_n \to L$ and $b_n \to L$. Define x_n by $x_{2n+1} = a_{n+1}$ and $x_{2n} = b_n$ (so that $x_1 = a_1$, $x_2 = b_2$, $x_3 = a_2$, etc). Prove that $x_n \to L$.

14. Suppose that $x_n \geq 0$ and $x_n \to x > 0$. Prove that $\sqrt{x_n} \to \sqrt{x}$.

15. Suppose that (x_n) is a bounded sequence. Prove that (x_n^m) is bounded for all $m \in I\!N$. Suppose that (x_n) and (y_n) are bounded. Prove that $(x_n + y_n)$ is bounded.

16. Suppose that $x_n \to x$. Prove that $x_n^m \to x^m$ for all $m \in I\!N$.

17. Suppose that $\{a_n\}$ is a sequence and define another sequence
$$s_n = \frac{a_1 + a_2 + \cdots + a_n}{n} = 1/n \sum_{i=1}^{n} a_i.$$
Prove that if $a_n \to A$ then $s_n \to A$ as well. Give an example of a sequence such that s_n converges but a_n does not converge.

18. Let
$$D = \{\frac{n}{2^m} \ : \ m \in I\!N, 0 \leq n \leq 2^m\} = \{0, \frac{1}{2}, 1, \frac{1}{4}, \frac{3}{4}, \frac{1}{8}, \frac{3}{8}, \frac{5}{8}, \frac{7}{8}, \ldots\}.$$
Prove that for any $x \in [0, 1]$ there is a sequence (x_n) with $x_n \in D$ and $x_n \to x$.

19. Let $x_0 = 1$ and $x_{n+1} = 1 + x_n^2/4$. Prove that (x_n) is an increasing sequence that is bounded above by 2. Find the limit of x_n.

20. Suppose that the sequence $1 < a_n \nearrow 2$. Let $x_1 = 1$ and $x_{n+1} = \sqrt{a_n x_n}$. Prove that (x_n) is an increasing sequence which is bounded above by 2. Find the limit of x_n.

21. Suppose that $1 < a < 2$ and define $f(x) = ax(1-x)$. Notice that for $x = (a-1)/a$, we have $f(x) = x$. Prove that for $0 < x < (a-1)/a$ we have $x < f(x)$. Finally, prove that if $0 < x < (a-1)/a$ then $0 < f(x) < (a-1)/a$. Let $x_0 \in (0, (a-1)/a)$ and define $x_{n+1} = f(x_n)$. Prove that (x_n) is an increasing sequence whose limit is $(a-1)/a$.

22. Let $x_1 = 1$ and $x_{n+1} = \sqrt{x_n + a}$ for $n > 1$. Show that (x_n) is an increasing sequence bounded above by $1 + \sqrt{1+a}$. To do this, use for induction hypothesis that $1 \le x_{n-1} \le x_n \le 1 + \sqrt{1+a}$ and prove $x_n \le x_{n+1} \le 1 + \sqrt{1+a}$. Also, it might be useful to notice that

$$x_n^2 - 2x_n - a = (x_n - 1 - \sqrt{1+a})(x_n - 1 + \sqrt{1+a}).$$

23. Let $x_1 = 2$ and $x_{n+1} = (2/3)\ x_n + 1/x_n^2$. Show that this sequence is decreasing and bounded below by $\sqrt[3]{3}$. To do this, use the fact that

$$(2/3)\ x + 1/x^2 - \sqrt[3]{3} = \frac{(2x + 3^{1/3})(x - 3^{1/3})^2}{3x^2}$$

to show that $(2/3)x + 1/x^2 \ge \sqrt[3]{3}$ for any $x > 1$ (this will help you show that $x_n \ge \sqrt[3]{3}$ for all n). Why does the limit of (x_n) exist and what is the limit?

24. Let $x_1 = 2/3$ and

$$x_{2n+1} = \frac{x_{2n}}{3} + \frac{2}{3} \qquad x_{2n} = \frac{x_{2n-1}}{3}$$

(thus, $x_2 = 2/9$ and $x_3 = 20/27$ and $x_4 = 20/81$, etc). Prove that $0 < x_n < 1$ and (x_n) does not converge. Notice that

$$x_{2n+2} = \frac{x_{2n}}{9} + \frac{2}{9} \qquad x_{2n+3} = \frac{x_{2n+1}}{9} + \frac{2}{3}.$$

Use this to show that

$$|x_{2n+2} - 1/4| = \frac{4}{36}|x_{2n} - 1/4| \quad \text{and} \quad |x_{2n+3} - 3/4| = \frac{4}{36}|x_{2n+1} - 3/4|$$

and thus $x_{2n} \to 1/4$ while $x_{2n+1} \to 3/4$.

25. Let $A > 0$ be fixed and $x_1 = 2A$ and

$$x_{n+1} = (\frac{2}{3})x_n + \frac{A}{3x_n^2}.$$

Show that (x_n) is a decreasing sequence by showing that $\sqrt[3]{A} < x_{n+1} < x_n$ for all n. To show that $x_{n+1} > \sqrt[3]{A}$, it might be helpful to know that

$$2x^3 - (3\sqrt[3]{A})x^2 + A = (2x + A^{1/3})(x - A^{1/3})^2.$$

26. Suppose that $x_n > 0$ with

$$\sup_{n \geq N_0} \frac{x_{n+1}}{x_n} < 1.$$

Prove that $x_n \to 0$. (**Hint:** show that there is an $r < 1$ and a constant C with $x_n < Cr^n$).

27. Suppose $f : \mathbb{R} \to \mathbb{R}$ satisfies $f(0) = 0$ and $|f(x)| \leq |x|/2$ for all $x \in \mathbb{R}$. Define the sequence (x_n) by choosing some $x_1 \in \mathbb{R}$ and letting $x_{n+1} = f(x_n)$. Prove that $x_n \to 0$.

28. Let $p \in \mathbb{N}$. Prove that $(1 + 1/n)^p \to 1$ as $n \to \infty$. Now, let $0 < r < 1$ and show that the sequence $x_n = r^n n^p$ is eventually decreasing (that is, there is some N so that if $n \geq N$ then $x_{n+1} \leq x_n$).

29. Let $x_1 > 1$ and $x_{n+1} = 2 - 1/x_n$. Prove that (x_n) converges. (**Hint:** Is (x_n) monotone?).

30. Let $x_1 = 1$ and $x_{n+1} = 1 - x_n^2/4$. Show that $x_2 \leq x_4 \leq x_6 \leq \cdots x_5 \leq x_3 \leq x_1$. To do this, show that $x_{n-2} \leq x_n \leq x_{n-1}$ if n is even and $x_{n-1} \leq x_n \leq x_{n-2}$ if n is odd. Prove that the subsequence of even terms converges and the subsequence of odd terms converges and find their commmon limit.

31. Suppose that $E \subset \mathbb{R}$ and that $x = \sup E$ exists. Prove that there is a sequence $x_n \in E$ for $n \in \mathbb{N}$ with $x_n \leq x_{n+1}$ and $|x_n - x| < 1/n$.

32. Suppose $x_n \to x$ and let $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots a_k x^k$ be a polynomial. Prove that $p(x_n) \to p(x)$.

33. Suppose that $x_n \to 0$. Let $f : \mathbb{N} \to \mathbb{N}$ be any bijection and define $y_n = x_{f(n)}$ (that is, we use the function f to scramble around the terms of the sequence (x_n) to obtain (y_n)). Prove that $y_n \to 0$ as well.

34. Let $a > 1$ and

$$x_n = \sum_{i=1}^{n} a^{-i}.$$

Prove that (x_n) is Cauchy and hence converges.

35. Let the sequence (x_n) be defined by $x_0 = 1$, $x_1 = 2$ and $x_{n+1} = (x_n + x_{n-1})/2$. Prove, directly from the definition, that (x_n) is a Cauchy sequence.

36. Let $a_n > 1$ for all $n \in \mathbb{N}$ and define the sequence (p_n) by $p_0 = 1$ and $p_{n+1} = a_{n+1} p_n$.

    (a) Prove that (p_n) is an increasing sequence.

    (b) Prove that if (p_n) converges then $a_n \to 1$.

    (c) Prove that if $a_n \geq 1 + 1/n$ then $p_n \to \infty$.

37. Let $a, b > 0$. Let $x_1 = \sqrt{ab}$ and $y_1 = (a+b)/2$ and recursively define $x_{n+1} = \sqrt{x_n y_n}$ and $y_{n+1} = (x_n + y_n)/2$.

    (a) Prove that $x_n < y_n$ for all n.

    (b) Prove that (x_n) is an increasing sequence and (y_n) is a decreasing sequence and both are bounded. Thus, $x_n \nearrow x$ and $y_n \searrow y$.

    (c) Prove that $y_{n+1} - x_{n+1} < (y_1 - x_1)/2^n$ for all $n > 1$.

    (d) Prove $x = y$.

38. Since
$$\frac{3}{2}x - \frac{x^2}{4} = \frac{-1}{4}(x-3)^2 + \frac{9}{4},$$
show that $(3/2)x - x^2/4 \leq 9/4$ for all x. Furthermore, show that if $0 < x < 2$ then $x < (3/2)x - x^2/4$. Let $x_0 = 1$ and $x_{n+1} = -x_n^2/4 + \frac{3}{2}x_n$. Prove that (x_n) is an increasing sequence that is bounded above. What is the limit of x_n?

39. Find an example of a Cauchy sequence (x_n) and a function $f(x)$ so that $(f(x_n))$ is NOT a Cauchy sequence. What if f is a bounded function? Is it still possible to find such an example?

40. Suppose that $a_n \to a \in \mathbb{R}$. Let $x_0 \in \mathbb{R}$ be some point and define the sequence (x_n) by $x_{n+1} = x_n/2 + a_n$. Prove that (x_n) converges.

41. Prove that if $x_n \to \infty$ and $y_n \to -\infty$ then $x_n y_n \to -\infty$.

42. Suppose that $x_n \to x$. Show that if (x_{n_k}) is any subsequence of (x_n) then $x_{n_k} \to x$ as well (as $k \to \infty$).

43. Let (x_n) be a real sequence so that any subsequence converges. Show that (x_n) converges. On the other hand, find an example of a real sequence (y_n) which does not converge but has the property that every subsequence (y_{n_k}) has a further subsequence $(y_{n_{k_l}})$ which converges.

44. Suppose that (x_n) is a real sequence and that there is an $x \in \mathbb{R}$ so that for any subsequence (x_{n_k}) we have $x_{n_k} \to x$. Show that $x_n \to x$ as well.

45. Let (x_n) be a real sequence and define $t_n = \sup\{x_n, x_{n+1}, x_{n+2}, \ldots\}$ and $s_n = \inf\{x_n, x_{n+1}, x_{n+2}, \ldots\}$. Prove that $s_n \leq s_{n+1}$ for all n and $s_n \leq t_m$ for n, m and that $t_{n+1} \leq t_n$ for all n. Suppose that (x_n) is bounded and prove that (s_n) and (t_n) converge with $\lim_n s_n \leq \lim_n t_n$.

46. (This question is related to the previous one). Suppose that $x_n \to x$. Prove that $s_n \to x$ and $t_n \to x$ as well (where s_n and t_n are defined as in the previous question).

47. Let $t = \lim_n t_n$ and $s = \lim_n s_n$ (as in the previous two questions). Suppose that $s = t$. Prove that (x_n) converges to s.

48. Let (s_n) and (t_n) be as in the previous problems and suppose that (x_n) is bounded.

    (a) For each $\epsilon > 0$ define the set $S_\epsilon = \{n \in I\!N : x_n > t - \epsilon\}$. Prove that S_ϵ is infinite for each ϵ. Thus for $m \in S_\epsilon$ we have $t - \epsilon < x_m \leq t_m$.

    (b) Prove that you can choose a sequence of integers $m_1 < m_2 < m_3 < \ldots$ so that $t - 1/k < x_{m_k} \leq t_{m_k}$.

    (c) Prove that $x_{m_k} \to t$ as $k \to \infty$.

49. In this problem we will prove that any bounded sequence (x_n) has either an increasing or a decreasing subsequence. Thus, suppose (x_n) is a bounded sequence.

    (a) If $\{x_n : n \in I\!N\}$ is finite, show that (x_n) has a constant subsequence.

    (b) Suppose that (x_n) has no decreasing subsequence and define $s = \inf\{x_n : n \in I\!N\}$. Show that $s = x_{n_1}$ for some $n_1 \in I\!N$.

    (c) Repeating the previous step inductively, obtain a subsequence (x_{n_k}) with $x_{n_k} \leq x_{n_{k+1}}$ and $n_k < n_{k+1}$.

    Now, put all these steps together to prove that any bounded sequence has either an increasing or decreasing subsequence.

50. In this problem, we define the standard Cantor set and prove some properties about it. Let $C_0 = [0, 1]$. We define C_1 by starting with C_0 and removing the middle 1/3 open interval from C_0, that is $C_1 = C_0 \setminus (1/3, 2/3) = [0, 1/3] \cup [2/3, 1]$. We continue this way, taking C_n (which consists of 2^n closed intervals of length $(1/3)^n$) and removing the middle 1/3 open interval from each o these intervals to obtain C_{n+1}. So, for example, $C_2 = [0, 1/9] \cup [2/9, 3/9] \cup [6/9, 7/9] \cup [8/9, 1]$ and

$$C_3 = [0, \frac{1}{27}] \cup [\frac{2}{27}, \frac{3}{27}] \cup [\frac{6}{27}, \frac{7}{27}] \cup [\frac{8}{27}, \frac{9}{27}] \cup [\frac{18}{27}, \frac{19}{27}] \cup [\frac{20}{27}, \frac{21}{27}] \cup [\frac{24}{27}, \frac{25}{27}] \cup [\frac{26}{27}, 1].$$

    We notice that $C_{n+1} \subset C_n$ and define $C = \bigcap_n C_n$. The set C is called the *middle 1/3 Cantor Set*.

    (a) Prove that C is uncountable.

    (b) Suppose that $x_n \in C$ and $x_n \to x$. Prove that $x \in C$.

Chapter 4

Functions: Convergence and Continuity

Definition 4.1 *Let $a \in \mathbb{R}$ and I be an open interval containing a. Let f be a real function defined on $I \setminus \{a\}$. Then f is said to* converge to L as $x \to a$ *if for every $\epsilon > 0$ there is a $\delta > 0$ so whenever $0 < |x - a| < \delta$ and $x \in I$ we have $|f(x) - L| < \epsilon$.*

We write this as $\lim_{x\to a} f(x) = L$.

Example For $f(x) = x^2$ we have that $\lim_{x\to 1} f(x) = 1$. To see this, let $\epsilon > 0$ be given. Since $|x - 1| < 1$ implies that $0 < x < 2$ which in turn implies that $|x + 1| < 3$, let $\delta = \min\{1, \epsilon/3\}$. Then if $0 < |x - 1| < \delta$ we have that

$$|x^2 - 1| = |x + 1||x - 1| < 3\delta \le \epsilon.$$

Notice the condition $x \in I$ and $0 < |x - a| < \delta$ (so $x \ne a$). Thus the value of f at a (if it exists) doesn't influence the value of the limit.

Theorem 4.1 *Let $a \in I$ with $I \subset \mathbb{R}$ an open interval and f be a real function which is defined on $I \setminus \{a\}$. Then $L = \lim_{x\to a} f(x)$ iff $f(x_n) \to L$ for every sequence $x_n \in I \setminus \{a\}$ with $x_n \to a$.*

Proof: Suppose that $\lim_{x\to a} f(x) = L$ and let $x_n \to a$ with $x_n \in I \setminus \{a\}$. Let $\epsilon > 0$ be given. Then since $\lim_{x\to a} f(x) = L$ there is a $\delta > 0$ so that whenever $0 < |x - a| < \delta$ and $x \in I$ we have $|f(x) - L| < \epsilon$. Now since $x_n \to a$ there is an $N \in \mathbb{N}$ so whenever $n \ge N$ we have $0 < |x_n - a| < \delta$ which implies that $|f(x_n) - L| < \epsilon$. Thus $f(x_n) \to L$ as desired.

Conversely, suppose that for any (x_n) with $x_n \to a$ and $x_n \in I \setminus \{a\}$ we know $f(x_n) \to L$. Suppose that $\lim_{x\to a} f(x) \ne L$. Then there is an $\epsilon > 0$ so that for all $\delta > 0$ there is some $x \in I \setminus \{a\}$ with $0 < |x - a| < \delta$ but $|f(x) - L| \ge \epsilon$. This means that for each $n \in \mathbb{N}$ we have an $x_n \in I \setminus \{a\}$ with

$0 < |x_n - a| < 1/n$ but $|f(x_n) - L| \geq \epsilon$. However, then $x_n \to a$ but $f(x_n) \not\to a$, which is a contradiction. ∎

Example For

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

we see that $f(x)$ has no limit as $x \to 0$ because if we let

$$x_n = \frac{2}{(2n+1)\pi} \quad \Rightarrow \quad f(x_n) = \sin\left(\frac{(2n+1)\pi}{2}\right) = \sin(\pi/2 + n\pi) = (-1)^n$$

which does not converge as $n \to \infty$. However, $x_n \to 0$.

Arithmetic and Limits

Theorem 4.2 *Suppose that $a \in \mathbb{R}$ and $I \subset \mathbb{R}$ is an open interval containing a. Suppose further that f, g are real valued functions that are defined on $I \setminus \{a\}$. If $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist then*

1. $\lim_{x\to a}(f(x) + g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$
2. $\lim_{x\to a}(f(x) \cdot g(x)) = (\lim_{x\to a} f(x)) \cdot (\lim_{x\to a} g(x))$
3. $\lim_{x\to a} \alpha f(x) = \alpha \lim_{x\to a} f(x)$ *for all* $\alpha \in \mathbb{R}$.
4. $\lim_{x\to a}(f(x)/g(x)) = (\lim_{x\to a} f(x)) \,/\, (\lim_{x\to a} g(x))$ *if* $\lim_{x\to a} g(x) \neq 0$.

Proof: We use the corresponding theorem (Thm 3.5) on sequential limits and the fact that sequential limits suffice for limits of functions (Thm 4.1). ∎

As an illustration, we prove the first part of the theorem, that is we prove that $\lim_{x\to a} f(x)+g(x) = \lim_{x\to a} f(x)+\lim_{x\to a} g(x)$. Suppose that $\lim_{x\to a} f(x) = L$ and $\lim_{x\to a} g(x) = M$. Let $x_n \in I \setminus \{a\}$ be such that $x_n \to a$. Then since $L = \lim_{x\to a} f(x)$ we know (by Thm 4.1) that $\lim_n f(x_n) = L$ and similarly that $M = \lim_n g(x_n)$. But then we have that $\lim_n f(x_n)+g(x_n) = L+M$. Since the sequence (x_n) was arbitrary, by Theorem 4.1, we know that $\lim_{x\to a} f(x)+g(x) = L + M$, as desired.

Theorem 4.3 *(Squeeze Theorem for Functions) Suppose that $a \in \mathbb{R}$ and $I \subset \mathbb{R}$ is an open interval with $a \in I$ and f, g, h are real-valued functions defined on $I \setminus \{a\}$.*

1. *If $f(x) \leq g(x) \leq h(x)$ for all $x \in I \setminus \{a\}$ and $L = \lim_{x\to a} f(x) = \lim_{x\to a} h(x)$ then $\lim_{x\to a} g(x) = L$ as well.*

2. *If* $|g(x)| \leq M$ *for all* $x \in I \backslash \{a\}$ *and* $\lim_{x\to a} f(x) = 0$ *then* $\lim_{x\to a} f(x)g(x) = 0$.

Again, we just prove the first part as an illustration.

Proof: (i) Let $x_n \in I \setminus \{a\}$ with $x_n \to a$. Then $f(x_n) \leq g(x_n) \leq h(x_n)$ and $f(x_n) \to L$ and $h(x_n) \to L$. Thus, by the Squeeze Theorem for sequences (Thm 3.3) we know that $\lim_n g(x_n) = L$ as well. Thus $\lim_{x\to a} g(x) = L$. ∎

Theorem 4.4 *(Comparison Theorem for Functions) Suppose that* $a \in I \subset I\!R$ *with* I *an open interval. and* f, g *are real-valued functions defined on* $I \setminus \{a\}$. *If* $\lim_{x\to a} f(x)$ *and* $\lim_{x\to a} g(x)$ *both exist and* $f(x) \leq g(x)$ *for all* $x \in I \setminus \{a\}$, *then* $\lim_{x\to a} f(x) \leq \lim_{x\to a} g(x)$.

Proof: Use Theorem 4.1 ∎

One-Sided Limits

Definition 4.2 *Let* $I = [a, b) \subset I\!R$ *with* $a < b$. *Suppose that* f *is defined on* (a, b). *We say that* L is the limit of f as $x \to a$ *if for all* $\epsilon > 0$ *there is a* $\delta > 0$ *so whenever* $x \in I$ *with* $0 < |x - a| < \delta$ *we have that* $|f(x) - L| < \epsilon$.

We write this as $\lim_{x\to a^+} f(x) = L$.

We note that we can also define $\lim_{x\to a^-} f(x) = L$ in a similar fashion. Furthermore, it is easy to modify the above definition to define $\lim_{x\to a^+} f(x) = \infty$.

As an example, let's show that for $f(x) = 1/x$ we have $\lim_{x\to 0^+} f(x) = \infty$. To see this, let $I = [0, 1]$ and $M > 0$. Then for $\delta = 1/M > 0$ we see that for any $x \in (0, 1]$ with $0 < |x - 0| < \delta$ we have that $f)x) = 1/x > M$. Thus, $\lim_{x\to 0^+} f(x) = \infty$, as desired.

It is also true that the sequential characterization (Theorem 4.1) holds for one-sided and infinite limits as well.

Continuity

Definition 4.3 *Let* $\emptyset \neq E \subset I\!R$ *and* $f : E \to I\!R$.

1. *f is said to be* continuous at the point $x_0 \in E$ *if for every* $\epsilon > 0$ *there is a* $\delta > 0$ *so whenever* $x \in E$ *and* $|x - x_0| < \delta$ *we have that* $|f(x) - f(x_0)| < \epsilon$.
2. *f is said to be* continuous on a set $A \subset E$ *if it is continuous at each point of* A.
3. *f is said to be* continuous *if it is continuous on its domain.*

Basically f is continuous at x_0 if $\lim_{x\to x_0} f(x) = f(x_0)$ (ALMOST). This isn't quite true, since if there isn't an open interval $I \subset E$ with $x_0 \in I$ then $\lim_{x\to x_0} f(x)$ isn't really defined.

Example Let $E = \{1, 1/2, 1/3, 1/4, \ldots\}$ and $f : E \to \mathbb{R}$ be defined by $f(1/n) = n$ is continuous. Why? Let $\epsilon > 0$ be given and choose some $n \in \mathbb{N}$. Let $\delta = \frac{1}{3n(n+1)}$. Then if $x \in E$ with $|x - 1/n| < \delta$ we have $x = 1/n$ so $|f(x) - f(1/n)| = 0 < \epsilon$.

That is, since E is a discrete set (each point $x \in E$ is *isolated* in the sense that there is some $\delta > 0$ so the only point of E in $(x - \delta, x + \delta)$ is x) it is automatically continuous.

There are some standard examples of strangely behaved functions. We give two of these examples. The first example is one which is discontinuous everywhere and the second is a function which is continuous at all irrational numbers but discontinuous at all rational numbers.

Example Let $f : \mathbb{R} \to \mathbb{R}$ be defined by

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

Then f is discontinuous at all points of $\mathbb{R}$. To see this, let $x \in \mathbb{R}$ and $\delta > 0$ be given. Then there are points $y, z \in (x - \delta, x + \delta)$ with y a rational number and z an irrational number which means that $f(y) = 1$ and $f(z) = 0$. Thus either $|f(x) - f(y)| = 1$ or $|f(x) - f(z)| = 1$.

Example Let $g : \mathbb{R} \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 0, & \text{if } x \text{ is an irrational number} \\ 1/q, & \text{if } x = p/q \text{ where } p \text{ and } q \text{ are relatively prime} \end{cases}$$

Then f is discontinuous at all rational numbers (since each rational number has a sequence of irrational numbers which converge to it) but continuous at each irrational number.

Theorem 4.5 *Let $\emptyset \neq E \subset \mathbb{R}$ and $x_0 \in E$ and $f : E \to \mathbb{R}$. Then f is continuous at x_0 iff for every sequence $x_n \to x_0$ with $x_n \in E$ we have $f(x_n) \to f(x_0)$.*

Proof: Suppose that f is continuous at x_0 and $x_n \in E$ converges to x_0. Let $\epsilon > 0$ be given. Then since f is continuous at x_0, there is a $\delta > 0$ so that whenever $|x - x_0| < \delta$ and $x \in E$ we have $|f(x) - f(x_0)| < \epsilon$. However, since $x_n \to x_0$ there is an $N \in \mathbb{N}$ so that for all $n \geq N$ we have $|x_n - x_0| < \delta$ which implies that $|f(x_n) - f(x_0)| < \epsilon$. Thus $f(x_n) \to f(x_0)$, as desired.

Conversely, suppose $f(x_n) \to f(x_0)$ for all sequences $x_n \in E$ with $x_n \to x_0$. Suppose further that f is not continuous at x_0. Then there is some $\epsilon > 0$ so for all $\delta > 0$ there is some $x \in E$ with $|x - x_0| < \delta$ but $|f(x) - f(x_0)| \geq \epsilon$. Thus for each $n \in I\!N$ there is some $x_n \in E$ with $|x_n - x_0| < 1/n$ but $|f(x_n) - f(x_0)| \geq \epsilon$. However, this means that we have a sequence $x_n \to x_0$ with $f(x_n) \not\to f(x_0)$, which is a contradiction. ▮

Combining continuous functions

Theorem 4.6 *Let $\emptyset \neq E \subset I\!R$ and $f, g : E \to I\!R$ and $x_0 \in E$. Suppose that f and g are both continuous at x_0. Then so are $f + g$, $f \cdot g$, $\alpha \cdot f$ (for any $\alpha \in I\!R$) and f/g (as long as $g(x_0) \neq 0$).*

Proof: The theorem follows directly from the properties of limits of sequences (Theorem 3.5) and the previous theorem. ▮

Theorem 4.7 *(Composition) Suppose $A, B \subset I\!R$, f is defined on A, g is defined on B and $f(A) \subset B$.*

1. *If $I \subset A$ is an open interval with $a \in I$ and if $L = \lim_{x\to a} f(x)$ and $L \in B$ and g is continuous at L. Then $\lim_{x\to a} g(f(x)) = g(L)$.*

2. *If f is continuous at $a \in A$ and g is continuous at $f(a)$ then $g \circ f$ is continuous at a.*

Proof: (of 1) Let $\epsilon > 0$. Since g is continuous at L there is a $\gamma > 0$ so that $0 < |y - L| < \gamma$ which implies that $|g(y) - g(L)| < \epsilon$. Since $\lim_{x\to a} f(x) = L$ there is a $\delta > 0$ so that whenever $0 < |x - a| < \delta$ and $x \in I$ we have that $|f(x) - L| < \gamma$ which implies that $|g(f(x)) - g(L)| < \epsilon$. Thus $\lim_{x\to a} g(f(x)) = g(L)$.

Part 2 follows directly from part 1. ▮

The next two theorems are fundamental theorems in the study of continuous functions. The Intermediate Value Theorem satisfies our intuition that a continuous function should have a graph that is unbroken.

Theorem 4.8 *(Extreme Value Theorem) Suppose $f : I \to I\!R$ is continuous where $I \subset I\!R$ is a closed bounded interval. Then f is bounded on I. In fact, if $M = \sup_{x\in I} f(x)$ and $m = \inf_{x\in I} f(x)$ then there are points $x^*, x_* \in I$ with $f(x^*) = M$ and $f(x_*) = m$.*

Proof: Suppose that f is not bounded on $I = [a, b]$. Then there is some sequence $x_n \in I$ with $f(x_n) \geq n$ for all $n \in I\!N$. Since $x_n \in I$, it is a bounded sequence so by the Bolzano-Weierstrass Theorem there is a subsequence $x_{n_k} \to x \in I$. Since f is continuous on I we have $f(x_{n_k}) \to f(x)$, however $f(x_{n_k}) \geq n_k$ so $f(x_{n_k}) \to \infty$, which contradicts the fact that $f(x_{n_k}) \to f(x)$. Thus f must be bounded on I.

This means that $M = \sup_{x \in I} f(x)$ and $m = \inf_{x \in I} f(x)$ both exist.

Suppose that there is no $x_* \in I$ with $f(x_*) = m$. Then $f(x) > m$ for all $x \in I$ so the function $g : I \to \mathbb{R}$ defined by $g(x) = 1/(f(x) - m)$ is well-defined (since $f(x) \neq m$) and continuous on I. But this means that g is bounded on I so $|g(x)| = g(x) \leq C$ for some $C > 0$. Rewriting this we get $f(x) \geq m + 1/c$ which contradicts $m = \inf_{x \in I} f(x)$. Thus there must be some $x_* \in I$ with $f(x_*) = m$.

A similar argument shows that there is an $x^* \in I$ with $f(x^*) = M$. ∎

Insert a nice picture of the Extreme Value Theorem here

Lemma 4.9 *(Sign Preserving Property) Suppose f is continuous at x_0 and $f(x_0) > 0$. Then there are positive numbers $\delta, \epsilon > 0$ with $|x - x_0| < \delta$ implying $f(x) > \epsilon$.*

Proof: Let $\epsilon = f(x_0)/2$. Then since f is continuous at x_0 there is some $\delta > 0$ so if $0 < |x - x_0| < \delta$ we have $|f(x) - f(x_0)| < \epsilon$ which implies that

$$f(x) > f(x_0) - \epsilon = f(x_0) - f(x_0)/2 = f(x_0)/2 = \epsilon$$

as desired. ∎

Theorem 4.10 *(Intermediate Value Theorem) Let $I = [a, b]$ be a closed bounded interval and $f : I \to R$ be continuous. If $f(a) \neq f(b)$ and y_0 is between $f(a)$ and $f(b)$ then there is an $x_0 \in [a, b]$ with $f(x_0) = y_0$.*

Proof: We suppose $f(a) \leq y_0 \leq f(b)$. Define $E = \{x \in [a, b] : f(x) \leq y_0\} \subset [a, b]$ and let $x_0 = \sup E$. Let $x_n \to x_0$ with $x_n \in E$, then since f is continuous on I we know that $f(x_n) \to f(x_0)$ so since $f(x_n) \leq y_0$ we know that $f(x_0) \leq y_0$ as well.

Suppose that $f(x_0) < y_0$. Then letting $\epsilon = (y_0 - f(x_0))/2 > 0$ we see that since f is continuous on I there is a $\delta > 0$ so if $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < \epsilon$ which implies $f(x) < (f(x_0) + y_0)/2 < y_0$ so all $x \in (x_0, x_0 + \delta)$ satisfy $f(x) < y_0$, which contradicts the fact that $x_0 = \sup E$. Thus $f(x_0) = y_0$. ∎

Insert a nice picture of the Intermediate Value Theorem here

Uniform Continuity

Definition 4.4 *Let $E \subset \mathbb{R}$ and $f : E \to \mathbb{R}$. We say that f is* uniformly continuous on E *if for every $\epsilon > 0$ there is a $\delta > 0$ so that of $|x - y| < \delta$ and $x, y \in E$ then $|f(x) - f(y)| < \epsilon$.*

Example The function $f(x) = x^2$ is uniformly continuous on $[0, 2]$. To see this, let $\epsilon > 0$ be given. Then for any $x, y \in [0, 2]$ with $|x - y| < \epsilon/4$ we have that

$$|x^2 - y^2| \leq |x - y||x + y| \leq 4|x - y| < \epsilon.$$

Example The function $f(x) = 1/x$ is not uniformly continuous on the set $(0, 1]$. To see this just notice that

$$|f(\frac{1}{n}) - f(\frac{1}{n-1})| = n - (n - 1) = 1.$$

Suppose that f were uniformly continuous. Then for $\epsilon = 1/2 > 0$ there is some $\delta > 0$ so that anytime we have $x, y \in (0, 1)$ with $|x - y| < \delta$ then $|f(x) - f(y)| < 1/2$. However, since $1/n \to 0$, there is some n so that $|1/n - 1/(n - 1)| < \delta$ but then $|f(1/n) - f(1/(n - 1))| = 1$, a contradiction.

Theorem 4.11 *Let $f : [a, b] \to \mathbb{R}$ be continuous where $[a, b]$ is a closed bounded interval. Then f is also uniformly continuous.*

Proof: Suppose that f is not uniformly continuous. Then there is an $\epsilon > 0$ so that for each $n \in \mathbb{N}$ there are two points $x_n, y_n \in [a, b]$ so that $f(x_n) - f(y_n) > \epsilon$ and $|x_n - y_n| < 1/n$. However, by the Bolzano-Weierstrass Theorem there are subsequences x_{n_k} and y_{n_k} which both converge. Since $|x_{n_k} - y_{n_k}| < 1/n_k$ we know that both of these subsequences converge to the same limit, call it z. Now, f is continuous so that $f(x_{n_k}) \to f(z)$ and $f(y_{n_k}) \to f(z)$. However, $f(x_{n_k}) - f(y_{n_k}) > \epsilon$, which is a contradiction. Thus it must be the case that f is uniformly continuous.$\blacksquare$

Sequentially compact sets

Definition 4.5 *A set $E \subset \mathbb{R}$ is said to be* sequentially compact *if all sequences (x_n) in E have a convergent subsequence which converges to an element of E.*

Usually we will simply call such sets *compact*.

Thus by the Bolzano-Weierstrass Theorem all closed and bounded intervals are sequentially compact. The following two properties of compact sets are easy to prove (see the problems).

Proposition 4.12 *1. Every finite union of compact sets is compact.*

2. The arbitrary intersection of compact sets is compact.

Definition 4.6 *We say that a set $O \subset \mathbb{R}$ is* open *if for every $x \in O$ there is an $\epsilon > 0$ so that $(x-\epsilon, x+\epsilon) \subset O$. We say that $K \subset \mathbb{R}$ is closed if $\mathbb{R} \setminus K$ is open.*

Theorem 4.13 *A subset $K \subset \mathbb{R}$ is closed iff for every convergent sequence (x_n) in K we have that the limit point of the sequence is also in K.*

Proof: Suppose that K is closed and that $x_n \in K$ with $x_n \to K$. Suppose that $x \notin K$. Then since K is closed, $\mathbb{R} \setminus K$ is open, so there is some $\epsilon > 0$ so that $(x-\epsilon, x+\epsilon) \subset \mathbb{R} \setminus K$. Now since $x_n \to x$ there is some $N \in \mathbb{N}$ so that for all $n \geq N$ we have that $|x - x_n| < \epsilon$ which implies that $x_n \in (x-\epsilon, x+\epsilon)$ so that $x_n \notin K$, which is a contradiction. Thus $x \in K$.

Conversely, suppose that for all sequences $x_n \in K$ with $x_n \to x$ we have that $x \in K$ as well. We wish to prove that K is closed. Suppose that K is not closed, then $\mathbb{R} \setminus K$ is not open so there is some $x \notin K$ so that for all $\epsilon > 0$ we have that $(x-\epsilon, x+\epsilon) \cap K \neq \emptyset$. Thus for all $n \in \mathbb{N}$ there is some $x_n \in K$ with $|x_n - x| < 1/n$. But then we have a sequence (x_n) with $x_n \in K$ and $x_n \to x$ but $x \notin K$. ∎

Theorem 4.14 *(Heine-Borel) A subset $K \subset \mathbb{R}$ is compact iff it is closed and bounded.*

Proof: Suppose that K is compact. We wish to prove that K is closed and bounded. First suppose that K is not bounded. Then there is some sequence $x_n \in K$ with $|x_n| \geq n$ for each n. But then the sequence (x_n) cannot have a Cauchy subsequence, so it cannot have a convergent subsequence. Thus K must be bounded.

To show that K is closed, let (x_n) be a convergent sequence in K with limit $x \in \mathbb{R}$. Then since K is compact, there is a subseqence (x_{n_k}) of the sequence (x_n) which converges with $x_{n_k} \to y \in K$. However, since $x_n \to x$ we know that $x_{n_k} \to x$ as well, so $x = y \in K$, so K is closed.

Conversely, suppose that K is closed and bounded. Then there is some $M > 0$ so that $K \subset [-M, M]$. Let (x_n) be a sequence in K. Then $x_n \in [-M, M]$ for all n, so by the Bolzano-Weierstrass Theorem there is a subsequence (x_{n_k}) and a point $x \in [-M, M]$ so that $x_{n_k} \to x$. But since K is closed, we must have that $x \in K$. ∎

Theorem 4.15 *Suppose that $f : K \to \mathbb{R}$ is continuous and K is compact. Then f is uniformly continuous on K.*

Proof: Suppose that f is not uniformly continuous on K. Then there is an ϵ so that for any $\delta > 0$ we have points $x, y \in K$ so that $|x - y| < \delta$ but $|f(x) - f(y)| \geq \epsilon$. Thus for all $n \in I\!N$ there are points $x_n, y_n \in K$ with $|x_n - y_n| < 1/n$ but $|f(x_n) - f(y_n)| \geq \epsilon$.

Now since K is compact there is some subsequence (x_{n_k}) with $x_{n_k} \to x$. However, then there is a subsequence of (y_{n_k}) which converges. Call this subsequence $(y_{n_{k_l}})$ so that $y_{n_{k_l}} \to y \in K$. Thus we have that $x_{n_{k_l}} \to x$ and $y_{n_{k_l}} \to y$. Now by our choice of x_n and y_n we know that

$$|x_{n_{k_l}} - y_{n_{k_l}}| < 1/n_{k_l} \qquad \Rightarrow \qquad x = y.$$

Finally, since f is continuous, we see that $f(x_{n_{k_l}}) \to f(x)$ and $f(y_{n_{k_l}}) \to f(y)$ with $f(x) = f(y)$. However, this contradicts the fact that $|f(x_{n_{k_l}}) - f(y_{n_{k_l}})| \geq \epsilon$. ∎

The proof of the following theorem is virtually identical to the proof of Theorem 4.8.

Theorem 4.16 *Suppose that $f : K \to I\!R$ is continuous and K is compact. Then there are points $x_*, x^* \in K$ so that $m = f(x_*) \leq f(x) \leq f(x^*) = M$ for all $x \in K$.*

Proof: First we show that f is bounded on K. Suppose it is not bounded. Then for every $n \in I\!N$ there is a point $x_n \in K$ so that $|f(x_n)| \geq n$. However, since K is compact there is a subsequence (x_{n_k}) of (x_n) so that $x_{n_k} \to x \in K$. But then since f is continuous we know that $f(x_{n_k}) \to f(x)$, which contradicts the fact that $|f(x_{n_k})| \to \infty$. Thus f is bounded.

Since f is bounded both $m = \inf_{x \in K} f(x)$ and $M = \sup_{x \in K} f(x)$ exist. Suppose that there is no point $x^* \in K$ so that $f(x^*) = M$. Then we have that $f(x) < M$ for all $x \in K$ so the function $g : K \to I\!R$ defined by $g(x) = 1/(M - f(x))$ is well-defined on K and continuous. But then we know that g is bounded on K, so there is a number $C > 0$ so that $g(x) < C$ for all $x \in K$. Rearranging this, we see that this implies that $f(x) < M - 1/C$ for all $x \in K$, which contradicts the fact that $M = \sup_{x \in K} f(x)$. ∎

It is instructive to see an alternate proof of this theorem, so we sketch it now. Since K is compact, then K is bounded. It is possible to show that since f is uniformly continuous on K, then it is bounded on the bounded set K (see the problems below). Thus we know that f is bounded on K, so that $M = \sup_{x \in K} f(x)$ exists. By the approximation property for suprema, there is a sequence $x_n \in K$ so that $f(x_n) \to M$. Since K is compact, there is a subsequence (x_{n_k}) so that $x_{n_k} \to x \in K$ and thus $f(x_{n_k}) \to f(x)$ so $f(x) = M$.

Compact sets are special in many ways. The following theorem illustrates one particularly nice feature of continuous bijections on compact sets.

Theorem 4.17 *Let $f : K \to I\!R$ be continuous where $K \subset I\!R$ is a compact set. Suppose that f is an injection. Then $f^{-1} : f(K) \to K$ is also continuous.*

Proof: It is possible to show that $f(K)$ is also compact (see the problems). To show that f^{-1} is continuous, let (y_n) be a convergent sequence in $f(K)$ with $f(x_n) = y_n \to y$. Then since $f(K)$ is compact, we know that there is some $x \in K$ with $f(x) = y$. We wish to prove that $x_n \to x$. Suppose that $x_n \not\to x$. Then there is an $\epsilon > 0$ and a subsequence (x_{n_k}) of (x_n) for which $|x_{n_k} - x| \geq \epsilon$. Since K is compact, there must be some subsequence $(x_{n_{k_l}})$ of (x_{n_k}) which converges, say $x_{n_{k_l}} \to z \in K$. But then since f is continuous we know that $y_{n_{k_l}} = f(x_{n_{k_l}}) \to f(z)$. However, $y_n \to y$, which means that $f(z) = y$ as well. Since f is a bijection, we have that $z = x$. But this contradicts the fact that $|x_{n_{k_l}} - x| \geq \epsilon$. Thus we must have that $x_n \to x$ so f^{-1} is continuous. ∎

Sequences of functions

In this section we discuss the convergence of sequences of functions. Unlike the case for sequences of numbers, there are several natural ways in which a sequence of functions could converge.

Definition 4.7 *Let $E \subset I\!R$ and $f_n : E \to I\!R$ be a sequence of functions. We say that f_n* converges pointwise *on E if for every $x \in E$ we have that $\lim_{n\to\infty} f_n(x)$ exists.*

Notice that if f_n converges pointwise on E, we can define a function $f : E \to I\!R$ by $f(x) = \lim_{n\to\infty} f_n(x)$. In this case, we say that $f_n \to f$ pointwise on E.

Example Let $A = [0, 1]$ and $f_n(x) = x^n$. Then f_n converges pointwise on A to the function

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

Example Let $A = [-\pi, \pi]$ and $f_n = \sin(nx)$. Then f_n does not converge pointwise on A.

Example Let $A = I\!R$ and $f_n(x) = x^n$. Again, f_n does not converge pointwise on A.

Example Let $A = I\!R$ and $f_n(x) = 1 + x + x^2/2 + x^3/6 + \ldots + x^n/n!$. Then f_n converges pointwise on R to the function $f(x) = e^x$.

We see from the examples that each f_n can be continuous while the limit function f not be continuous. When will the limit of continuous functions be continuous? We provide one condition for this to be the case.

Definition 4.8 *Let $f_n : A \to \mathbb{R}$ be a sequence of functions where $A \subset \mathbb{R}$. We say that $f_n \to f$ (a function on A)* uniformly *if for all $\epsilon > 0$ there is an $N \in \mathbb{N}$ so that for all $n \geq N$ and all $x \in A$ we have $|f_n(x) - f(x)| < \epsilon$.*

How is this different from pointwise convergence? We can write pointwise convergence as: *$f_n \to f$ pointwise if for each $x \in A$ and $\epsilon > 0$ there is an $N \in \mathbb{N}$ so that for all $n \geq N$ we have $|f_n(x) - f(x)| < \epsilon$.*

Thus, for pointwise convergence, the N can depend on both x and ϵ while for uniform convergence the N depends on only ϵ (and NOT on x).

Example For $A = [0, 1]$ and $f_n(x) = x^n$ we have that f_n converges pointwise but not uniformly.

Example Let $A = [0, 1]$ and $f_n(x) = x/n$. Then $f_n \to 0$ uniformly on A. To show this, let $\epsilon > 0$ be given then there is an $N \in \mathbb{N}$ so that $1/N < \epsilon$ which implies that for all $n \geq N$ we have

$$|f_n(x) - 0| = x/n < 1/n < 1/N < \epsilon$$

for all $x \in [0, 1]$. The functions f_n get uniformly close to f on the set A.

The proof of the next proposition is left as an exercise.

Proposition 4.18 *Suppose that $f_n \to f$ uniformly. Then $f_n \to f$ pointwise as well.*

Proposition 4.19 *(Cauchy criterion for uniform convergence) The sequence of functions $f_n : A \to \mathbb{R}$ converges uniformly iff for all $\epsilon > 0$ there is an $N \in \mathbb{N}$ so for any two $n, m \geq N$ we have $|f_n(x) - f_m(x)| < \epsilon$ for all $x \in A$.*

Proof: Suppose that $f_n \to f$ uniformly on A and let $\epsilon > 0$ be given. Then there is an $N \in \mathbb{N}$ so that for all $n \geq N$ we have $|f_n(x) - f(x)| < \epsilon/2$ for all $x \in A$. Thus for all $n, m \geq N$ we have

$$|f_n(x) - f_m(x)| \leq |f_n(x) - f(x)| + |f(x) - f_m(x)| < \epsilon.$$

Conversely, suppose that for all $\epsilon > 0$ there is an $N \in \mathbb{N}$ so for all $n, m \geq N$ we have $|f_n(x) - f_m(x)| < \epsilon$ for all $x \in A$. This means that for each $x \in A$ the numerical sequence $(f_n(x))$ is a Cauchy sequence and thus converges to some real number, call it $f(x)$. Now, given an $\epsilon > 0$ we know that there is an $N \in \mathbb{N}$ so whenever $n, m \geq N$ we have that $|f_n(x) - f_m(x)| < \epsilon/2$ for all $x \in A$. If we take the limit of this inequality as $m \to \infty$ we get $|f_n(x) - f(x)| \leq \epsilon/2$ for all $x \in A$ and thus $f_n \to f$ uniformly on A. ∎

Theorem 4.20 *Suppose $f_n \to f$ pointwise on A. Then $f_n \to f$ uniformly on A iff*

$$\sup_{x \in A} |f_n(x) - f(x)| \to 0 \quad \textit{as} \quad n \to \infty$$

Continuity and uniform limits

Theorem 4.21 *Let $A \subset I\!R$ and $f_n : A \to I\!R$ converge uniformly to $f : A \to I\!R$. Suppose that $z \in A$ and for each n we have $\lim_{x\to z} f_n(x) = L_n$ exists. Then (L_n) converges and $\lim_{x\to z} f(x) = \lim_{n\to\infty} L_n$.*

Proof: Let $\epsilon > 0$ be given. Since (f_n) converges uniformly, there is an $N \in I\!N$ so for all $n, m \geq N$ we have $|f_n(x) - f_m(x)| < \epsilon$. Taking limits as $x \to z$ we have that $|L_n - L_m| < \epsilon$ for all $n, m \geq N$. Thus, (L_n) is a Cauchy sequence so converges, say to L.

Now,

$$|f(x) - L| \leq |f(x) - f_n(x)| + |f_n(x) - L_n| + |L_n - L|.$$

We first choose $N \in I\!N$ so that for any $n, m \geq N$ we have $|f_n(x) - f(x)| < \epsilon/3$ for all $x \in A$ and so that $|L_n - L| < \epsilon/3$. Now fix an $n \geq N$. Then choose $\delta > 0$ so that if $|x - z| < \delta$ we have $|f_n(x) - L_n| < \epsilon/3$. Thus we have $|f(x) - L| < \epsilon$ for $|x - z| < \delta$, so $\lim_{x\to z} f(x) = L = \lim_{n\to\infty} f_n(z)$. ∎

We mention that it is possible to modify the above proof for one-sided limits. Finally, the above theorem has the following nice corollary.

Corollary 4.22 *Let $f_n : A \to I\!R$ converges uniformly to $f : A \to I\!R$. If each f_n is continuous, then so is f.*

Problems

1. Define $f : I\!R \to I\!R$ by

$$f(x) = \begin{cases} x - 1, & \text{if } x < 0 \\ x + 1, & \text{if } x \geq 0 \end{cases}$$

   If (x_n) is a monotone sequence which converges show that $(f(x_n))$ is also a monotone sequence which converges. Find an example of a sequence (y_n) which converges but $(f(y_n))$ does not converge.

2. Suppose that $f : I\!R \to I\!R$ is continuous. Prove that $f(x) = 0$ for all x iff $f(q) = 0$ for all rational q.

3. Let $S = \{x \in [0, 1] : x = i/2^n, \text{ for some } i, j \in Z\!\!Z\}$. Suppose that $f : [0, 1] \to I\!R$ satisfies $f(x) = x$ for all $x \in S$. Show that $f(x) = x$ for all $x \in [0, 1]$.

4. Find an example of a continuous function $f : I\!R \to I\!R$ and a sequence (x_n) and a point x so that $f(x_n) \to f(x)$ but $x_n \not\to x$.

5. Suppose that $f : I \to I\!R$ is continuous where I is a closed bounded interval. Prove that there is a $c \in I\!R$ so that the function $g : I \to I\!R$ defined by $g(x) = 1/(c - f(x))$ is continuous.

6. Suppose that $f : A \subset I\!R \to I\!R$ is continuous and $I \subset A$ is an interval. Prove that $f(I)$ is also an interval.

7. Suppose that $f : [0, 1] \to [0, 1]$ is continuous. Prove that there is a point $x \in [0, 1]$ so that $f(x) = x$.

8. Define $f : I\!R \to I\!R$ by

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

Show that $f(f(x)) = x$ for all x and that f is not continuous. Furthermore, show that there is no interval $[a, b]$ (with $a < b$) so that $f([a, b]) \subset [a, b]$.

9. Suppose that $f : I\!R \to I\!R$ is continuous and satisfies $f(f(x)) = x$. Let $x_0 \in I\!R$ and suppose (wolog) that $x_0 < f(x_0)$. Show that $f([x_0, f(x_0)]) = [x_0, f(x_0)]$.

10. Suppose that $f : I\!R \to I\!R$ is continuous and that $f(x+y) = f(x)+f(y)$ for all $x, y \in I\!R$. Define $a = f(1)$ and $g(x) = ax$. We show that $f(x) = g(x)$ for all $x \in I\!R$.

    (a) Show that $f(n) = g(n)$ for all $n \in Z\!\!\!Z$ (that is, $f(n) = nf(1) = na$).

    (b) First show that $2f(x/2) = f(x)$ and hence $f(x/2) = f(x)/2$. Use this to show $f(\frac{x+y}{2}) = \frac{1}{2}(f(x) + f(y))$.

    (c) Define $S = \{i/2^j : i \in Z\!\!\!Z, j \in I\!N\}$. Show that $f(x) = g(x)$ for all $x \in S$.

11. Suppose that $f : I\!R \to I\!R$ satisfies $f(x+y) = f(x)f(y)$ and that $\lim_{x\to 0} f(x) = L \neq 0$ exists. Prove that $L = 1$.

12. Suppose that $f : I\!R \to I\!R$ is continuous and increasing and that $E \subset I\!R$ is bounded. Prove that $f(\sup E) = \sup f(E)$.

13. Prove that if $A, B \subset I\!R$ are both compact sets then so is $A \cup B$.

14. Suppose that $\{A_\lambda : \lambda \in \Lambda\}$ is a collection of compact subsets of $I\!R$. Prove that $\cap_\lambda A_\lambda$ is compact as well.

15. Suppose that $f : A \to I\!R$ is continuous and $K \subset A$ is compact. Prove that $f(K)$ is also compact.

16. Suppose that $f : A \to I\!R$ is uniformly continuous and A is bounded. Prove that f is bounded as well.

17. Prove that if $f_n \to f$ uniformly on A then $f_n \to f$ pointwise on A.

18. Give an example of a bounded continuous function $f : I\!R \to I\!R$ with $\sup\{f(x) : x \in I\!R\} \notin Range(f)$ and $\inf\{f(x) : x \in I\!R\} \notin Range(f)$.

19. Let $f : [0,1] \to I\!R$ be continuous with the maximum and minimum values of f not occuring at either 0 or 1. Show that f is not injective.

20. Let the function $f : [0,1] \to I\!R$ be defined by

$$f(x) = \begin{cases} 0, & \text{if } x = 0 \\ (-1)^n, & \text{if } \frac{1}{n+1} < x \leq \frac{1}{n}. \end{cases}$$

At which points $a \in [0,1]$ does $\lim_{x \to a} f(x)$ exist? What about the function $g : [0,1] \to I\!R$ given by

$$g(x) = \begin{cases} 0, & \text{if } x = 0 \\ (-1)^n/n, & \text{if } \frac{1}{n+1} < x \leq \frac{1}{n} \end{cases}$$

21. Let

$$f(x) = \lim_{n \to \infty} \frac{x^n}{1 + x^n}$$

when this limit exists.

(a) For what values of x does the limit exist?

(b) For which values of x is the function continuous?

(c) Compute $f(x)$ for all x in the domain of f.

22. In this problem we define a function $g : I\!R \to I\!R$ which is continuous but is not monotone on any interval (and so doesn't have a derivative anywhere). We break the proof up into several small steps.

(a) Using the fact that $\sum_{i=k}^{n} (2/3)^i = 3((2/3)^k - (2/3)^{n+1})$ show that $a_n = \sum_{i=0}^{n} (2/3)^i$ is a bounded increasing sequence and that

$$(2/3)^k < \sum_{i=k+1}^{\infty} (2/3)^i.$$

(b) For any $x \in I\!R$ let $\mathtt{frac}(x) = x - n$ where $n \in I\!N$ and $n \leq x < n+1$ (so that $\mathtt{frac}(x)$ is like the fractional part of x). Then define $f : I\!R \to I\!R$ by

$$f(x) = \begin{cases} \mathtt{frac}(x), & \text{if } 0 \leq \mathtt{frac}(x) \leq 1/2 \\ 1 - \mathtt{frac}(x), & \text{if } 1/2 \leq \mathtt{frac}(x) < 1 \end{cases}.$$

Show that f is continuous and that $f(2^n x)$ is continuous for each $n \in I\!N$.

(c) For each $n \in I\!N$ define $g_n : I\!R \to I\!R$ by

$$g_n(x) = \sum_{i=0}^{n} (2/3)^i f(2^i x).$$

Show that g_n is continuous for all $n \in I\!N$.

(d) Prove that for any fixed $x \in \mathbb{R}$, $0 \leq g_n(x) \leq 3$ and the sequence $(g_n(x))$ is increasing, so therefore must converge.

Since $\lim_n g_n(x)$ exists for each x, it makes sense to define $g(x) = \lim_n g_n(x)$.

(e) Show that g is a continuous function.

(f) Show that for $i \in \mathbb{Z}$ and $j \in \mathbb{N}$, $g(x)$ is not monotone on the interval $[i/2^j, (i+1)/2^j]$.

23. Define the sequence of functions $f_n : [0,1] \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} n^2 x, & \text{if } 0 \leq x \leq 1/n \\ -n^2(x - 2/n), & \text{if } 1/n < x \leq 2/n \\ 0, & \text{if } 2/n < x \leq 1 \end{cases}$$

Does f_n converge pointwise? If so, find the function it converges to pointwise and prove it. Does f_n converge uniformly? If so, find the function it converges uniformly to and prove the convergence. If it converges both pointwise and uniformly, are these functions the same?

24. Suppose that $f_n \to f$ pointwise and $f_n \to g$ uniformly on $[a, b]$. Prove that $f = g$.

25. Let $f_n(x) = x + 1/n$ for all $n \in \mathbb{N}$ and $f(x) = x$.

(a) Show that $f_n \to f$ uniformly on $\mathbb{R}$.

(b) Show that $(f_n)^2 \to (f)^2$ pointwise on $\mathbb{R}$ but not uniformly.

26. Suppose that $\{f_n\}$ is a sequence of bounded functions on $E = [0, 1]$.

(a) If $f_n \to f$ pointwise on E, is f necessarily bounded? Either prove or give a counterexample.

(b) If $f_n \to f$ uniformly on E, is f necessarily bounded? Either prove or give a counterexample.

Chapter 5

Derivatives

Definition 5.1 *Let f be a real-valued function of a real variable.*

1. *f is said to be* differentiable at the point x_0 *if f is defined on an open interval containing x_0 and*

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

exists.

2. *f is said to be* differentiable *if it is differentiable at each point of its domain.*

The limit (above) is called the *derivative of f at the point x_0.*

Theorem 5.1 *If f is differentiable at x_0 then f is continuous at x_0.*

Proof: For all $x \neq x_0$ and such that $f(x)$ is defined, we have

$$f(x) = \left(\frac{f(x) - f(x_0)}{x - x_0} \right) (x - x_0) + f(x_0).$$

As $x \to x_0$ the right hand side converges to $f(x_0)$ by the assumption that $f'(x_0)$ exists. Thus f is continuous at x_0. ∎

The example of $f(x) = |x|$ shows that the converse is not true since for $x < 0$ we have that

$$\frac{f(x) - f(0)}{x - 0} = -1$$

while for $x > 0$ we have that

$$\frac{f(x) - f(0)}{x - 0} = 1$$

so the limit cannot exist.

Derivatives and Arithmetic

This next theorem establishes all the simple rules of the calculus of derivatives. Using these rules allows one to reduce the computation of the derivative of a complicated function to computing the derivatives of simpler functions.

Theorem 5.2 *If f and g are differentiable at x_0 then so are $f+g$ and fg and αf (for all real α) and f/g as long as $g(x_0) \neq 0$. Furthermore, we have that*

- $(f+g)'(x_0) = f'(x_0) + g'(x_0)$
- $(fg)'(x_0) = f'(x_0)g(x_0) + f(x_0)g'(x_0)$
- $(\alpha f)'(x_0) = \alpha f'(x_0)$
- $(f/g)'(x_0) = \dfrac{f'(x_0)g(x_0) - f(x_0)g'(x_0)}{(g(x_0))^2)}$

Proof: The proofs of all these are basically all the same, so we only prove the last.

Let $x_n \to x_0$ be such that both $f(x_n)$ and $g(x_n)$ are defined for all n. Then we see that

$$\begin{aligned}
\frac{\frac{f(x_n)}{g(x_n)} - \frac{f(x_0)}{g(x_0)}}{x_n - x_0} &= \frac{1}{g(x_n)g(x_0)} \left(\frac{f(x_n)g(x_0) - f(x_0)g(x_0) + f(x_0)g(x_0) - f(x_0)g(x_n)}{x_n - x_0} \right) \\
&= \frac{1}{g(x_n)g(x_0)} \left[g(x_0) \left(\frac{f(x_n) - f(x_0)}{x_n - x_0} \right) - f(x_0) \left(\frac{g(x_n) - g(x_0)}{x_n - x_0} \right) \right]
\end{aligned}$$

which converges to

$$\frac{1}{g(x_0)g(x_0)} \left(g(x_0)f'(x_0) - f(x_0)g'(x_0) \right)$$

as $x_n \to x_0$, as desired. ∎

The following lemma gives an alternative characterization of the derivative of a function at a point. This characterization is useful in proving the chain rule, which will be its first application.

Lemma 5.3 *Let (a, b) be an open interval in $\mathbb{R}$, $x_0 \in (a, b)$ and $f : (a, b) \to \mathbb{R}$. Then f is differentiable at x_0 iff there is a function $\phi : (a, b) \to \mathbb{R}$ which is continuous at x_0 and such that for all $x \in (a, b)$ we have*

$$f(x) = \phi(x)(x - x_0) + f(x_0)$$

in which case $f'(x_0) = \phi(x_0)$.

Proof: First suppose that f is differentiable at x_0. Define the function $\phi : (a,b) \to I\!R$ by

$$\phi(x) = \begin{cases} \frac{f(x)-f(x_0)}{x-x_0}, & \text{if } x \neq x_0 \\ f'(x_0), & \text{if } x = x_0 \end{cases}$$

Then $\lim_{x\to x_0} \phi(x) = f'(x_0) = \phi(x_0)$ by the definition of $f'(x_0)$, so that ϕ is continuous at x_0. We see by the definition of ϕ that for all $x \neq x_0$ we have that

$$f(x) = \phi(x)(x - x_0) + f(x_0)$$

while clearly $f(x_0) = \phi(x_0)(x_0 - x_0) + f(x_0)$, so the formula holds for all $x \in (a,b)$.

Conversely, suppose that there is some function $\phi : (a,b) \to I\!R$ so that for all $x \in (a,b)$ we have

$$f(x) = \phi(x)(x - x_0) + f(x_0)$$

and so that ϕ is continuous at x_0. Then we see that for $x \neq x_0$ we have

$$\phi(x) = \frac{f(x) - f(x_0)}{x - x_0}$$

which converges to $\phi(x_0)$ as $x \to x_0$ (since we assumed that ϕ is continuous at x_0). Thus $f'(x_0) = \phi(x_0)$ exists so f is differentiable at x_0. ▮

Theorem 5.4 *(The Chain Rule) Let f and g be real valued functions of a real variable. Suppose that f is differentiable at x_0 and that g is differentiable at $f(x_0)$. Then $g \circ f$ is differentiable at x_0 with*

$$(g \circ f)'(x_0) = g'(f(x_0))f'(x_0).$$

Proof: By our assumption, f is defined on some open interval (a,b) (containing x_0) and g is defined on some open interval (c,d) (containing $f(x_0)$) and there are functions $f^* : (a,b) \to I\!R$ and $g^* : (c,d) \to I\!R$ with f^* continuous at x_0 and g^* continuous at $f(x_0)$ and so that

$$f(x) = f^*(x)(x - x_0) + f(x_0) \qquad \text{and} \qquad g(y) = g^*(y)(y - f(x_0)) + g(f(x_0))$$

for all $x \in (a,b)$ and $y \in (c,d)$. Combining these, we see that

$$g(f(x)) = g^*(f(x))\,(f(x) - f(x_0)) + g(f(x_0)) = [g^*(f(x))f^*(x)]\,(x - x_0) + g(f(x_0))$$

for all $x \in (a,b)$. Let $h = g \circ f$ and let $h^* = g^*(f(x))f^*(x)$. Then we see that h^* is continuous at x_0 and that

$$h(x) = h^*(x)(x - x_0) + h(x_0)$$

so $(g \circ f)'(x_0) = h'(x_0) = h^*(x_0) = g'(f(x_0))f'(x_0)$, as desired. ▮

The Mean Value Theorem

The Mean Value Theorem is perhaps the fundamental theorem in the study of differential calculus.

Theorem 5.5 *(Rolle's Theorem) Suppose f is continuous on the closed bounded interval $[a, b]$ and differentiable on (a, b). If $f(a) = f(b)$, then there is a point $c \in (a, b)$ with $f'(c) = 0$.*

Proof: By the Extreme Value Theorem, there are points x^* and x_* in $[a, b]$ with $m = f(x_*) \leq f(x) \leq f(x^*) = M$ for all $x \in [a, b]$. If $m = M$, then $f'(x) = 0$ for all $x \in (a, b)$ (since then f is a constant on this interval).

Thus suppose that $M > m$. Then without loss of generality we assume that $x_* \in (a, b)$. Since m is the minimum value of f over the interval $[a, b]$, we have that

$$f(x_* + h) - f(x_*) \geq 0$$

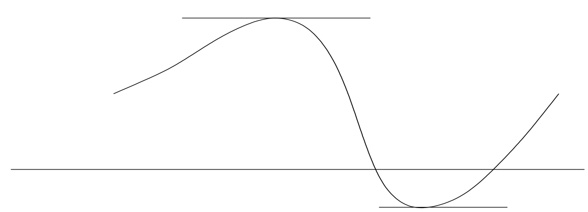
for all h so that $x_* + h \in (a, b)$. If $h > 0$ this means that

$$f'(x_*) = \lim_{h \to 0^+} \frac{f(x_* + h) - f(x_*)}{h} \geq 0$$

while if $h < 0$ then we have

$$f'(x_*) = \lim_{h \to 0^-} \frac{f(x_* + h) - f(x_*)}{h} \leq 0.$$

Thus $f'(x_*) = 0$. ∎



We mention that we need both the conditions that f is continuous on $[a, b]$ and that f is differentiable on (a, b). Here are two examples which show this.

Example The function $f : [0, 1] \to I\!R$ defined by

$$f(x) = \begin{cases} x, & \text{if } x \in [0, 1) \\ 0, & \text{if } x = 1 \end{cases}$$

has $f'(x) = 1$ for all $x \in (0, 1)$ but $f(0) = f(1) = 0$.

Example The function $g : [-1, 1] \to I\!R$ defined by $g(x) = |x|$ is continuous on $[-1, 1]$ and differentiable on $(-1, 0) \cup (0, 1)$ and $g(-1) = g(1)$ but $g'(x) \neq 0$ for all $x \in (-1, 1)$.

Theorem 5.6 *(The Mean Value Theorem) If f is continuous on the closed bounded interval $[a, b]$ and differentiable on (a, b) then*

$$f(b) - f(a) = f'(c)(b - a)$$

for some $c \in (a, b)$

Proof: Let $h(x) = f(x)(b - a) - x(f(b) - f(a))$. Then h is differentiable on (a, b) with $h'(x) = f'(x)(b - a) - (f(b) - f(a))$. Furthermore, h is continuous on $[a, b]$. Now

$$\begin{aligned} h(a) &= f(a)(b - a) - a(f(b) - f(a)) = bf(a) - af(b) \\ h(b) &= f(b)(b - a) - b(f(b) - f(a)) = bf(a) - af(b), \end{aligned}$$

so $h(a) = h(b)$. Thus by Rolle's Theorem there is a $c \in (a, b)$ with

$$0 = h'(c) = f'(c)(b - a) - (f(b) - f(a)) \quad \Rightarrow \quad f(b) - f(a) = f'(c)(b - a),$$

as desired. ∎

Theorem 5.7 *(Generalized Mean Value Theorem) Suppose f and g are continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b). Then there is a $c \in (a, b)$ with*

$$f'(c)(g(b) - g(a)) = g'(c)(f(b) - f(a)).$$

Proof: Use the function $h(x) = f(x)(g(b) - g(a)) - g(x)(f(b) - f(a))$ and repeat the proof of the Mean Value Theorem. ∎

We give some simple applications of the Mean Value Theorem. First, we show that if $f' \geq 0$ then f is increasing.

Proposition 5.8 *Suppose that $f'(x) \geq 0$ for all $x \in (a, b)$. Then f is increasing on (a, b).*

Proof: Let $a < x < y < b$. Then f is continuous on $[x, y]$ and differentiable on (x, y), so by the Mean Value Theorem we have that there is a $c \in (x, y)$ with

$$f(y) = f(x) + f'(c)(y - x) \geq f(x)$$

since $f'(c) \geq 0$. ∎

It is easily to modify this proof to show that if $f' > 0$ then f is strictly increasing. Clearly the corresponding statements concerning decreasing functions are also easy to prove.

We can also prove that the derivative is zero at a local maximum or local minimum.

Proposition 5.9 *Suppose that f is differentiable on (a, b) and continuous on $[a, b]$. Suppose $x \in (a, b)$ has the property that there is some $\delta > 0$ with $f(x) \leq f(z)$ for all $z \in (x - \delta, x + \delta)$. Then $f'(x) = 0$.*

Proof: For $z > x$ we see that

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

while for $z < x$ we see

$$f'(x) = \lim_{z \to x^-} \frac{f(z) - f(x)}{z - x} \leq 0$$

so $f'(x) = 0$. ▮

The Mean Value Theorem also allows one to estimate functions. The following inequality illustrates this technique.

Proposition 5.10 *Let $\alpha > 0$ and $\delta \geq -1$. If $0 < \alpha \leq 1$ then*

$$(1 + \delta)^\alpha \leq 1 + \alpha\delta$$

while if $\alpha \geq 1$ then

$$(1 + \delta)^\alpha \geq 1 + \alpha\delta.$$

Proof: We prove the second case since the details of the first are similar.

Let $f(x) = x^\alpha$. Then $f(1 + \delta) = f(1) + \alpha x_0^{\alpha-1}\delta$ for some x_0 between 1 and $1 + \delta$.

If $\delta > 0$ then $x_0 > 1$ which implies that $x_0^{\alpha-1} > 1$ so that $\delta x_0^{\alpha-1} > \delta$. On the other hand, if $1 < \delta \leq 0$, then $x_0 \leq 1$ so $x_0^{\alpha-1} \leq 1$ or $\delta x_0^{\alpha-1} \geq \delta$.

Thus, in either case we have that $(1 + \delta)^\alpha = f(1 + \delta) = f(1) + \alpha x_0^{\alpha-1} \geq 1 + \alpha\delta$. ▮

Theorem 5.11 *(L'Hopital's Rule) Let x_0 be an extended real number and I an open interval which either contains x_0 or has x_0 as an endpoint. Suppose that f and g are differentiable on I and $g(x) \neq 0 \neq g'(x)$ for all $x \in I$. Suppose further that*

$$A := \lim_{x \to x_0} f(x) = \lim_{x \to x_0} g(x) = 0(\text{ or } \infty).$$

If $B = \lim_{x \to x_0} f'(x)/g'(x)$ exists as an extended real number, then $B = \lim_{x \to x_0} f(x)/g(x)$.

Proof:

We only prove the case where $B \neq \pm\infty$ (that is, B is a real number). The cases $B = \pm\infty$ are similar. Let $\epsilon > 0$ be given.

If $x_0 \neq \pm\infty$ we can choose $\delta > 0$ so that whenever $0 < |z - x_0| < \delta$ then $|f'(z)/g'(z) - B| < \epsilon$. If $x_0 = \pm\infty$ then we will suppose with no loss of generality that $x_0 = \infty$. Then choose $M > 0$ so that if $z > M$ then we have $|f'(z)/g'(z) - B| < \epsilon$.

Now choose $x \in I$ so that either $x > M$ or $|x - x_0| < \delta$. Choose $y \in I$ between x and x_0. Then by the Generalized Mean Value Theorem we get a point x^* between x and y and such that

$$\frac{f(y) - f(x)}{g(y) - g(x)} = \frac{f'(x^*)}{g'(x^*)}.$$

This means (since x^* is also within δ of x_0 or $x^* > M$) that we have

$$B - \epsilon < \frac{f(y) - f(x)}{g(y) - g(x)} < B + \epsilon. \qquad (*)$$

Now we wish to link this to the value of A.

If $A = 0$, then when we take the limits of (*) as $y \to x_0$ we get

$$B - \epsilon \leq \frac{f(x)}{g(x)} \leq B + \epsilon$$

for all x with $|x - x_0| < \delta$ (if $x_0 \in I\!R$) or $x > M$ (if $|x_0| = \infty$).

However, since $\epsilon > 0$ was arbitrary, this means that

$$\frac{f(x)}{g(x)} \to B \quad \text{as} \quad x \to x_0.$$

On the other hand, if $A = \infty$ then we can assume that both $g(y)$ and $g(y) - g(x)$ are positive (that is, either make $\delta > 0$ small enough or $M > 0$ large enough depending on if $|x_0| = \infty$ or $|x_0| < \infty$).

By multiplying (*) by $g(y) - g(x)$ and dividing by $g(y)$ we get

$$\frac{f(x)}{g(y)} + (B - \epsilon)\left(1 - \frac{g(x)}{g(y)}\right) < \frac{f(y)}{g(y)} < \frac{f(x)}{g(y)} + (B + \epsilon)\left(1 - \frac{g(x)}{g(y)}\right)$$

for appropriate y. Now since $A = \infty$, we know that $g(x)/g(y) \to 0$ and $f(x)/g(y) \to 0$ as $y \to x_0$. Thus we can chose $\delta > 0$ small enough (or $M > 0$ large enough) so that

$$B - 2\epsilon < \frac{f(y)}{g(y)} < B + 2\epsilon$$

and thus

$$\frac{f(y)}{g(y)} \to B$$

as $y \to x_0$. ∎

Theorem 5.12 *(Taylor's Theorem) Let $f : [a, b] \to \mathbb{R}$ and $n \in \mathbb{N}$. Assume that $f^{(n-1)}$ is continuous on $[a, b]$ and differentiable on (a, b). Then there is a point $c \in (a, b)$ so that*

$$f(b) = f(a) + \sum_{i=1}^{i=n-1} \frac{f^{(i)}(a)}{i!}(b-a)^i + \frac{f^{(n)}(c)}{n!}(b-a)^n.$$

Proof: For $n = 1$, this reduces to the Mean Value Theorem. Thus suppose that $n > 1$.

Define the number $A \in \mathbb{R}$ to satisfy

$$f(b) = f(a) + \sum_{i=1}^{n-1} \frac{f^{(i)}(a)}{i!}(b-a)^i + \frac{A}{n!}(b-a)^n.$$

We show that there is a point $c \in (a, b)$ so that $f^{(n)}(c) = A$. Towards this end, define the function $g : [a, b] \to \mathbb{R}$ by

$$g(t) = -f(b) + f(t) + \sum_{i=1}^{n-1} \frac{f^{(i)}(t)}{i!}(b-t)^i + \frac{A}{n!}(b-t)^n.$$

Notice that $g(a) = g(b) = 0$ and that g is continuous on $[a, b]$ and differentiable on (a, b). Thus by Rolle's Theorem there is a point $c \in (a, b)$ for which $g'(c) = 0$. Computing we see that

$$g'(t) = f'(t) + \sum_{i=1}^{n-1} \left[\frac{f^{(i+1)}(t)}{i!}(b-t)^i - \frac{f^{(i)}(t)}{(i-1)!}(b-t)^{i-1} \right] - \frac{A}{(n-1)!}(b-t)^{n-1}.$$

Since the sum is a telescoping sum, we get that

$$g'(t) = \frac{(b-t)^{n-1}}{(n-1)!} \left[f^{(n)}(t) - A \right]$$

so that if $g'(c) = 0$ this implies that

$$0 = \frac{(b-c)^{n-1}}{(n-1)!} \left[f^{(n)}(c) - A \right]$$

which implies that $f^{(n)}(c) = A$, as desired (since $b - a \neq 0$). ∎

Monotone Functions

Definition 5.2 *A function $f : I \to \mathbb{R}$ (where I is an interval) is* monotone *on I if it is either increasing or decreasing on I.*

Notation 1 *We denote by $f(x-) = \lim_{z \to x^-} f(z)$, when it exists. Similarly, we denote by $f(x+) = \lim_{z \to x^+} f(z)$, when it exists.*

Theorem 5.13 *If f is monotone on an interval I then f has at most countably many points of discontinuity.*

Proof: Without loss of generality we suppose that f is increasing.

Since $\mathbb{R} = \cup_{n\in\mathbb{Z}}[n, n+1]$, if we can show that f has at most countably many points of discontinuity on an interval of the form $[n, n+1]$ we will know that f has at most countably many points of discontinuity on $\mathbb{R}$. Thus, suppose that $I = [n, n+1]$. Now f is discontinuous at $x \in I$ iff $f(x-) < f(x+)$.

Let $A_i = \{x \in I : f(x+) - f(x-) > 1/i\}$ for each $i \in \mathbb{N}$. Then the set of points of discontinuity of f on I is equal to $\cup_i A_i = E$. Suppose E is uncountable. Then there is some $j \in \mathbb{N}$ so that A_j is uncountable. But then

$$f(n+) - f(n-) \geq \sum_{x\in A_j} f(x+) - f(x-) = \infty$$

since each $f(x+) - f(x-) > 1/j$ and there are infinitely many $x \in A_j$. This contradiction shows (in fact) that each A_j is finite which implies that E is countable. ∎

Theorem 5.14 *If f is one-to-one and continuous on a closed bounded interval $[a, b]$, then f is strictly monotone on this interval. Furthermore, f^{-1} is continuous and strictly monotone on the closed bounded interval with endpoints $f(a)$ and $f(b)$.*

Proof: Without loss of generality we assume that $f(a) < f(b)$. Let $x \in (a, b)$. Suppose $f(x) \leq f(a)$. Then since f is continuous on $[x, b]$, by the Intermediate Value Theorem there is a $y \in (x, b)$ with $f(y) = f(a)$, which contradicts f being injective. Thus, $f(x) > f(a)$. Similarly, we can show that $f(x) < f(b)$ so that $f(a) < f(x) < f(b)$.

Repeating this argument with some $y \in (x, b)$ we show that $f(x) < f(y) < f(b)$, so that f is strictly increasing.

Let $c = f(a)$ and $d = f(b)$. By the Intermediate Value Theorem, we know that $range(f) = [c, d]$. Thus, $f^{-1} : [c, d] \to [a, b]$. Let $y_1 < y_2$, with $y_1, y_2 \in [c, d]$. If $x_1 = f^{-1}(y_1) \geq f^{-1}(y_2) = x_2$, then we would have $y_1 = f(x_1) \geq f(x_2) = y_2$, a contradiction. Thus, $f^{-1}(y_1) < f^{-1}(y_2)$ so f^{-1} is strictly increasing.

Finally, we want to show that f^{-1} is continuous. We show $f(y-) = f(y)$ for each $y \in [c, d]$ leaving the similar case of $f(y+) = f(y)$.

Thus, suppose $y_n \to y \in [c, d]$ where $y_n < y_{n+1} < y$ for all n. Then $f^{-1}(y_n) < f^{-1}(y_{n+1}) < f^{-1}(y)$ so $z = \sup\{f^{-1}(y_n) : n \in \mathbb{N}\}$ exists. Furthermore, $f^{-1}(y_n) \to z$ as $n \to \infty$ so, since f is continuous, we have

$$y_n = f(f^{-1}(y_n)) \to f(z)$$

as $n \to \infty$. However, $y_n \to y$ so $y = f(z)$ or $f^{-1}(y) = z$. Thus, $f^{-1}(y_n) \to f^{-1}(y)$ as desired. ∎

Theorem 5.15 *(Inverse Function Theorem) Let f be injective and continuous on the open interval I. If $f'(x_0)$ exists and is non-zero for some $x_0 \in I$, then f^{-1} is differentiable at $y_0 = f(x_0)$ and*

$$(f^{-1})'(y_0) = \frac{1}{f'(x_0)}.$$

Proof: Choose $a, b \in I$ so that $[a, b] \subset I$ and $x_0 \in (a, b)$. Then by Theorem 5.14, f is strictly monotone on $[a, b]$ (say increasing) and f^{-1} is continuous and strictly monotone on $[f(a), f(b)]$. Let $y_0 = f(x_0)$ and $h \neq 0$ small enough so that $y_0 + h = y \in (f(a), f(b))$. Let $x = f^{-1}(y_0 + h)$, then

$$f(x) - f(x_0) = y_0 + h - y_0 = h.$$

Now since f and f^{-1} are both continuous we know $x \to x_0$ iff $h \to 0$ so

$$\lim_{h \to 0} \frac{f^{-1}(y_0 + h) - f^{-1}(y_0)}{h} = \lim_{x \to x_0} \frac{x - x_0}{f(x) - f(x_0)} = \frac{1}{f'(x_0)}$$

as desired. ∎

This leads us to another version of the Inverse Function Theorem.

Theorem 5.16 *Let I be an open interval and f be differentiable on I with $f'(x) > 0$ for all $x \in I$. Then $f^{-1} : f(I) \to I$ exists.*

Proof: Since $f'(x) > 0$ on I, we see that f is injective on I, so f^{-1} exists. ∎

Uniform limits and derivatives

We go back and discuss the interplay between uniform convergence of functions and derivatives. When is the derivative of a limit equal to the limit of the derivatives? This next theorem gives a partial answer to this question.

Theorem 5.17 *Suppose that (f_n) is a sequence of functions differentiable on $[a, b]$ and such that $(f_n(x_0))$ converges for some point $x_0 \in (a, b)$. If (f_n') converges uniformly on $[a, b]$ then (f_n) converges uniformly on $[a, b]$ to some function f and*

$$\lim_{n \to \infty} f_n'(x) = f'(x)$$

for all $x \in (a, b)$.

Proof: Let $\epsilon > 0$ be given. Then there is an $N \in \mathbb{N}$ so that for all $n, m \geq N$ we have $|f_n(x_0) - f_m(x_0)| < \epsilon/2$ and $|f_n'(t) - f_m'(t)| < \epsilon/(2(b - a))$ for all $t \in [a, b]$. Using the Mean Value Theorem for $f_n - f_m$ we get

$$f_n(x) - f_m(x) \quad = \quad [f_n(x) - f_m(x) - f_n(x_0) + f_m(x_0)] + [f_n(x_0) - f_m(x_0)]$$

$$\begin{aligned}
&= f_n'(t)(x - x_0) - f_m'(t)(x - x_0) + [f_n(x_0) - f_m(x_0)] \\
&= (f_n'(t) - f_m'(t))(x - x_0) + (f_n(x_0) - f_m(x_0))
\end{aligned}$$

where t is some point between x and x_0. Thus we have that

$$|f_n(x) - f_m(x)| < \epsilon$$

for all $x \in [a, b]$ so (f_n) is uniformly Cauchy and thus converges uniformly to some function $f(x)$.

Now fix some point $x \in [a, b]$ and define for $t \neq x$

$$\phi_n(t) = \frac{f_n(t) - f_n(x)}{t - x} \qquad \text{and} \qquad \phi(t) = \frac{f(t) - f(x)}{t - x}$$

Then since f_n is differentiable we know that $\lim_{t \to x} \phi_n(t) = f_n'(x)$ for all n. Furthermore,

$$\phi_n(t) - \phi_m(t) = \frac{f_n(t) - f_n(x) - f_m(t) + f_m(x)}{t - x} = f_n'(s) - f_m'(s)$$

for some s between x and t. However we know that (f_n') converges uniformly on $[a, b]$ which implies that ϕ_n also converges uniformly for $t \neq x$. From the definition of ϕ_n we can see that, in fact, $\phi_n \to \phi$ uniformly for $t \neq x$. However, then we see that

$$\lim_{t \to x} \phi(t) = \lim_{n \to \infty} \lim_{t \to x} \phi_n(t) = \lim_{n \to \infty} f_n'(x)$$

exists so we have that $f'(x) = \lim_{n \to \infty} f_n'(x)$ as desired. ▮

Later we will prove the corresponding result for uniform convergence and integrals. That is if $f_n \to f$ uniformly on $[a, b]$ then

$$\int_a^b f_n(x)\, dx \to \int_a^b f(x)\, dx.$$

Problems

1. Let $f(x) = \frac{1}{x}$. Prove from the definition that $f'(x) = \frac{-1}{x^2}$.

2. Let $f_n(x) = x^n$. Prove using only the definition and the arithmetical rules of derivatives that $f_n'(x) = nx^{n-1}$.

3. Suppose that f is differentiable on the open interval I. Prove that f' is bounded on I iff there is a constant M so that $|f(x) - f(y)| \leq M|x - y|$ for all $x, y \in I$.

4. Define $f : I\!R \to I\!R$ by

$$f(x) = \begin{cases} x^2, & \text{if } x \text{ is rational} \\ 0, & \text{if } x \text{ is irrational} \end{cases}$$

Show that f is only continuous at $x = 0$. Does $f'(0)$ exist? Prove your answer.

5. Suppose that I is an open interval and f is differentiable on I. Prove that if f is increasing on I then $f'(x) \geq 0$ for all $x \in I$.

6. Suppose that f, g are both differentiable on the interval (a, b) and continuous on the closed interval $[a, b]$. Suppose that $f(a) = g(a)$ and $f'(x) \leq g'(x)$ for all $x \in (a, b)$. Prove that $f(x) \leq g(x)$ for all $x \in [a, b]$.

7. Show that $\ln(1+x) < x$ for all $x > 0$. Use the fact that $\frac{d}{dx} \ln(1+x) = \frac{1}{1+x}$.

8. Let $f : I\!R \to I\!R$ be differentiable with $f(0) = 0$ and $|f'(x)| < 1$ for all $x \in I\!R$. Show that $f(x) \neq x$ for all $x \neq 0$.

9. Let $p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$ be a polynomial and $I = [a, b]$ a closed interval. Prove that there is a $K > 0$ so that $|p^{(m)}(x)| \leq K$ for all $x \in [a, b]$ and all m.

10. Let $p(x)$ be a polynomial and $I = [a, b]$ be a closed bounded interval. Prove that there is a $K > 0$ so that

$$|p(x) - p(y)| \leq K|x - y| \qquad \text{for all } x, y \in I.$$

11. Let $p(x)$ be a polynomial with real coefficients. Prove that the roots of $p'(x)$ lie between the roots of $p(x)$, *i.e.*, for any two roots $x_1 < x_2$ of $p(x)$, there is a root y of $p'(x)$ with $x_1 \leq y \leq x_2$. Prove that if x is a multiple root of p then x is also a root of p'.

12. Suppose that $f : I\!R \to I\!R$ satisfies

$$|f(x) - f(y)| \leq (x - y)^2 \quad \text{for all } x, y \in I\!R.$$

Show that $f'(x) = 0$ for all $x \in I\!R$. Show that f is a constant function.

13. Let $g(0) = 0$ and suppose that $g'(0) = 0$. Show that there is a $\delta > 0$ so that $|g(x)| < |x|$ for $|x| < \delta$ (notice that we are not assuming that $g'(x)$ exists for any x other than $x = 0$). Now suppose that f is differentiable and that $f(0) = f'(0) = f''(0)$, show that there is a $\delta > 0$ so that for all $|x| < \delta$ we have $|f(x)| < x^2$.

14. Let

$$f_n(x) = \frac{x}{1 + nx^2}$$

for all $n \in \mathbb{N}$ and $x \in \mathbb{R}$. Show that $\{f_n\}$ converges uniformly to a function f and that

$$f'(x) = \lim_{n\to\infty} f_n'(x)$$

holds for $x \neq 0$ but not for $x = 0$.

15. Let $f : (-1, 1) \to \mathbb{R}$ be differentiable. Prove that for every $\epsilon > 0$ there is a function $g : (-1, 1) \to \mathbb{R}$ so that $|f(x) - g(x)| < \epsilon$ for all $x \in (-1, 1)$ with $g'(x_0)$ not existing for at least one point $x_0 \in (-1, 1)$.

16. Let $f : (-1, 1) \to \mathbb{R}$ be defined by

$$f(x) = \begin{cases} 0, & \text{if } x = 0 \\ x^2 \sin(1/x), & \text{if } x \neq 0 \end{cases}$$

Prove that $f'(x)$ exists for all $x \neq 0$ (you can use the fact that the derivative of $\sin(x)$ is $\cos(x)$) and find a formula for $f'(x)$ for these x. Prove, by using the definition, that $f'(0)$ exists. Prove that f' is NOT continuous at 0.

17. Find an example of a function $f : (-1, 1) \to \mathbb{R}$ so that f is differentiable on $(-1, 0) \cup (0, 1)$ and continuous on $(-1, 1)$ for which there is a sequence (x_n) with $f'(x_n) = 0$, $x_n \to 0$ but $f'(0)$ does not exist.

18. For a function g, we define $g^{\circ 0} = g$ and $g^{\circ(n+1)} = g \circ g^{\circ n}$ (so that $g^{\circ n}$ is the composition of g with itself n times). Let $f(x) = \sin(x)$. Prove that for any fixed x we have that $f^{\circ n}(x) \to 0$. Prove as well that $\frac{d}{dx} f^{\circ n}(x) \to 0$.

19. Suppose that $f : (a, b) \to \mathbb{R}$ is differentiable and $|f'(x)| \leq M$ for all $x \in (a, b)$. Prove that f is uniformly continuous on (a, b).

20. Suppose that $f : (0, 1) \to \mathbb{R}$ is continuously differentiable with $f'(x) \neq 0$ for all $x \in (0, 1)$. Prove that for any open interval $(a, b) \subset (0, 1)$ we have that $f(\ (a, b)\ )$ is an open set.

21. For each n, let $p_n(x) = x^n + x^{n-1} + x^{n-2} + \cdots x^2 + x - 1$. Show that for each n, there is a unique number $\gamma_n > 0$ so that $p_n(\gamma_n) = 0$. Show that $\gamma_{n+1} \leq \gamma_n$ and that $\gamma_n \to 1/2$ as $n \to \infty$.

22. For each n, let $p_n(x) = x^n + x^{n-1} + x^{n-2} + \cdots x^2 + x - \theta$. Show that for each n, there is a unique number $\gamma_n > 0$ so that $p_n(\gamma_n) = 0$. Show that $\gamma_{n+1} \leq \gamma_n$ and that the sequence (γ_n) converges as $n \to \infty$. What is the limit?

23. In this problem we define a function $f : \mathbb{R} \to \mathbb{R}$ which has infinitely many derivatives at $x = 0$ but whose nth degree Taylor Remainder does not converge to zero as $n \to \infty$.

(a) Define

$$f(x) = \begin{cases} 0, & \text{if } x \leq 0 \\ e^{-1/x^2}, & \text{if } x > 0. \end{cases}$$

Show that $f'(x)$ exists for all x and that $f'(0) = 0$.

(b) Let $p(x)$ be a polynomial of degree n. Show that

$$\lim_{x \to 0} p(x) e^{-1/x^2} = 0.$$

Use this to show that $f^{(n)}(0) = 0$.

(c) Show that when you expand around $a = 0$, the Taylor Remainder cannot converge to zero.

Chapter 6

Integration

Our discussion of integration will be a little different than most treatments. In particular, we do not use Riemann Sums. We will define the integral for a smaller class of functions than is standard in the theory of Riemann integration. Our main reason for doing this is to try to make the development more transparent and to try to eliminate some of the technical difficulties. Another reason is that this approach is closer in spirit to the theory of Lebesgue integration, which uses the measurable functions and simple function instead of our class $\mathcal{I}[a,b]$ and step functions.

Preliminaries: Step functions and $\mathcal{I}[a,b]$

To begin, we define the class of step functions. First we must define what we mean by a characteristic function of a set. Given any set A, the *characteristic function of* A is the function

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

Now we will, very formally, define what we mean by a step function. Let $I = [a,b]$ be a closed bounded interval and $\mathcal{P} = \{A_1, A_2, \ldots, A_n\}$ be a collection of subintervals of I so that

1. $A_i \cap A_j = \emptyset$ for $i \neq j$ and
2. $\bigcup_i A_i = I$

(such a collection of subintervals is called a *Partition of* $[a,b]$).

Then a step function based on the partition $\mathcal{P}$ is any function of the form

$$\phi(x) = \sum_i \alpha_i \chi_{A_i} \tag{6.1}$$

where the $\alpha_i's$ are real numbers.

As an example, the function that has the value 1 on the interval $[0, \frac{1}{2})$ and the value -1 on the interval $[\frac{1}{2}, 1]$ is a step function based on the partition $\{0, \frac{1}{2}, 1\}$.

If ϕ is a step function then it is pretty clear what we mean by the integral of ϕ. Suppose that our partition is given by $\mathcal{P} = \{a = x_0 < x_1 < x_2 < \cdots < x_n = b\}$ and we have

$$\phi(x) = \sum_{i=1}^{i=n} \alpha_i \chi_{[x_{i-1}, x_i)}(x).$$

Then the integral of ϕ over the interval $[a, b]$ should be the number

$$\int_a^b \phi(x)\ dx = \sum_{i=1}^{i=n} \alpha_i (x_i - x_{i-1})$$

since the height of ϕ on the interval $[x_{i-1}, x_i]$ is α_i.

Notice that if we write the step function ϕ as

$$\phi(x) = \sum_i \alpha_i \chi_{A_i}$$

then we have

$$\int_a^b \phi(x)\ dx = \sum_i \alpha_i\ length(A_i)$$

where $length(A_i)$ is the length of the subinterval A_i.

Now, if we could somehow approximate a function f by step functions then perhaps we could say that the integral of f is the limit of the integrals of the step functions. This is the idea behind the definition of our class of integrable functions.

Definition 6.1 *Let $[a, b]$ be a closed bounded interval in $I\!R$. We define the class of functions $\mathcal{I}[a, b]$ as the set of all functions $f : [a, b] \to I\!R$ so that there is a sequence $\{\phi_n\}$ of step functions which converge uniformly to f on $[a, b]$.*

We will first prove a couple of properties of functions in $\mathcal{I}[a, b]$.

Theorem 6.1 *If $f \in \mathcal{I}[a, b]$ then f is bounded.*

Proof: Suppose that $\{\phi_n\}$ is a sequence of step functions which converge to f. Let $N \in I\!N$ be such that for all $n, m \geq N$ we have that $|\phi_n(x) - \phi_m(x)| < 1$ for all $x \in [a, b]$. Now since ϕ_N is a step function, there is an $M > 0$ so that $|\phi_N(x)| \leq M$ for all $x \in [a, b]$. However, then this means that for all $n > N$ we have that $|\phi_n(x)| \leq M + 1$ which implies that $|f(x)| \leq M + 1$ as well (since f is the limit of ϕ_n). ∎

Theorem 6.2 *If $f, g \in \mathcal{I}[a, b]$ then so is fg.*

Proof: Suppose that $\phi_n \to f$ uniformly and $\psi_n \to g$ uniformly. Let $\epsilon > 0$. Now since $f, g \in \mathcal{I}[a,b]$ there are M, K so that $|f(x)| \leq M$ and $|g(x)| \leq K$ for all $x \in [a,b]$. Since $\phi_n \to f$ uniformly and $\psi_n \to g$ uniformly, there is an $N \in I\!N$ so that for all $n \geq N$ we have that

$$|f(x) - \phi_n(x)| < \min(\frac{\epsilon}{2K}, 1)$$

and

$$|g(x) - \psi_n(x)| < \frac{\epsilon}{2(M+1)}$$

for all $x \in [a,b]$. Notice that this means that $|\phi_n(x)| < M+1$ for all $x \in [a,b]$. Thus we have

$$\begin{aligned}
|\phi_n(x)\psi_n(x) - f(x)g(x)| &\leq |\phi_n(x)\psi_n(x) - \phi_n(x)g(x)| + |\phi_n(x)g(x) - f(x)g(x)| \\
&= |\phi_n(x)||\psi_n(x) - g(x)| + |g(x)||\phi_n(x) - f(x)| \\
&\leq (M+1)\frac{\epsilon}{2(M+1)} + K\frac{\epsilon}{2K} \\
&= \epsilon
\end{aligned}$$

so that $\phi_n\psi_n \to fg$ uniformly. ▮

Theorem 6.3 *Suppose $f_n \in \mathcal{I}[a,b]$ converges uniformly to f. Then $f \in \mathcal{I}[a,b]$ as well.*

Proof: Since $f_n \in \mathcal{I}[a,b]$, there is a sequence $\{\phi_m^n\}$ of step functions so that $\phi_m^n \to f_n$ uniformly. For each $n \in I\!N$ there is an $N_n \in I\!N$ so that whenever $k \geq N_n$ we have $|\phi_k^n(x) - f_n(x)| < 1/n$ for all $x \in [a,b]$. We claim that $\phi_{N_n}^n \to f$ uniformly. To prove this, let $\epsilon > 0$ be given. Then there is an $N \in I\!N$ so that for $k \geq N$ we have

$$|\phi_{N_k}^k - f_k(x)| < \epsilon/2 \quad \text{and} \quad |f_k(x) - f(x)| < \epsilon/2 \quad \text{for all } x \in [a,b].$$

Then we have that

$$|\phi_{N_k}^k - f(x)| \leq |\phi_{N_k}^k - f_k(x)| + |f_k(x) - f(x)| < \epsilon$$

for all $x \in [a,b]$, so the Theorem is proved. ▮

Continuous Functions are in $\mathcal{I}[a,b]$

We will want to prove that all continuous functions on $[a,b]$ are also in $\mathcal{I}[a,b]$. In order to do this, we first define a stronger form of continuity called *uniform continuity.*

The following definition and theorem are in Chapter 4. We repeat them here as a reminder to the reader and to make this chapter a little more independent of the previous chapters.

Definition 6.2 *Let $f : A \to \mathbb{R}$ where $A \subset \mathbb{R}$. Then we say that f is* uniformly continuous *if for all $\epsilon > 0$ there is a $\delta > 0$ so that whenever $|x - y| < \delta$ and $x, y \in [a, b]$ we have that $|f(x) - f(y)| < \epsilon$.*

If a function is uniformly continuous then it is also continuous. The converse is not true as can be seen from the example $f : (0, 1] \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$. This function is not uniformly continuous since

$$f(\frac{1}{n+1}) - f(\frac{1}{n}) = n + 1 - n = 1$$

(why does this show that f is not uniformly continuous?)

Theorem 6.4 *Let $f : [a, b] \to \mathbb{R}$ be continuous where $[a, b]$ is a closed bounded interval. Then f is also uniformly continuous.*

Proof: Suppose that f is not uniformly continuous. Then there is an $\epsilon > 0$ so that for each $n \in \mathbb{N}$ there are two points $x_n, y_n \in [a, b]$ so that $f(x_n) - f(y_n) > \epsilon$ and $|x_n - y_n| < 1/n$. However, by the Bolzano-Weierstrass Theorem there are subsequences x_{n_k} and y_{n_k} which both converge. Since $|x_{n_k} - y_{n_k}| < 1/n_k$ we know that both of these subsequences converge to the same limit, call it z. Now, f is continuous so that $f(x_{n_k}) \to f(z)$ and $f(y_{n_k}) \to f(z)$. However, $f(x_{n_k}) - f(y_{n_k}) > \epsilon$, which is a contradiction. Thus it must be the case that f is uniformly continuous.$\blacksquare$

Now we are ready to prove that all continuous functions are also in the class of functions $\mathcal{I}[a, b]$.

Theorem 6.5 *Suppose that $f : [a, b] \to \mathbb{R}$ is continuous. Then $f \in \mathcal{I}[a, b]$.*

Proof: Fix $n \in \mathbb{N}$ and let $\delta > 0$ be such that whenever $x, y \in [a, b]$ and $|x - y| < \delta$ we have that $|f(x) - f(y)| < 1/n$ (we can do this since f is uniformly continuous). Now, let $\mathcal{P}_n = \{a = x_0 < x_1 < x_2 < \cdots < x_m = b\}$ be a partition of $[a, b]$ with the maximum size of a subinterval smaller than δ. Define the step function ϕ_n by

$$\phi_n(x) = \sum_{i=1}^{m-1} f(x_{i-1})\chi_{[x_{i-1}, x_i)}(x) + f(x_{m-1})\chi_{[x_{m-1}, x_m]}(x).$$

Now, for all $z \in [x_{i-1}, x_i)$ we know that $|f(z) - f(x_{i-1})| < 1/n$ (since $|z - x_{i-1}| < \delta$). Thus, $|f(x) - \phi_n(x)| < 1/n$ for all $x \in [a, b]$.

However, this means that we can construct a sequence of step functions $\phi_n(x)$ which converge uniformly to f and thus that $f \in \mathcal{I}[a, b]$. $\blacksquare$

Definition of the Integral

Now for $f \in \mathcal{I}[a, b]$ and $\{\phi_n\}$ a sequence of step functions which converge to f, we want to define

$$\int_a^b f(x)\ dx = \lim_n \int_a^b \phi(x)\ dx$$

(assuming that this limit exists). In order to do this, we need to first show that this limit does exist for any given choice of ϕ_n and that this limit does not depend on the choice of the sequence $\{\phi_n\}$.

Theorem 6.6 *Suppose that $\{\phi_n\}$ and $\{\psi_n\}$ are sequences of step functions which both converge to the function f. Then*

$$\lim_n \int |\phi_n(x) - \psi_n(x)|\ dx = 0.$$

Proof: Since $\phi_n \to f$ and $\psi_n \to f$ uniformly, for all $\epsilon > 0$ there is an $N \in I\!N$ so that whenever $n \geq N$ we have that $|\phi_n(x) - \psi_n(x)| \leq |\phi_n(x) - f(x)| + |\psi_n(x) - f(x)| < \epsilon$. Fix some $n \geq N$. Since both ϕ_n and ψ_n are step functions, they are both based on partitions of $[a, b]$, say ϕ_n is based on $\mathcal{P}_1$ and ψ_n is based on $\mathcal{P}_2$. Then if we "combine" these two partitions (by taking the union) we will get some common partition $\mathcal{P}$ that BOTH are based on. Say that $\mathcal{P} = \{a = x_0 < x_1 < \cdots < x_m = b\}$. Then we have that

$$\begin{aligned} \int_a^b |\phi_n(x) - \psi_n(x)|\ dx &= \sum_i |\phi_n(x_i) - \psi_n(x_i)|(x_i - x_{i-1}) \\ &< \sum_i \epsilon(x_i - x_{i-1}) = \epsilon(b - a). \end{aligned}$$

Clearly this can be made as small as we wish.∎

Thus we know that if we wish to define the integral of $f \in \mathcal{I}[a, b]$ we can do it using any sequence of step functions which converge to f uniformly. So now we want to prove that the limit does exist.

Theorem 6.7 *Let $f \in \mathcal{I}[a, b]$ and $\{\phi_n\}$ be a sequence of step functions which converge to f uniformly. Then the limit*

$$\lim_n \int_a^b \phi_n(x)\ dx$$

exists.

Proof: We show that the sequence $\{\int \phi_n(x)\ dx\}$ is a Cauchy sequence, which means that it converges. Thus, let $\epsilon > 0$. Since $\phi_n \to f$ uniformly, there is an $N \in I\!N$ so that whenever $n, m > N$ we have that

$$|\phi_n(x) - \phi_m(x)| < \frac{\epsilon}{b - a}.$$

Fix $n, m > N$. Then, as in the proof of the previous theorem, there is a partition $\mathcal{P} = \{a = x_0 < x_1 < \cdots < x_k = b\}$ so that both ϕ_n and ϕ_m are based on $\mathcal{P}$. Then computing we get that

$$\begin{aligned}
\left| \int_a^b \phi_n(x) - \phi_m(x)\ dx \right| &= \left| \sum_i (\phi_n(x_i) - \phi_m(x_i))(x_i - x_{i-1}) \right| \\
&\leq \sum_i |\phi_n(x_i) - \phi_m(x_i)|\,(x_i - x_{i-1}) \\
&< \sum_i \frac{\epsilon}{b-a}(x_i - x_{i-1}) \\
&= \epsilon.
\end{aligned}$$

Thus, the sequence is a Cauchy sequence and so converges.∎

Now we can finally define the integral of a function in $\mathcal{I}[a, b]$.

Definition 6.3 *Let $f \in \mathcal{I}[a, b]$ and let $\{\phi_n\}$ be any sequence of step functions which converges uniformly to f. We define the* integral of f over $[a, b]$ *to be the limit*

$$\int_a^b f(x)\ dx = \lim_n \int_a^b \phi_n(x)\ dx.$$

Let's do an example to show how one would use the definition to compute the integral of a given function.

Example

Let's construct a sequence of step functions which converge to the function $f(x) = 2x^2 - x + 1$ on the interval $[0, 1]$ and use them to compute the value of the integral of f over this interval.

First, we need to construct the sequence of step functions. To do this, let's just consider for each $n \in \mathbb{N}$ the partition

$$\mathcal{P}_n = \{[\frac{i}{n}, \frac{i+1}{n})\ :\ i = 0, 1, \ldots, n-2\} \cup \{[\frac{n-1}{n}, 1]\}.$$

Define the step function $\phi_n(x) : [0, 1] \to \mathbb{R}$ by letting ϕ_n have the value $f(\frac{i}{n})$ on the interval $[\frac{i}{n}, \frac{i+1}{n})$ for $i = 0, 1, \ldots, n-2$ and the value $f(\frac{n-1}{n})$ on the interval $[\frac{n-1}{n}, 1]$.

We claim that $\phi_n \to f$ uniformly on the interval $[0, 1]$.

To see this, notice that $f'(x) = 4x - 1$ so that $|f'(x)| \leq 3$ for all $x \in [0, 1]$. However, then by the Mean Value Theorem we know that for all $a, b \in [0, 1]$ we have that

$$|f(a) - f(b)| = |f'(c)||a - b| \leq 3|a - b|.$$

Let $\epsilon > 0$ be given. Choose $N \in \mathbb{N}$ so that $\frac{1}{N} < \frac{\epsilon}{3}$. Then for any $n \geq N$ we have that if $x, y \in [0, 1]$ with $|x - y| \leq \frac{1}{n}$ then we have that $|f(x) - f(y)| < \epsilon$.

However, for $z \in [\frac{i}{n}, \frac{i+1}{n}]$ we know that $|z - \frac{i}{n}| \leq \frac{1}{n}$ so that $|f(z) = f(\frac{i}{n})| < \epsilon$. However, this means that $|f(x) - \phi_n(x)| < \epsilon$ for all $x \in [0,1]$. Thus, $\phi_n \to f$ uniformly on $[0,1]$.

Ok, the next thing to do is to compute

$$\int_0^1 \phi_n(x)\ dx$$

and take the limit as $n \to \infty$. So, computing we see that

$$\begin{aligned}
\int_0^1 \phi_n(x)\ dx &= \sum_{i=0}^{i=n-1} f(\frac{i}{n})\frac{1}{n} \\
&= \sum_{i=0}^{n-1} \left(\frac{2\,i^2 - in + n^2}{n^2}\right)\frac{1}{n} \\
&= \frac{1}{n^3}\left(2\sum_i i^2 - n\sum_i i + n^2\sum_i 1\right).
\end{aligned}$$

Here we must use the formula

$$\sum_{i=1}^{n-1} i^2 = \frac{1}{3}n^3 - \frac{1}{2}n^2 + \frac{1}{6}n$$

and

$$\sum_{i=1}^{n-1} i = \frac{1}{2}n^2 - \frac{1}{2}n.$$

Using these, the sum simplifies to

$$\frac{1}{n^3}\left(2(\frac{1}{3}n^3 - \frac{1}{2}n^2 + \frac{1}{6}n) - n(\frac{1}{2}n^2 - \frac{1}{2}n) + n^2(n)\right)$$

which simplifies to

$$\frac{7\,n^2 - 3\,n + 2}{6n^2}.$$

Thus, we know that

$$\int_0^1 \phi_n(x)\ dx = \frac{7\,n^2 - 3\,n + 2}{6n^2}.$$

Taking the limit as $n \to \infty$ of this expression we get $7/6$.

So, we have computed that

$$\int_0^1 (2x^2 - x + 1)\ dx = \frac{7}{6}.$$

Clearly we wish to avoid doing all this work each time we want to compute an integral! Of course, this is where properties of the integral and the Fundamental Theorem of Calculus come in.

Elementary properties of the integral

Now we are ready to proceed with proving some of the usual properties of the integral. For instance, first we wish to prove that the integral linear in its arguments.

Theorem 6.8 *Suppose that $f, g \in \mathcal{I}[a,b]$ and $c \in I\!R$. Then we have that*

$$\int_a^b (f+g)(x)\ dx = \int_a^b f(x)\ dx + \int_a^b g(x)\ dx$$

and that

$$\int_a^b cf(x)\ dx = c\int_a^b f(x)\ dx.$$

Proof: To prove this is suffices to prove that these two properties are true for integrals of step functions. Thus, let ϕ, ψ be two step functions on $[a,b]$ so that

$$\phi(x) = \sum_i \alpha_i \chi_{A_i}$$

(where $\{A_1, A_2, \ldots, A_n\}$ is a partition of $[a,b]$) and similarly

$$\psi(x) = \sum_i \beta_i \chi_{B_i}$$

(where $\{B_1, B_2, \ldots, B_m\}$ is a partition of $[a,b]$). Now, as above, there is a partition $\{C_1, C_2, \ldots, C_N\}$ of $[a,b]$ so that

$$\phi(x) = \sum_i \hat{\alpha}_i \chi_{C_i} \qquad \text{and} \qquad \psi(x) = \sum_i \hat{\beta}_i \chi_{C_i}.$$

Then we have that

$$\begin{aligned}
\int_a^b \phi(x) + \psi(x)\ dx &= \sum_i (\hat{\alpha}_i + \hat{\beta}_i)\ length(C_i) \\
&= \sum_i \hat{\alpha}_i\ length(C_i) + \sum_i \hat{\beta}_i\ length(C_i) \\
&= \int_a^b \phi(x)\ dx + \int_a^b \psi(x)\ dx.
\end{aligned}$$

Similarly,

$$\begin{aligned}
\int_a^b c\phi(x)\ dx &= \sum_i (c\alpha_i)\ length(A_i) \\
&= c\sum_i \alpha_i\ length(A_i) \\
&= c\int_a^b \phi(x)\ dx.
\end{aligned}$$

Thus the Theorem is true for all step functions. However, this means that it is also true for all functions in $\mathcal{I}[a,b]$ since such functions are uniform limits of step functions and the integrals of such functions are the limits of the integrals of step functions. ∎

We next prove a result that relates our definition of integrals with the more usual definition in terms of Riemann sums. That is, if we have a sequence of partitions $\{\mathcal{P}_n\}$ that are getting "fine enough" (in the sense that some step functions based on $\mathcal{P}_n$ are converging uniformly to f) then ANY sequence of step functions defined like in a Riemann sum based on $\mathcal{P}_n$ will converge uniformly to f. So, in defining our sequence of step functions which converges uniformly to f, we can use any values of f inside each of the subintervals as the values of the step functions.

Theorem 6.9 *Suppose that $f \in \mathcal{I}[a,b]$ and let $\{\phi_n\}$ be a sequence of step functions which converge uniformly to f. For each n, let $\mathcal{P}_n = \{a = x_0^n < x_1^n < x_2^n < \cdots < x_{k_n}^n = b\}$ be the partition of ϕ_n. For each n and $i = 0, 1, \ldots, k_n - 1$ choose a point $\bar{x_i^n}$ with $x_i^n < \bar{x_i^n} < x_{i+1}^n$. Finally define the step function*

$$\psi_n(x) = \sum_{i=0}^{k_n - 1} f(\bar{x_i^n}) \chi_{[x_i^n, x_{i+1}^n]}(x).$$

Then $\psi_n \to f$ uniformly.

Proof: Let $\epsilon > 0$. Then since $\phi_n \to f$ uniformly, there is an $N \in \mathbb{N}$ so that for all $n \geq N$ we have that $|\phi_n(x) - f(x)| < \frac{\epsilon}{2}$ for all $x \in [a,b]$. However, this means that for all $x \in (x_i^n, x_{i+1}^n)$ we have

$$|f(\bar{x_i^n}) - f(x)| \leq |f(\bar{x_i^n}) - \phi_n(\bar{x_i^n})| + |\phi_n(x) - f(x)| < \epsilon.$$

Thus, $|\psi_n - f| < \epsilon$ for all $x \in [a,b]$ so $\psi_n \to f$ uniformly as claimed. ∎

We now prove some inequalities for integrals.

Theorem 6.10 *Let $f \in \mathcal{I}[a,b]$ so that $f(x) \geq 0$ for all $x \in [a,b]$. Then*

$$\int_a^b f(x)\ dx \geq 0.$$

Proof: Using the previous Theorem, we can find a sequence $\{\psi_n\}$ of step functions where the values of the function ψ_n are defined using the function f. Since $f(x) \geq 0$ for all x we know that $\psi_n(x) \geq 0$ for all x so this means that $\int \psi_n(x)\ dx \geq 0$ as well for all n. Thus, $\int f(x)\ dx \geq 0$. ∎

Corollary 6.11 *(Comparison Theorem) Let $f, g \in \mathcal{I}[a,b]$ so that $f(x) \leq g(x)$ for all $x \in [a,b]$. Then*

$$\int_a^b f(x)\ dx \leq \int_a^b g(x)\ dx.$$

Proof: We just use the preceding theorem with the function $g(x) - f(x) \geq 0$. ∎

Theorem 6.12 *Let $f \in \mathcal{I}[a,b]$. Then*

$$\left| \int_a^b f(x)\ dx \right| \leq \int_a^b |f(x)|\ dx.$$

Proof: It is enough to prove it for all step functions. For

$$\phi(x) = \sum_i \alpha_i \chi_{A_i}(x)$$

we have that

$$\begin{aligned} \left| \int_a^b \phi(x)\ dx \right| &= \left| \sum_i \alpha_i\ length(A_i) \right| \\ &\leq \sum_i |\alpha_i|\ length(A_i) \\ &= \int_a^b |\phi(x)|\ dx \end{aligned}$$

so the Theorem is proved. ∎

Now we will prove a theorem about the behavior of integrals with respect to uniform limits.

Theorem 6.13 *Suppose that $f_n \in \mathcal{I}[a,b]$ converges uniformly to f (necessarily in $\mathcal{I}[a,b]$). Then*

$$\int_a^b f_n(x)\ dx \to \int_a^b f(x)\ dx.$$

Proof: Let $\epsilon > 0$. Then there is an $N \in \mathbb{N}$ so that for all $n \geq N$ we have $|f_n(x) - f(x)| < \frac{\epsilon}{b-a}$ for all $x \in [a,b]$. However, this implies that

$$f(x) - \frac{\epsilon}{b-a} < f_n(x) < f(x) + \frac{\epsilon}{b-a}$$

for all $x \in [a,b]$. Then by the Comparison Theorem for integrals, we have

$$\int_a^b f(x)\ dx - \epsilon < \int_a^b f_n(x)\ dx < \int_a^b f(x)\ dx + \epsilon$$

so that

$$\left| \int_a^b f(x)\ dx - \int_a^b f_n(x)\ dx \right| < \epsilon$$

as was desired. ∎

The corresponding result for pointwise limits is not necessarily true, as the following example illustrates.

Example Define $f_n : (0, 1] \to I\!R$ by

$$f_n(x) = \begin{cases} 2^n & \text{if } x \le \frac{1}{2^n} \\ 0 & \text{if } \frac{1}{2^n} < x \le 1 \end{cases}$$

Then f_n converges pointwise to the zero function. However,

$$\int_0^1 f_n(x)\ dx = 1$$

for all n and this does not converge to 0.

Fundamental Theorem(s) of Calculus

We now want to prove the Fundamental Theorem(s) of Calculus. These theorems relate the derivative to the integral in that they say that the process of taking an integral is basically the inverse process to that of taking a derivative. Using the FTC, it is easier to calculate integrals.

First, we prove a Mean Value Theorem for integrals.

Theorem 6.14 *(Mean Value Theorem for Integrals) Suppose that $f, g \in \mathcal{I}[a, b]$ with $g(x) \ge 0$ for all $x \in [a, b]$. If*

$$m = \inf_{x \in [a,b]} f(x) \quad \textit{and} \quad M = \sup_{x \in [a,b]} f(x)$$

then there is a $c \in [m, M]$ so that

$$\int_a^b f(x)g(x)\ dx = c \int_a^b g(x)\ dx.$$

In particular, if f is continuous, then there is an $x_0 \in [a, b]$ so that

$$\int_a^b f(x)g(x)\ dx = f(x_0) \int_a^b g(x)\ dx.$$

Proof: Since $g(x) \geq 0$ we have by the Comparison Theorem that

$$m \int_a^b g(x)\ dx \leq \int_a^b f(x)g(x)\ dx \leq M \int_a^b g(x)\ dx.$$

Now, if $\int_a^b g(x)\ dx = 0$, then there is nothing to prove. If it is not, we set

$$c = \frac{\int_a^b f(x)g(x)\ dx}{\int_a^b g(x)\ dx}$$

and we see that $m \leq c \leq M$. If f is continuous, by the Intermediate Value Theorem, there is some $x_0 \in [a, b]$ so that $f(x_0) = c$. ▮

Notice that if $g(x) = 1$ and f is continuous we get that there is an $x_0 \in [a, b]$ so that

$$\int_a^b f(x)\ dx = f(x_0)(b - a)$$

(since $\int_a^b g(x)\ dx = b - a$ in this case).

Ok, now we prove the two parts of the Fundamental Theorem of Calculus.

Theorem 6.15 *(Fundamental Theorem of Calculus, part I) Suppose that $f : [a, b] \to I\!R$ is continuous and we define $F : [a, b] \to I\!R$ by*

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

Then $F'(x) = f(x)$ for all $x \in (a, b)$.

Proof: Notice that for $h \geq 0$ we have that

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

By the MVT for integrals, we know that there is some $x_h \in [x, x + h]$ so that

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

Thus,

$$\frac{F(x + h) - F(x)}{h} = f(x_h).$$

Now $x_h \to x$ as $h \to 0$ so since f is continuous this means that $f(x_h) \to f(x)$ as $h \to 0$ so

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

Similarly we can prove that

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

Thus, $F'(x) = f(x)$ as claimed. ∎

Theorem 6.16 *(Fundamental Theorem of Calculus, part II) Suppose that $f' \in \mathcal{I}[a,b]$. Then*

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

for all $x \in [a,b]$.

Proof: Let $\{\phi_n\}$ be a sequence of step functions which converge uniformly to f'. Suppose that $\mathcal{P}_n = \{a = x_0^n < x_1^n < \cdots < x_{k_n}^n = x\}$ is the partition associated with ϕ_n. Using the Mean Value Theorem we have points $\bar{x_i^n} \in (x_i^n, x_{i+1}^n)$ with

$$f(x_{i+1}) - f(x_i) = f'(\bar{x_i^n})(x_{i+1} - x_i).$$

Now, using Theorem 6.9 we have a sequence of step functions $\{\psi_n\}$ defined by

$$\psi_n(t) = \sum_{i=0}^{k_n - 1} f'(\bar{x_i^n})\chi_{[x_i^n, x_{i+1}^n]}(t)$$

and such that $\psi_n \to f'$ uniformly. However, then we compute that

$$\begin{aligned}
\int_a^x \psi_n(t)\ dt &= \sum_i f'(\bar{x_i^n})(x_{i+1} - x_i) \\
&= \sum_i f(x_{i+1}) - f(x_i) \\
&= f(x) - f(a)
\end{aligned}$$

since the second series is a telescoping series. ∎

Some more integration theorems

Two of the very useful theorems on integrals are the theorem on integration by parts and the change of variables formula. We prove these two next.

Theorem 6.17 *(Integration by parts) Suppose that $f', g' \in \mathcal{I}[a,b]$. Then*

$$\int_a^b f'(x)g(x)\ dx = f(b)g(b) - f(a)g(a) - \int_a^b f(x)g'(x)\ dx.$$

Proof: We know that $f(x)g(x)$ is differentiable and

$$(fg)'(x) = f'(x)g(x) + f(x)g'(x) \in \mathcal{I}[a, b]$$

by the product rule and the fact that sums and products of functions in $\mathcal{I}[a, b]$ are in $\mathcal{I}[a, b]$. Thus, by the Fundamental Theorem of Calculus we have

$$f(b)g(b) - f(a)g(a) = \int_a^b (fg)'(x)\ dx = \int_a^b f'(x)g(x)\ dx + \int_a^b f(x)g'(x)\ dx$$

and we are done with the proof. ∎

Theorem 6.18 *(Change of Variables for Integrals) Let Φ be continuously differentiable on $[a, b]$. Suppose that f is continuous on $\Phi([a, b])$. Then*

$$\int_{\Phi(a)}^{\Phi(b)} f(u)\ du = \int_a^b f(\Phi(x))\ \Phi'(x)\ dx.$$

Proof: Define the functions

$$G(x) = \int_a^x f(\Phi(t))\Phi'(t)\ dt \quad \text{and} \quad F(u) = \int_{\Phi(a)}^u f(t)\ dt.$$

Then by the FTC we know that $G'(x) = f(\Phi(x))\Phi'(x)$ and $F'(u) = f(u)$. Therefore, by the Chain Rule we have that

$$\frac{d}{dx}(G(x) - F(\Phi(x))) = 0$$

for all $x \in [a, b]$. However, then this implies that $G(x) - F(\Phi(x))$ is a constant function (using the MVT) on $[a, b]$. Since $G(a) - F(\Phi(a)) = 0 - 0 = 0$ we see that $G(x) = F(\Phi(x))$ for all $x \in [a, b]$. In particular,

$$\int_a^b f(\Phi(t))\Phi'(t)\ dt = G(b) = F(\Phi(b)) = \int_a^b f(u)\ du$$

as desired. ∎

Improper integrals

Now, suppose that we want to integrate over an infinite interval, such as $[0, \infty)$. What do we do? We do the standard thing of defining an *Improper Integral*.

Definition 6.4 *If $I = [a, \infty)$ and if for all $b > a$ we have that f restricted to $[a, b]$ is in $\mathcal{I}[a, b]$, then we define*

$$\int_a^\infty f(x)\ dx = \lim_{b \to \infty} \int_a^b f(x)\ dx$$

whenever this limit exists.

Obviously, we can define integrals over intervals $(-\infty, a]$ or $(-\infty, \infty)$ in a similar manner.

Finally, suppose that f is NOT the uniform limit of step functions on some interval $[a, b]$ but for all $c \in (a, b)$ we have that f restricted to $[c, b]$ is in $\mathcal{I}[c, b]$, then we could define

$$\int_a^b f(x)\ dx = \lim_{c \to a} \int_c^b f(x)\ dx$$

if the limit exists.

As an example of this last type, the function $f(x) = \sin(1/x)$ is not in $\mathcal{I}[0, 1]$ but when restricted to $[b, 1]$ for $b > 0$ it is in $\mathcal{I}[b, 1]$. Furthermore, it can be shown that

$$\lim_{b \to 0} \int_b^1 \sin(1/x)\ dx = 1/2$$

so we would say that

$$\int_0^1 \sin(1/x)\ dx = 1/2.$$

Limits to our integral

The integral we have defined is certainly a limited integral, in the sense that there are many functions which cannot be integrated using our construction. One can prove that any $f \in \mathcal{I}[a, b]$ has both a left limit and right limit at all points of $[a, b]$.

Theorem 6.19 *Let $f \in \mathcal{I}[a, b]$ and $x \in [a, b]$. Then*

$$\lim_{z \to x^+} f(z) \qquad \textit{and} \qquad \lim_{z \to x^-} f(z)$$

both exist.

Proof: Let $\phi_n \to f$ uniformly on $[a, b]$ where ϕ_n is a step function for each n. Now, for each n, we know that $\lim_{z \to x^+} \phi_n(z) = c_n$ exists since ϕ_n is a step function. We claim that $c_n \to c$ for some c. To show this, we first prove that $\{c_n\}$ is a Cauchy sequence. To this end, let $\epsilon > 0$ be given. Then since $\phi_n \to f$ uniformly there is an $N \in \mathbb{N}$ so that for all $n, m \geq N$ we have that

$$|\phi_n(z) - \phi_m(z)| < \frac{\epsilon}{3}$$

for all $z \in [a, b]$. Now since ϕ_N is a step function there is a $\delta > 0$ so that ϕ_N is constant on the interval $(x, x + \delta)$. But then this constant must be $c_N = \lim_{z \to x^+} \phi_N(z)$. However, then for all $n \geq N$ we have

$$|\phi_n(z) - c_N| < \frac{\epsilon}{3}$$

for all $z \in (x, x+\delta)$. However, then

$$|c_n - C_N| \leq \frac{\epsilon}{3}$$

which implies that for any $n, m \geq N$ we have

$$|c_n - c_m| \leq |c_n - c_N| + |c_N - c_m| \leq \frac{2\epsilon}{3} < \epsilon.$$

Thus, $\{c_n\}$ is a Cauchy sequence so there is a $c \in I\!R$ so $c_n \to c$.

Now we claim that $\lim_{z \to x^+} f(z) = c$. To show this, we again let $\epsilon > 0$ be given. Then since $\phi_n \to f$ uniformly and $c_n \to c$ there is some $N \in I\!N$ so whenever $n \geq N$ we have

$$|c_n - c| < \frac{\epsilon}{2} \quad \text{and} \quad |\phi_n(z) - f(z)| < \frac{\epsilon}{2}$$

for all $z \in [a, b]$. Since ϕ_N is a step function there is a $\delta > 0$ so that $\phi_N(z) = c_N$ for all $z \in (x, x+\delta)$ which implies that

$$|f(z) - c| \leq |f(z) = \phi_N(z)| + |\phi_N(z) - c| \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all $z \in (x, x+\delta)$. Thus, $\lim_{z \to x^+} f(z) = c$ as claimed.

Clearly the case of $\lim_{z \to x^-} f(z)$ is similar. ∎

This excludes many functions that one might wish to integrate, as the next example shows.

Example Let $f : [0, 1] \to I\!R$ be defined by $f(x) = \sin(1/x)$ for $x \neq 0$ and $f(0) = 0$. Then f is continuous on $(0, 1]$ but has no limit at 0. Thus, $f \notin \mathcal{I}[a, b]$.

The more common construction of an integral is that of the Riemann Integral using Riemann sums. There is a very nice theorem of Lebesgue which states that a function f is Riemann integrable if and only if f is continuous except on a set of measure zero. (A set $A \subset I\!R$ is said to *have measure zero* if for all $\epsilon > 0$ there is a countable collection of disjoint intervals I_n such that $A \subset \cup_n I_n$ and $\sum_n length(I_n) < \epsilon$).

One way we could try to extend our integral is to extend the definition of step function to include countably many different values. That is, we could try:

Definition 6.5 *Let $A_1, A_2, \ldots,$ be a sequence of subintervals of $[a, b]$ such that*

1. $\bigcup_i A_i = [a, b]$ *and*
2. $A_i \cap A_j = \emptyset$ *whenever* $i \neq j$.

Then an extended step function based on the partition $\mathcal{P} = \{A_i\}$ is any function of the form

$$\phi(x) = \sum_i \alpha_i \chi_{A_i}(x)$$

where $\{\alpha_i\}$ is any sequence of real numbers such that $\sum_i |\alpha_i| < \infty$.

The integral of $\phi(x)$ (as given above) would be the number

$$\sum_i \alpha_i \; length(A_i)$$

(we know this series converges since $\sum_i length(A_i) = b - a$ and $\sum_i \alpha_i$ converges absolutely).

We could then define the class $\mathcal{J}[a,b]$ to be all those functions that are uniform limits of extended step functions. For $f \in \mathcal{J}[a,b]$ with $\phi_n \to f$ (where ϕ_n is an extended step function for all n) we could then define

$$\int_a^b f(x)\; dx = \lim_n \int_a^b \phi_n(x)\; dx.$$

Clearly $\mathcal{I}[a,b] \subset \mathcal{J}[a,b]$. What functions are in $\mathcal{J}[a,b] \setminus \mathcal{I}[a,b]$?

Riemann-Stieltjes Integrals

There is a slight extension of the Riemann integral that is quite useful, especially in probability theory. Furthermore, the basic idea is rather simple and leads naturally to the more general integration theory based on measures. For these reasons we include an introductory discussion.

As an illustration, suppose that we have the function $f(x)$ on the interval $[-1,1]$ and we wish to compute the integral of f. However, for some reason we think that the interval $[-1,0]$ is twice important as the interval $[0,1]$ so should contribute twice as much to the integral. One way of computing this "integral" would be to compute

$$2\int_{-1}^0 f(x)\; dx + \int_0^1 f(x)\; dx = \int_{-1}^0 f(x)\; d(2x) + \int_0^1 f(x)\; dx.$$

We write $d(2x)$ for the integral over the interval $[-1,0]$ and use this to indicate that we weight this interval by a factor of 2.

This example was very simplistic in that we had a constant weights (at least over distinct intervals). Suppose our weights are not constant? Suppose that we wish to weight things further from 0 less heavily? For example, we might believe that the point x should be weighted by a factor proportional to e^{-x^2}. How can we formalize this?

Definition 6.6 *Let $g : I\!R \to I\!R$ be an increasing function and $I \subset I\!R$ be an interval with endpoints $a < b$. Then the g* length *of I is*

$$g_*(I) = g(b) - g(a).$$

The intuition behind this is that we assign the interval $[a,b]$ to have mass or "importance" $g(b) - g(a)$. Notice that if $g(x) = x$ then $g(b) - g(a) = b - a$, the length of the interval. We warn the reader that the name (g length) and notation ($g_*(I)$) is not standard.

If $\{A_i\}$ is a partition of $[a,b]$ and we have the step function

$$\phi(x) = \sum_i \alpha_i \chi_{A_i}$$

we define the integral of ϕ with respect to g to be

$$\int_a^b \phi(x)\ dg(x) = \sum_i \alpha_i g_*(A_i). \tag{6.2}$$

Notice that if our partition $\{A_i\}$ is equal to $\{a = x_0 < x_1 < \cdots < x_n = b\}$ then we can write this integral as

$$\int_a^b \phi(x)\ dg(x) = \sum_{i_1}^{n} \alpha_i (g(x_i) - g(x_{i-1})).$$

Now if we have a function $f \in \mathcal{I}[a,b]$ and a sequence of step functions (ϕ_n) which converge uniformly to f on $[a,b]$, we want to define

$$\int_a^b f(x)\ dg(x) = \lim_n \int_a^b \phi_n(x)\ dg(x).$$

In order to know that this is well-defined we must know that this limit exists and is independent of the choice of sequence of step functions converging to f.

Proposition 6.20 *Let $g : [a,b] \to I\!\!R$ be increasing and $\phi_n \to f$ uniformly with ϕ_n step functions on $[a,b]$. Then the sequence*

$$\left(\int_a^b \phi_n(x)\ dg(x) \right)$$

is a convergent sequence.

Proof: Let $\epsilon > 0$ be given. We see that (ϕ_n) is uniformly Cauchy, so there is an $N \in I\!\!N$ so that for any $n, m \geq N$ we have $|\phi_n(x) - \phi_m(x)| < \epsilon/(g(b) - g(a))$ for all $x \in [a,b]$. Let $n, m > N$ be fixed and $P = \{A_i\}$ be a partition adapted to both ϕ_n and ϕ_m so that

$$\phi_n(x) = \sum_i a_i^n \chi_{A_i}(x) \qquad \phi_m(x) = \sum_i a_i^m \chi_{A_i}(x).$$

Then

$$\left| \int_a^b \phi_n(x)\ dg(x) - \int_a^b \phi_m(x)\ dg(x) \right| \quad = \quad \left| \sum_i (a_i^n - a_i^m) g_*(A_i) \right|$$

$$
\begin{aligned}
&\leq \sum_i |a_i^n - a_i^m| g_*(A_i) \\
&\leq \sum_i \left(\frac{\epsilon}{g(b) - g(a)} \right) g_*(A_i) \\
&= \frac{\epsilon}{g(b) - g(a)} \sum_i g_*(A_i) \\
&= \frac{\epsilon}{g(b) - g(a)} (g(b) - g(a)) \;=\; \epsilon
\end{aligned}
$$

as desired. ∎

Proposition 6.21 *Suppose that $\phi_n \to f$ and $\psi_n \to f$, both uniformly on $[a, b]$ where ϕ_n and ψ_n are step functions for each n and that $g : [a, b] \to I\!R$ is an increasing function. Then*

$$\lim_n \left| \int_a^b \phi_n(x)\ dg(x) - \int_a^b \psi_n(x)\ dg(x) \right| = 0.$$

Proof: Let $\epsilon > 0$ be given. Then there is an $N \in I\!N$ so

$$|\phi_n(x) - f(x)| < \frac{\epsilon}{2(g(b) - g(a))} \quad \text{and} \quad |\psi_n(x) - f(x)| < \frac{\epsilon}{2(g(b) - g(a))}$$

for all $n \geq N$. Fix $n \geq N$ and let $\{A_i\}$ be a partition of $[a, b]$ adapted to both ϕ_n and ψ_n so that

$$\phi_n(x) = \sum_i a_i \chi_{A_i}(x) \quad \text{and} \quad \psi_n(x) = \sum_i b_i \chi_{A_i}(x).$$

Then

$$
\begin{aligned}
\left| \int_a^b \phi_n(x)\ dg(x) - \int_a^b \psi_n(x)\ dg(x) \right| &= \left| \sum_i (a_i - b_i) g_*(A_i) \right| \\
&\leq \sum_i |a_i - b_i| g_*(A_i) \\
&\leq \sum_i \left(\frac{\epsilon}{(g(b) - g(a))} \right) g_*(A_i) \\
&= \frac{\epsilon}{g(b) - g(a)} \sum_i g_*(A_i) = \epsilon
\end{aligned}
$$

as desired. ∎

Now we can define the integral of $f \in \mathcal{I}[a, b]$ with respect to an increasing g.

Definition 6.7 *Let $g : [a,b] \to \mathbb{R}$ be increasing and $f \in \mathcal{I}[a,b]$. Then the integral of f with respect to g is*

$$\int_a^b f(x)\ dg(x) = \lim_n \int_a^b \phi_n(x)\ dg(x)$$

where (ϕ_n) is any sequence of step functions on $[a,b]$ which converges uniformly to f on $[a,b]$.

Example Suppose that $a < b < c$ and

$$g(x) = \begin{cases} 0, & \text{if } x < b \\ 1, & \text{if } x \geq b \end{cases}.$$

Then

$$\int_a^c f(x)\ dg(x) = \lim_{x \to b^-} f(x).$$

This is easy to see if f is continuous at b. Otherwise, suppose that f is a step function with a step at b and let $\lim_{x\to b^-} f(x) = \alpha$ while $\lim_{x \to b^+} f(x) = \beta$. Then we see that

$$\int_a^c f(x)\,dg(x) = 0+\cdots+0+\alpha(g(b)-g(b-\delta))+\beta(g(b+\gamma)-g(b))+0+\cdots+0 = \alpha(1) = \alpha.$$

Example If $g(x) = x$, then we get the same integral that we defined in Definition 6.3.

Proposition 6.22 *Let $f_1, f_2 \in \mathcal{I}[a,b]$ and $g : [a,b] \to \mathbb{R}$ be increasing and $\alpha \in \mathbb{R}$ and $c \in (a,b)$. Then*

1. $\displaystyle\int_a^b (f_1(x) + f_2(x))\ dg(x) = \int_a^b f_1(x)\ dg(x) + \int_a^b f_2(x)\ dg(x)$
2. $\displaystyle\int_a^b \alpha f_1(x)\ dg(x) = \alpha \int_a^b f_1(x)\ dg(x)$
3. $\displaystyle\int_a^b f_1(x)\ dg(x) = \int_a^c f_1(x)\ dg(x) + \int_c^b f_1(x)\ dg(x).$

Proof: These properties are all clearly true for step functions. ▮

The following theorem should be no surprise; we leave its proof to the exercises.

Proposition 6.23 *Let (f_n) be a sequence in $\mathcal{I}[a,b]$ which converges uniformly to f and let $g : [a,b] \to I\!R$ be an increasing function. Then*

$$\lim_n \int_a^b f_n(x)\ dg(x) = \int_a^b f(x)\ dg(x).$$

In the special case that g is differentiable, there is a nice and simple relationship between the Riemann-Stieltjes integral and the plain vanilla Riemann integral. The basic idea is simply that when x_{i+1} is very close to x_i, $g(x_{i+1}) - g(x_i) \approx g'(x_i)(x_{i+1} - x_i)$.

Theorem 6.24 *Suppose that $g : [a,b] \to I\!R$ is differentiable and increasing and that $f, g' \in \mathcal{I}[a,b]$. Then*

$$\int_a^b f(x)\ dg(x) = \int_a^b f(x)g'(x)\ dx.$$

Proof: We can construct sequences of step functions (ϕ_n) and (ψ_n) so that $|\phi_n(x) - f(x)| < 1/n$ and $|\psi_n(x) - g'(x)| < 1/(2n)$ for all $x \in [a,b]$. Furthermore, we can assume that $\mathcal{P}_n = \{a = x_0^n < x_1^n < x_2^n < \cdots < x_{k_n}^n = b\}$ is a partition of $[a,b]$ so that both ϕ_n and ψ_n are based on $\mathcal{P}_n$. By the Mean Value Theorem, there are points $\hat{x_i^n} \in (x_i^n, x_{i+1}^n)$ so that $g(x_{i+1}^n) - g(x_i^n) = g'(\hat{x_i^n})(x_{i+1}^n - x_i^n)$. Define the step functions η_n by

$$\eta_n(t) = \sum_i g'(\hat{x_i^n})\chi_{[x_i^n, x_{i+1}^n]}(t).$$

Then we know that $|\eta_n(x) - g'(x)| < 1/n$ (by the proof of Theorem 6.9). This yeilds

$$\int_a^b \phi_n(x)\, dg(x) = \sum_i \alpha_i(g(x_{i+1}^n) - g(x_i^n)) = \sum_i \alpha_i g'(\hat{x_i^n})(x_{i+1}^n - x_i^n) = \int_a^b \phi_n(x)\eta_n(x)\, dx.$$

Furthermore, as $n \to \infty$ we see that

$$\int_a^b \phi_n(x)\, dg(x) \to \int_a^b f(x) dg(x) \quad \text{and} \quad \int_a^b \phi_n(x)\eta_n(x)\, dx \to \int_a^b f(x)g'(x)\, dx$$

as desired. ∎

What if we wish to integrate a function with respect to a more general function g? Well, clearly if $g(x) = g_1(x) - g_2(x)$ where g_1, g_2 are both increasing functions on $[a,b]$, then we can define

$$\int_a^b f(x)\ dg(x) = \int_a^b f(x)\ dg_1(x) - \int_a^b f(x)\ dg_2(x).$$

So, what kinds of functions are differences of two increasing functions? In fact, the collection of *functions of bounded variation* are exactly this class (see the projects for more on this). For the moment, we will define the class of functions of bounded variation as those functions which are the difference of two increasing functions. The projects give an alternate definition (which also explains the term "bounded variation").

Definition 6.8 *The class of* functions of bounded variation on the interval $[a, b]$ *is defined to be the collection*

$$BV[a, b] = \{f_1(x) - f_2(x) : f_1, f_2 \text{ are increasing functions } \}.$$

We will prove that every function of bounded variation is in $\mathcal{I}[a, b]$.

Theorem 6.25 *Suppose that $f : [a, b] \to I\!R$ is the difference of two bounded increasing functions, then $f \in \mathcal{I}[a, b]$.*

Proof: Since the class $\mathcal{I}[a, b]$ is closed under linear combinations, it suffices to assume that f is an increasing function.

Let $\epsilon > 0$ be given. Choose a partition $Q = \{f(a) = y_0 < y_1 < y_2 < \cdots < y_n = f(b)\}$ of $[f(a), f(b)]$ sufficiently fine so that $y_i - y_{i-1} < \epsilon/2$. Define $x_i = \sup\{x \in [a, b] : f(x) \le y_i\}$. Notice that it is possible to have $x_i = x_{i+1}$, but we ignore this since we can just eliminate these repeats. Define the step function $\phi : [a, b] \to I\!R$ by

$$\phi(t) = \sum_i y_i \chi_{[x_{i-1}, x_i]}(t).$$

Then we have $|f(x) - \phi(x)| < \epsilon$ by construction.

We have just shown that for any $\epsilon > 0$ there is a step function ϕ so that $|\phi(x) - f(x)| < \epsilon$ for all $x \in [a, b]$. Thus, we can construct a sequence of step functions (ϕ_n) with $\phi_n \to f$ uniformly on $[a, b]$ which implies that $f \in \mathcal{I}[a, b]$. ∎

Why do we care? Well, for one we would like to prove a version of the of integration by parts for these Riemann-Stieltjes integrals.

Theorem 6.26 *(Integration by Parts) Suppose that both $f, g \in BV[a, b]$. Then*

$$\int_a^b f(x)\ dg(x) + \int_a^b g(x)\ df(x) = f(b)g(b) - f(a)g(a).$$

Proof: We can use Theorem 6.9 to see that there are sequences of step functions $\phi_n \to f$ and $\psi_n \to g$ uniformly and with $\mathcal{P}_n = \{a = x_0^n < x_1^n < \cdots < x_{k_n}^n = b\}$ a partition of $[a, b]$ and

$$\phi_n(t) = \sum_{i=1}^{k_n} f(\eta_i^n)\chi_{[x_{i-1}^n, x_i^n]}(t)$$

with $x_{i-1}^n < \eta_i^n < x_i^n$. Furthermore, we can assume that

$$\psi_n(t) = \sum_{i=1}^{k_n} g(x_i^n)\chi_{[\eta_{i-1}^n, \eta_i^n]}(t).$$

Then ϕ_n and ψ_n are increasing for all n. Now, computing we see that

$$\int_a^b \phi_n(t)\ dg(t) = \sum_{i=1}^{k_n} f(\eta_i^n)(g(x_i^n) - g(x_{i-1}^n))$$

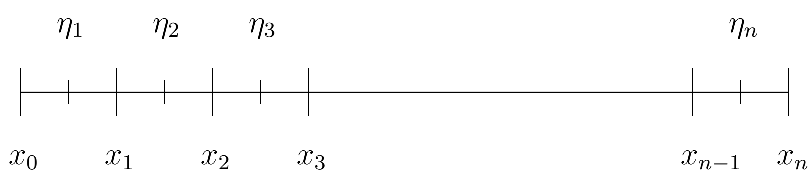
and

$$\int_a^b \psi_n(t)\ df(t) = \sum_{i=1}^{k_n} g(x_i^n)(f(\eta_i^n) - f(\eta_{i-1}^n)).$$

Adding these together and simplifying by removing the cancelling terms we get

$$\int_a^b \phi_n(t)\ dg(t) + \int_a^b \psi_n(t)\ df(t) = f(b)g(b) - f(a)g(a).$$

Taking limits as n tends to ∞ we get the desired result. ∎



The next theorem we present without proof.

Theorem 6.27 *(Bounded Convergence Theorem) Let $g : [a,b] \to \mathbb{R}$ be increasing and $f_n \to f$ pointwise on $[a,b]$ with $f_n, f \in \mathcal{I}[a,b]$. Suppose further that there is an $M > 0$ so that $|f_n(x)| \leq M$ for all n and $x \in [a,b]$. Then*

$$\lim_n \int_a^b f_n(x)\ dg(x) = \int_a^b f(x)\ dg(x).$$

As a simple corollary, we have the Monotone Convergence Theorem.

Theorem 6.28 *(Monotone Convergence Theorem) Let $g : [a,b] \to \mathbb{R}$ be increasing and $f_n, f \in \mathcal{I}[a,b]$ with $f_n \to f$ pointwise. Suppose further that $f_1(x) \leq f_2(x) \leq \cdots \leq f_n(x) \leq \cdots \leq f(x)$ for all $x \in [a,b]$. Then*

$$\lim_n \int_a^b f_n(x)\ dg(x) = \int_a^b f(x)\ dg(x).$$

Proof: We simply see that $|f_n(x)| \leq M$ for all n and $x \in [a,b]$ where

$$M = \sup_{x\in[a,b]} |f_1(x)| + \sup_{x\in[a,b]} |f(x)|$$

and so we use the Bounded Convergence Theorem. ∎

Riemann-Stieltjes integrals and change-of-variables: what does that dx mean?

Take an integral like

$$\int_0^{\pi/2} \sin(2x)\ dx.$$

Riemann-Stieltjes integrals help explain why we use the dx. There are really two different reasons. The first is that it clearly indicates what the variable is that is varying in the integral (so that any other variables involved which are independent of x can be "brought out" of the integral). The other reason is more subtle and is involved in the funny buissiness with du in u-substitution. That is, if we want to integrate

$$\int_0^1 \sin(\pi x^2) 2x\ dx$$

we set $u = x^2$ and then $du = 2xdx$ so this integral becomes

$$\int_0^1 \sin(u)\ du.$$

Changing from dx to du makes sense, since we are now "integrating with respect to u" instead of x. However, what about that funny $du = 2x\ dx$? Let's look at a simpler example first.

We know that

$$\int_0^{\pi/2} \sin(2x)\ dx = \frac{1}{2}\int_0^{\pi} \sin(u)\ du = \int_0^{\pi} \sin(u)\ d(u/2) = \int_0^{\pi} \sin(u)\ dg(u)$$

where $g(u) = u/2$.

It helps to take a specific Riemann sum to understand this. Take the partition $\{0, \pi/8, \pi/4, 3\pi/8, \pi/2\}$ of $[0, \pi/2]$ and the corresponding Riemann sum

$$\sin(2\times\frac{\pi}{8})\left(\frac{\pi}{8}-0\right)+\sin(2\times\frac{\pi}{4})\left(\frac{\pi}{4}-\frac{\pi}{8}\right)+\sin(2\times\frac{\pi}{4})\left(\frac{3\pi}{8}-\frac{\pi}{4}\right)+\sin(2\times\frac{3\pi}{8})\left(\frac{\pi}{2}-\frac{3\pi}{8}\right)$$

This Riemann sum is illustrated in the top image in Figure 6.1.

Compare this to the Riemann sum for the integral $\int_0^{\pi} \sin(u)\ du$ for the partition $\{0, \pi/4, \pi/2, 3\pi/4, \pi\}$

$$\sin(\frac{\pi}{4})\left(\frac{\pi}{4}-0\right)+\sin(\frac{\pi}{2})\left(\frac{\pi}{2}-\frac{\pi}{4}\right)+\sin(\frac{3\pi}{4})\left(\frac{3\pi}{4}-\frac{\pi}{2}\right)+\sin(\pi)\left(\pi-\frac{3\pi}{4}\right)$$

(illustrated in the bottom image of Figure 6.1). The only differnce between these two Riemann sums is in the width of the subintervals; the widths for the second sum are twice the widths of the corresponding widths for the first sum. Clearly the factor of 1/2 corrects for this difference. Furthermore, this correspondence cleary works for ANY Riemann sum you care to put on $[0, \pi/2]$ – that is, for

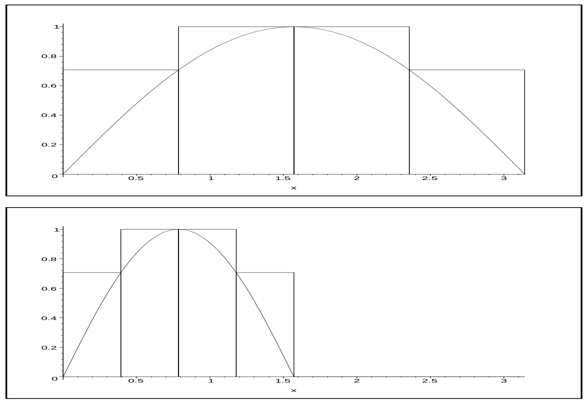


Figure 6.1: Two different approximations to $\int_0^{\pi/2} \sin(2x)\ dx$

any such Riemann sum there is an associated one on $[0, \pi]$ which has exactly twice the value.

This is all very simple. The Riemann-Stieltjes integral comes in when we take the correspondance

$$\int_0^{\pi/2} \sin(2x)\ dx = \int_0^{\pi} \sin(u)\ dg(u)$$

for $g(u) = u/2$ (a perfectly good bounded increasing function on $[0, \pi]$). The interpretation here is that in order to make the make the corresponding Riemann sums have the same value we must somehow measure "length" on $[0, \pi]$ differently than on $[0, \pi/2]$ (in fact, we must divide by a factor of 2).

To say this more precisely, for any partition $\{0 = x_0 < x_1 < \cdots < x_n = \pi/2\}$ of $[0, \pi/2]$ and Riemann sum

$$\sum_i \sin(2x_i^*)\ (x_{i+1} - x_i)$$

there is a corresponding partition $\{0 = y_0 < y_1 < \cdots < y_n = \pi\}$ of $[0, \pi]$ (with $y_i = 2x_i$ and $y_i^* = 2x_i^*$) and a Riemann-Stieltjes sum

$$\sum_i \sin(y_i^*)\ (g(y_{i+1}) - g(y_i))$$

where $g(u) = u/2$. These two Riemann sums clearly have exactly the same value since we are evaluating sin at exactly the same points and we have the same "measure of length" of the corresponding subintervals in the two partitions.

Ok, so that is what happens in the simple case of substituting $u = 2x$. What if the substitution is more complicated?

Well, let's now look at a slightly more complicated example. Suppose we wish to compute the integral

$$\int_0^1 \sin(\pi x^2) 2x \; dx$$

Clearly we will use the substitution $u = x^2$ and see that $du = 2x$ so we would then integrate

$$\int_0^1 \sin(\pi u) \; du.$$

Notice that if we let $g(x) = x^2$, we have

$$\int_0^1 \sin(\pi u) \; du = \int_0^1 \sin(\pi x^2) \; 2x \; dx = \int_0^1 \sin(\pi x^2) \; g'(x) \; dx = \int_0^1 \sin(\pi x^2) \; dg(x)$$

(by Theorem 6.24). We use Riemann-Stieltjes sums to illustrate the equality of the first and last integral (out of the four immediately above).

This time, we take the partition $\{0, 1/4, 1/2, 3/4, 1\}$ of $[0, 1]$ and the Riemann-Stieltjes sum

$$\sin\left(\pi\left(\tfrac{1}{4}\right)^2\right)\left(g(\tfrac{1}{4}) - g(0)\right) + \sin\left(\pi\left(\tfrac{1}{2}\right)^2\right)\left(g(\tfrac{1}{2}) - g(\tfrac{1}{4})\right) + \sin\left(\pi\left(\tfrac{1}{\sqrt{2}}\right)^2\right)\left(g(\tfrac{3}{4}) - g(\tfrac{1}{2})\right) + \sin\left(\pi\left(\tfrac{3}{4}\right)^2\right)\left(g(1) - g(\tfrac{3}{4})\right).$$

This corresponds exactly to the Riemann sum

$$\sin(\pi\frac{1}{16})\left(\frac{1}{16} - 0\right) + \sin(\pi\frac{1}{4})\left(\frac{1}{4} - \frac{1}{16}\right) + \sin(\pi\frac{1}{2})\left(\frac{9}{16} - \frac{1}{4}\right) + \sin(\pi\frac{9}{16})\left(1 - \frac{9}{16}\right)$$

associated to the partition $\{0, 1/16, 1/4, 9/16, 1\}$. These two Riemann sums are illustrated in Figure 6.2.

Again, the thing to notice is that we use the term $dg(x)$ to correct for the distortion in the lengths of the subintervals of the partition introduced by the change of variable.

Application to Probability Theory

One important area of application of Riemann-Stieltjes integrals is in probability theory. Suppose that we have a probability density function (*a pdf*) $g(x)$ on an interval $[a, b]$. That is, $g(x) \geq 0$ and

$$\int_a^b g(x) d(x) = 1.$$

This pdf g could represent some random variable or process.

Define $G : [a, b] \to I\!R$ by

$$G(x) = \int_a^x g(t) \; dt$$

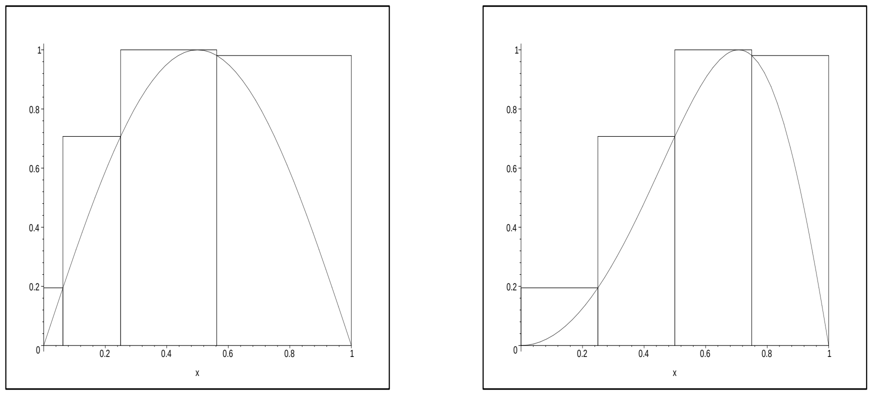


Figure 6.2: Two different approximations for $\int_0^1 \sin(\pi u)\ du$

(so that $G(x)$ is the accumulated "density" or the cdf (cumulative density function) of the distribution). Then if we want to know the expected value of our random variable, we compute

$$\int_a^b xg(x)\ dx = \int_a^b x\ dG(x).$$

Thus, we are using our pdf $g(x)$ as a way of weighting the interval $[a, b]$. We assign more "importance" to those areas where $g(x)$ is larger and less "importance" to those areas of $[a, b]$ where $g(x)$ is smaller.

Ok, so what? Why do we need the full power of the Riemann-Stieltjes integral? If all we want to do is integrate with respect to a density function, then we can just integrate $f(x)$ (some function) times $g(x)$, as in

$$\int f(x)g(x)\ dx.$$

What is new?

Well, sometimes our random variable has a distribution which does not have a density function. For example, if the cdf has discontinuities, then it cannot have a density, a pdf, (as the density should be the derivative of the cdf). So in this case we cannot use the form

$$\int f(x)g(x)\ dx$$

and we need the full Riemann-Stieltjes integral. The Riemann-Stieltjes integral allows us to handle discrete and continuous random variables in the same manner, with the same formalism. This makes handling the calculus of mixtures of discrete and continuous random variables easier.

In fact the situation is even more complicated than this. We now construct an example of a function $f : [0, 1] \to I\!R$ which is continuous and increasing with $f'(x) = 0$ for all x not in the standard middle third Cantor set. If we use this as a cdf for a random variable, the random variable is "supported" only on the Cantor set (that is, the probability "density" is only positive on the Cantor set). This probability distribution is not a discrete distribution but it isn't really what we think of as a continuous distribution either. We say that it is a *singular* distribution.

Recall the construction of the standard middle third Cantor set where we start with the interval $[0, 1]$ and successively remove the open middle third of each remaining closed interval. Based on this construction, we define a sequence of functions $f_n : [0, 1] \to [0, 1]$ which are all increasing and such that the sequence is uniformly Cauchy (so the limit exists). This limit, call it f, has the desired properties (see the excercises).

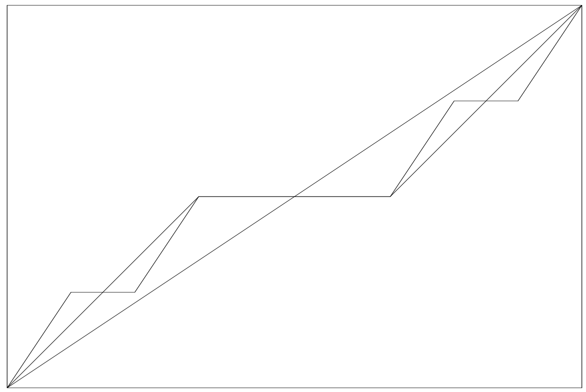
To see how to do this, we define $f_1(x) = x$. For $f_2(x)$, we define

$$f_2(x) = \begin{cases} (3/2)x, & \text{if } 0 \leq x \leq 1/3 \\ 1/2, & \text{if } 1/3 \leq x \leq 2/3 \\ (3/2)x - 1/2, & \text{if } 2/3 \leq x \leq 1 \end{cases}$$

so that f_2 linearly increases from 0 to 1/2 over the interval $[0, 1/3]$, stays constant on the interval $[1/3, 2/3]$ and then linearly increases from 1/2 to 1 over the interval $[2/3, 1]$. For f_3, we set $f_3(x) = f_2(x)$ for all $x \in [1/3, 2/3]$ (that is, where f_2 is constant) and only modify f_2 where it is non-constant. We do this in such a way that f_3 has the constant value 1/4 on the middle third of the interval $[0, 1/3]$ and has the constant value of 3/4 on the middle third of the interval $[2/3, 1]$. More formally,

$$f_3(x) = \begin{cases} (9/4)x, & \text{if } 0 \leq x \leq 1/9 \\ 1/4, & \text{if } 1/9 \leq x \leq 2/9 \\ (9/4)x - 1/4, & \text{if } 2/9 \leq x \leq 3/9 \\ 1/2, & \text{if } 1/3 \leq x \leq 2/3 \\ (9/4)x - 1, & \text{if } 2/3 \leq x \leq 7/9 \\ 3/4, & \text{if } 7/9 \leq x \leq 8/9 \\ (9/4)x - 5/4, & \text{if } 8/9 \leq x \leq 1 \end{cases}$$

We illustrate these first three functions in the figure. Using this same procedure, we construct f_n for all n. Since we only modify f_n on smaller and smaller sets (and by a smaller and smaller amount), it is at least believable that f_n converges.



This function is often called the *Cantor Ternary Function.*

Another way to construct the functions f_n is recursive. We define $f_1(x) = x$, as before and define

$$f_{n+1}(x) = \begin{cases} (1/2)f(3x), & \text{if } 0 \le x < 1/3 \\ 1/2, & \text{if } 1/3 \le x \le 2/3 \\ (1/2)f(3x-2) + 1/2, & \text{if } 2/3 < x \le 1 \end{cases}$$

for $n \ge 1$. One uses this recursive definition to show that f_n is continuous and increasing for all n and is a uniformly Cauchy sequence. This recursive definition expresses the fact that f_{n+1} on $[0, 1/3]$ "looks" like f_n on all of $[0, 1]$, except scaled down. Similarly for f_{n+1} on $[2/3, 1]$ – it "looks" like f_n on $[0, 1]$, except scaled down and shifted up by $1/2$.

Some comments on the Lebesgue Integral

The basic idea underlying the Lebesgue integral is to use a more general class of functions than the step functions to approximate a given function. As an example, suppose we want to integrate the function $f : [0, 1] \to I\!R$ given by

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

We cannot uniformly approximate this function with step functions since in any interval there are both irrational and rational numbers (thus we have to approximate both 0 and 1 in the same place). However, suppose we let A be the set of rational numbers in $[0, 1]$ and $B = [0, 1] \setminus A$ be the set of irrational numbers in $[0, 1]$. Then $f(x) = 0\chi_A(x) + 1\chi_B(x)$ so we would expect that

$$\int_0^1 f(x)\ dx = 0\ length(A) + 1\ length(B)$$

for some appropriate definition of $length(A)$ and $length(B)$.

The first idea is to make precise what we mean by the *length* of some general set. It turns out that if we want our *length* function to have nice properties, not all sets have "length". Those that do we call *measurable sets*. Then a *simple function* $\phi(x)$ is any function of the form

$$\phi(x) = \sum_i \alpha_i \chi_{A_i}(x)$$

where $\{A_1, A_2, A_3, \ldots, A_n\}$ is a finite collection of pairwise disjoint measurable sets and $\alpha_1, \alpha_2, \ldots, \alpha_n$ are any real numbers.

Then one integrates those functions which are limits of simple functions.

Problems

1. Find the integral of the function $f(x) = x^3$ on the interval $[0, 1]$. To do this construct a sequence $\{\phi_n\}$ of step functions which converge uniformly to f and compute the limit of the integrals of ϕ_n.

2. Suppose that ϕ and ψ are both step functions. Is $\phi\psi$ a step function as well? If it is, prove it. If it is not, give a counter-example. What about $1/\phi$?

3. Define the function

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

for $x \in [0, 1]$. Show that $f \notin \mathcal{I}[0, 1]$. (That is, show that there is no sequence of step functions which converges uniformly to f on $[0, 1]$).

4. Suppose that $f : [a, b] \to [0, \infty)$ is continuous with $f(x_0) > 0$ for some $x_0 \in [a, b]$. Prove that

$$\int_a^b f(x)\, dx > 0.$$

Hint: Show that there are $\epsilon, \delta > 0$ and a step function ϕ so that $0 \leq \phi(x) \leq f(x)$ for all x and

$$\int_a^b \phi(x)\, dx = \epsilon\delta.$$

5. Show that if $f \in \mathcal{I}$ then $|f| \in \mathcal{I}$ (that is, that if there is some sequence $\{\phi_n\}$ of step functions which converge uniformly to f then there is a sequence $\{\psi_n\}$ of step functions which converge uniformly to $|f|$).

6. A *rectangle* in $\mathbb{R}^2$ is a set of the form $R = \{(x, y) : a \leq x \leq b, c \leq y \leq d\}$ (or we could leave out one or more boundaries, that is, the $\leq$ could become $<$). A step function $\phi : \mathbb{R}^2 \to \mathbb{R}$ is a finite linear combination of characteristic functions of rectangles.

(a) Let $\phi : [a, b] \times [c, d] \to I\!R$ be a step function. Show that for any fixed $y \in [c, d]$ the function $f(x) = \phi(x, y)$ is a step function from $[a, b] \to I\!R$.

(b) Suppose that $\phi_n : [a, b] \times [c, d] \to I\!R$ is a sequence of step functions which converge uniformly to $f : [a, b] \times [c, d] \to I\!R$. Show that for any fixed $y \in [c, d]$ we have that $f_n(x) = \phi_n(x, y)$ is a sequence of step functions which converge uniformly to $f(x, y)$.

7. Find an example of an $f \in \mathcal{I}$ with $g \notin \mathcal{I}$ where we define

$$g(x) = \begin{cases} -1, & \text{if } f(x) < 0 \\ 0, & \text{if } f(x) = 0 \\ 1, & \text{if } f(x) > 0 \end{cases}$$

8. Show that if f is continuous and for all $g \in \mathcal{I}$ we have that

$$\int_a^b f(x)g(x)\ dx = 0$$

then $f(x) = 0$ for all $x \in [a, b]$.

9. Let $f \in \mathcal{I}$. Suppose that $c \in (a, b)$. Prove that $f \in \mathcal{I}[a, c]$ and $f \in \mathcal{I}[c, b]$. Furthermore, prove that

$$\int_a^b f(x)\ dx = \int_a^c f(x)\ dx + \int_c^b f(x)\ dx.$$

10. Let $C[a, b]$ be the set of continuous functions $f : [a, b] \to I\!R$. Define

$$\langle f, g\rangle = \int_a^b f(x)g(x)\ dx.$$

Prove that for all $f, g, h \in C[a, b]$ and $\alpha \in I\!R$

- $\langle f, f\rangle \geq 0$.
- $\langle f, f\rangle = 0$ only if $f(x) = 0$ for all x.
- $\langle f, g\rangle = \langle g, f\rangle$.
- $\langle f, g + h\rangle = \langle f, g\rangle + \langle f, h\rangle$.
- $\langle \alpha f, g\rangle = \alpha\langle f, g\rangle = \langle f, \alpha g\rangle$.

(A function $\langle \cdot, \cdot\rangle : V \times V \to I\!R$ which satisfies these properties is called an *inner product* on the vector space V).

11. Suppose that $f', g' \in \mathcal{I}$ and $f(a) \leq g(a)$ and $f'(x) \leq g'(x)$ for all $x \in [a, b]$. Prove that $f(x) \leq g(x)$ for all $x \in [a, b]$.

12. Let $f_n \in \mathcal{I}[0,1]$ with $f_n \to f$ uniformly on $[0,1]$. Define

$$F_n(x) = \int_0^x f_n(t)\ dt \quad \text{and} \quad F(x) = \int_0^x f(t)\ dt.$$

Does $F_n \to F$ uniformly on $[0,1]$? If so, prove it. If not, find a counterexample.

13. Consider the function $f : [0,1] \to I\!R$ defined by $f(0) = 0$ and

$$f(x) = (-1)^n \quad \text{for } \frac{1}{n+1} < x \leq \frac{1}{n}$$

(so that $f(x) = -1$ for $x \in (1/2, 1]$ and $f(x) = 1$ for $x \in (1/3, 1/2]$ etc). Prove that $f \notin \mathcal{I}[0,1]$ but that

$$\int_0^1 f(x)\ dx$$

exists as an improper integral. (As a note, the same proof can show that $g(x) = \sin(1/x)$ with $g(0) = 0$ can be integrated via an improper integral).

14. Suppose that $f : [0,1] \to I\!R$ is bounded and that f restricted to $[1/n, 1]$ is integrable for all $n \in I\!N$. Show that

$$\int_0^1 f(x)\ dx$$

exists as an improper integral.

15. If $|f(x)| \leq M$ for all $x \in [a,b]$, prove that the function F defined by

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

is uniformly continuous on $[a,b]$.

16. Suppose that $f : [a,b] \to [0,\infty)$ is an integrable function. Prove that there is a $c \in [a,b]$ so that

$$\int_a^c f(x)\ dx = \int_c^b f(x)\ dx.$$

17. Find the polynomial $p(x) = a + bx + cx^2$ so that

$$\|p(x) - \sin(x)\|_2 = \left(\int_{-\pi}^{\pi} (p(x) - \sin(x))^2\ dx \right)^{1/2}$$

is minimized.

18. Suppose that $f \in PC(2\pi)$ is continuous and that

$$0 = \int_{-\pi}^{\pi} f(x)\ dx = \int_{-\pi}^{\pi} f(x)\sin(nx)\ dx = \int_{-\pi}^{\pi} f(x)\cos(nx)\ dx$$

for all $n \in I\!N$. Prove that $f(x) = 0$ for all x.

19. Suppose that f is continuous on $[a, b]$ and that there are two nonzero numbers $\alpha \neq -\beta$ so that for all $c \in (a, b)$ we have that

$$\alpha \int_{a}^{c} f(x)\ dx = \beta \int_{c}^{b} f(x)\ dx.$$

Prove that $f(x) = 0$ for all $x \in [a, b]$.

20. Provide the details for the proof of Proposition 6.22.

21. Prove Proposition 6.23.

22. Prove that the Cantor Ternary function f (as constructed in the text) is increasing and continuous. Show further that $f'(x)$ exists and equals to 0 for all x not in the Cantor set.

23. Define $g : [0, 6] \to I\!R$ by

$$g(x) = \begin{cases} x, & \text{if } 0 \leq x < 1 \\ 1, & \text{if } 1 \leq x < 2 \\ 2, & \text{if } 2 \leq x < 3 \\ x/3 + 1, & \text{if } 3 \leq x < 6 \end{cases}$$

Compute the integrals

$$\int_{0}^{5} x\ dg(x) \quad \text{and} \quad \int_{0}^{5} (3 - x)\ dg(x).$$

24. Let $g : [0, 1] \to I\!R$ be defined by $g(x) = \sqrt{x}$. Compute

$$\int_{0}^{1} x\ dg(x) \quad \text{and} \quad \int_{0}^{1} 1\ dg(x) \quad \text{and} \quad \int_{0}^{1} 1/\sqrt{x}\ dg(x)$$

25. Let $g(x) = \tan(x)$ and consider the integral

$$\int_{-\pi/2}^{\pi/2} f(x)\ dg(x).$$

Assuming that this integral exists, show that it is the same as

$$\int_{-\infty}^{\infty} f\left(\arctan(z)\right)\ dz.$$

Compute the integral

$$\int_{-\pi/2}^{\pi/2} \cos^2(x)\ dg(x).$$

26. As a generalization of the previous problem, suppose that $g : (a, b) \to I\!R$ is continuous and strictly increasing with range (c, d) (including the possibility that $c = -\infty$ and/or $d = \infty$). Show that

$$\int_a^b f(x)\ dg(x) = \int_c^d f(g^{-1}(t))\ dt.$$

What happens if g is continuous and increasing, but not strictly increasing (that is, g might be constant on some interval and so g^{-1} doesn't exist)?

27. Let $f \in \mathcal{I}[0, 2]$ and $g : [0, 2] \to I\!R$ be defined by

$$g(x) = \begin{cases} g_1(x), & \text{if } 0 \le x \le 1 \\ g_2(x), & \text{if } 1 < x \le 2 \end{cases}$$

where $g_1 : [0, 1] \to [1, 2]$ is continuous and strictly increasing and $g_2(x) : (1, 2] \to (3, 4]$ is continuous and strictly increasing (so $g_1(0) = 1, g_1(1) = 2, g_2(2) = 4$ and $g_2(x) \searrow 3$ as $x \searrow 1$). Show that

$$\int_0^2 f(x)\ dg(x) = \int_1^2 f(g^{-1}(z))\ dz + \int_2^3 f(g^{-1}(2))\ dz + \int_3^4 f(g^{-1}(z))\ dz.$$

28. Let $f \in \mathcal{I}[0, 10]$ and $g; [0, 10] \to I\!R$ be given by $g(x) = x^2$. Show that there is an $M > 0$ so that

$$\left| \int_c^d f(x)\ dg(x) \right| \le M(d - c)$$

for any $[c, d] \subset [a, b]$.

29. Let $f \in \mathcal{I}[0, 1]$ with $m = \inf_x f(x) = f(1/3)$ and $M = \sup_x f(x) = f(2/3)$. Let $\mathcal{G}$ be the set of all possible increasing functions $g : [0, 1] \to I\!R$ with $g(0) = 0$ and $g(1) = 1$. Show that

$$m \le \int_0^1 f(x)\ dg(x) \le M$$

for all $g \in \mathcal{G}$. Find $g_m, g_M \in \mathcal{G}$ so that

$$M = \int_0^1 f(x)\ dg_M(x) \quad \text{and} \quad m = \int_0^1 f(x)\ dg_m(x).$$

30. Suppose that $g : [0, 1] \to I\!R$ is differentiable. Find a necessary condition on g so that

$$\int_0^1 f(x)\ dx \le \int_0^1 f(x)\ dg(x)$$

for all $f \in \mathcal{I}[0, 1]$ with $f(x) \ge 0$.

Chapter 7

Infinite Series

Definition 7.1 *Let (a_n) be a sequence of real numbers. An* infinite series *with* terms a_n *is the expression*

$$\sum_{i=1}^{\infty} a_i.$$

Note that we are NOT defining this as a number. It is formally just a collection of symbols.

Definition 7.2 *The* partial sums *of the infinite series $\sum_i a_i$ is the sequence of real numbers*

$$s_n = \sum_{i=1}^{i=n} a_i.$$

We say that the infinite series $\sum_i a_i$ converges *if the sequence (s_n) of partial sums of the series has a limit, s. In this case, we write*

$$\sum_{i=1}^{\infty} a_i = s$$

and we say that s is the sum of the series. We say that an infinite series $\sum_i a_i$ diverges *if it does not converge.*

Example If $a_k = 0$ for all $k \geq N$, then $\sum_{i=1}^{\infty} a_i$ converges to s_N.

Example The series $\sum_{i=1}^{\infty}(-1)^i$ diverges since in this case

$$s_n = \begin{cases} 0, & \text{if } n \text{ is even} \\ -1, & \text{if } n \text{ is odd.} \end{cases}$$

Theorem 7.1 *Let (a_k) be a sequence of real numbers such that the series $\sum_{i=1}^{\infty} a_i$ converges. Then $\lim_{i\to\infty} a_i = 0$.*

Proof: Since $\sum_i a_i$ converges we have that $s_n \to s$ as $n \to \infty$. But then (s_n) is a Cauchy sequence so $a_n = s_n - s_{n-1} \to 0$ as $n \to \infty$. ▮

There are three very important series whose convergence are easy to determine.

Telescoping series

Suppose that $a = \lim_{k\to\infty} a_k$ exists. Then

$$\sum_{i=1}^{\infty} a_i - a_{i-1}$$

converges to $a_1 - a$.

It is easy to see that this is the case, since $s_n = a_1 - a_{n+1} \to a_1 - a$.

Geometric series

Let $|r| < 1$, then the series

$$\sum_{i=0}^{\infty} r^i$$

converges to $1/(1-r)$. To see this, just note that since

$$s_n = 1 + r + r^2 + \cdots + r^n$$

we know that

$$rs_n = r + r^2 + r^3 + \cdots + r^{n+1}$$

so that

$$(1-r)s_n = 1 - r^{n+1} \qquad \Rightarrow \qquad s_n = \frac{1-r^{n+1}}{1-r} \to \frac{1}{1-r}.$$

On the other hand, if $|r| \geq 1$ then the series diverges.

Harmonic series

The series

$$\sum_{i=1}^{\infty} 1/i$$

diverges to ∞. To see this, just notice that

$$\begin{aligned} s_{2^n} &= 1 + 1/2 + 1/3 + 1/4 + 1/5 + \cdots + 1/8 + 1/9 + \cdots + 1/16 + 1/17 + \cdots + 1/32 \\ &+ 1/33 + \cdots + 1/2^{n-1} + \cdots + 1/2^n \\ &\geq 1 + 1/2 + 2(1/4) + 4(1/8) + 8(1/16) + 16(1/32) + \cdots + 2^{n-1}(1/2^n) \\ &= 1 + 1/2 + 1/2 + 1/2 + 1/2 + 1/2 + \cdots + 1/2 = 1 + n/2 \end{aligned}$$

Proposition 7.2 *If $a_n \geq 0$ for all n then $\sum_{n=1}^{\infty} a_n$ converges iff (s_n) is a bounded sequence.*

Proof: In this case (s_n) is an increasing sequence so it converges iff it is bounded. ∎

Theorem 7.3 *(Cauchy Criteria for Series) Let (a_n) be a real sequence. Then $\sum_n a_n$ converges iff for all $\epsilon > 0$ there is an $N \in I\!N$ so that for all $m > n \geq N$ we have*

$$\left|\sum_{k=n}^{m} a_k\right| < \epsilon.$$

Proof: The condition is the same as that the sequence of partial sums is a Cauchy sequence. ∎

Corollary 7.4 *Let (a_n) be a real sequence. Then the series $\sum_k a_k$ converges iff for all $\epsilon > 0$ there is an $N \in I\!N$ so that $n \geq N$ implies that*

$$\left|\sum_{k=n}^{\infty} a_k\right| < \epsilon.$$

Theorem 7.5 *(Arithmetic Properties) Let (a_k) and (b_k) be real sequences. If $\sum_k a_k$ and $\sum_k b_k$ are both convergent series, then*

$$\sum_k (a_k + b_k) = \sum_k a_k + \sum_k b_k$$

and

$$\sum_k \alpha\ a_k = \alpha \sum_k a_k$$

for all $\alpha \in I\!R$.

Proof: These are just statements about the corresponding sequences of partial sums. ∎

Example Let $a_k = (-1)^{k+1}$ and $b_k = (-1)^k$. Then $a_k + b_k = 0$ so $\sum_k a_k + b_k = 0$ but $\sum_k a_k$ and $\sum_k b_k$ both diverge.

Convergence Tests

We now turn to the discussion of tests of convergence of an infinite series.

Theorem 7.6 *(Comparison Test) Suppose there is some $N \in I\!N$ so that $0 \leq a_k \leq b_k$ for all $k \geq N$.*

1. *If $\sum_k b_k$ converges, then so does $\sum_k a_k$.*
2. *If $\sum_k a_k$ diverges, then so does $\sum_k b_k$.*

Proof: Set $s_n = \sum_{k=1}^n a_k$ and $t_n = \sum_{k=1}^n b_k$. Then for $n > N$ we have that $0 \leq s_n - s_N \leq t_n - t_N$. Thus if (t_n) is bounded (and thus convergent) so is (s_n) while if $s_n \to \infty$ so does (t_n). ∎

Theorem 7.7 *(Limit Comparison Test) Suppose $a_k \geq 0$ and $b_k \geq 0$ for all k and $\lim_k a_k/b_k = L$ exists.*

1. *If $0 < L < \infty$ then $\sum_k a_k$ converges iff $\sum_k b_k$ converges.*
2. *If $0 \leq L < \infty$ and $\sum_k b_k$ converges then so does $\sum_k a_k$.*
3. *If $0 < L \leq \infty$ and $\sum_k b_k$ diverges then so does $\sum_k a_k$.*

Proof: We prove the first statement, since the other ones have similar proofs.
We know that there is an $N \in I\!N$ so that for all $k \geq N$ we have

$$\left|\frac{a_k}{b_k} - L\right| < L/2 \quad \Rightarrow \quad \frac{L}{2} < \frac{a_k}{b_k} < \frac{3L}{2} \quad \Rightarrow \quad \left(\frac{L}{2}\right) b_k < a_k < \left(\frac{3L}{2}\right) b_k.$$

Thus by the Comparison Theorem we know that $\sum_k a_k$ converges iff $\sum_k b_k$ converges. ∎

Theorem 7.8 *(Integral Test) Suppose $f : [1, \infty) \to (0, \infty)$ is decreasing and integrable. Then the infinite series $\sum_k f(k)$ converges iff*

$$\lim_{N\to\infty} \int_1^N f(x)\ dx < \infty.$$

Proof: Define the step functions

$$\phi_n(x) = \sum_{i=1}^n f(i)\chi_{[i,i+1)}(x)$$

and

$$\psi_n(x) = \sum_{i=2}^{n+1} f(i)\chi_{[i-1,i)}(x).$$

Then $\phi_n(x) \geq f(x) \geq \psi_n(x)$ on $[1, n]$ so

$$\int_1^n \phi_n(x)\ dx \geq \int_1^n f(x)\ dx \geq \int_1^n \psi_n(x)\ dx$$

which implies

$$\sum_{i=1}^{n} f(i) \geq \int_1^n f(x)\ dx \geq \sum_{i=2}^{n+1} f(i).$$

Thus, if $\lim_{n\to\infty} \int_1^n f(x)\ dx < \infty$ then $\lim_{n\to\infty} \sum_{i=1}^{n} f(i) < \infty$ and conversely. ▮

Corollary 7.9 *(P-series) The series $\sum_{n=1}^{\infty} 1/n^p$ converges iff $p > 1$.*

Proof: Use the integral test for $p \neq 1$. ▮

Theorem 7.10 *(Ratio Test) Let (a_n) be a non-negative sequence with $a_k \neq 0$ for large k. Suppose that*

$$\lim_{k\to\infty} \frac{a_{k+1}}{a_k} = r$$

exists.

1. *If $r < 1$, then $\sum_{k=1}^{\infty} a_k$ converges.*
2. *If $r > 1$, then $\sum_{k=1}^{\infty} a_k$ diverges.*

Proof: If $r > 1$ then for large enough k we have

$$\frac{a_{k+1}}{a_k} > r - \epsilon > 1 \quad \Rightarrow \quad a_{k+n} > a_k(r-\epsilon)^n \to \infty$$

so $a_k \to \infty$ and thus $\sum_k a_k$ diverges.

Suppose that $r < 1$ and let $x \in (r, 1)$. Then there is an $N \in I\!N$ so for all $n \geq N$ we have that

$$\frac{a_{k+1}}{a_k} < x = \frac{x^{k+1}}{x^k} \quad \Rightarrow \quad \frac{a_{k+1}}{x^{k+1}} < \frac{a_k}{x^k}.$$

This means that the sequence $b_k = a_k/x^k$ is an eventually decreasing sequence. However, then there is some $M > 0$ so that $a_k/x^k < M$ for all k which implies that $a_k < Mx^k$. Now the series $\sum_k x^k$ converges since $0 < x < 1$, thus $\sum_k a_k$ converges by the Comparison Test. ▮

The proof of the next theorem is simple, so we leave it out.

Theorem 7.11 *(Root Test) Let (a_n) be a non-negative sequence. Suppose that*

$$R = \lim_{n\to\infty} \sqrt[n]{a_n}$$

exists. Then $\sum_n a_n$ converges if $R < 1$ and diverges if $R > 1$.

Definition 7.3 *Let (a_k) be a real sequence. If the series*

$$\sum_k |a_k|$$

converges, then we say that $\sum_k a_k$ converges absolutely*.*

Proposition 7.12 *If $\sum_k a_k$ converges absolutely then it converges.*

Proof: For each $m > n \in \mathbb{N}$, we have that

$$\left|\sum_{k=n}^{m} a_k\right| \le \sum_{k=n}^{m} |a_k|$$

so if $\sum_k a_k$ converges absolutely, then $\sum_k a_k$ converges by the Cauchy criteria. ∎

Definition 7.4 *If $\sum_k a_k$ converges but does not converge absolutely, we say that it* converges conditionally*.*

Definition 7.5 *If $a \in \mathbb{R}$ we define $a^+ = \max(a, 0)$ and $a^- = -\min(a, 0)$.*

Proposition 7.13 *If $\sum_k a_k$ converges conditionally, then*

$$\sum_k a_k^+ = \sum_k a_k^- = +\infty.$$

Proof: Suppose that $\sum_k a_k^+$ converges. Then both $\sum_k a_k$ and $\sum_k a_k^+$ converge and $a_k^- = a_k^+ - a_k$ so we know that $\sum_k a_k^-$ converges as well. However, then we have $|a_k| = a_k^+ + a_k^-$ so $\sum_k |a_k|$ must converge, which is a contradiction. ∎

Theorem 7.14 *(Riemann) Let $x \in \mathbb{R}$. If $\sum_k a_k$ converges conditionally, then there is some rearrangement of $\sum_k a_k$ which converges to x.*

Proof: (STRATEGY) We know $a_k^+ \to 0$ and $a_k^- \to 0$ and $\sum_k a_k^+ = \sum_k a_k^- = +\infty$.

Thus, start adding terms from (a_k^+) to get a value just above x and then start adding values from $(-a_k^-)$ to get below x. Since $a_k^+ \to 0$ and $a_k^- \to 0$, we should converge to x. ∎

There are many more convergence tests, but we content ourselves by just proving one more.

Theorem 7.15 *(Logarithm Test) Suppose $a_k > 0$ for large k and*

$$p = \lim_{k\to\infty} \frac{\ln(1/a_k)}{\ln(k)}$$

exists as an extended real number. If $p > 1$ then $\sum a_k$ converges. If $p < 1$ then $\sum_k a_k$ diverges.

Proof: Suppose that $p > 1$. Choose some q so that $1 < q < p$ and choose $N \in I\!N$ so that for all $n \geq N$ we have

$$\frac{\ln(1/a_k)}{\ln(k)} > q \quad \Rightarrow \quad \ln(1/a_k) > q\ln(k) \quad \Rightarrow \quad \ln(1/a_k) > \ln(k^q) \quad \Rightarrow \quad a_k < 1/k^q.$$

But then by the Comparison Theorem we know that $\sum_k a_k$ converges since $\sum_k 1/k^q$ converges.

On the other hand, if $p < 1$ then in a similar fashion we see that $a_k > 1/k$ for large k so $\sum_k a_k$ diverges. ∎

Convergence of Series of Functions

If (f_n) is a sequence of functions we say that the infinite series

$$\sum_k f_k$$

converges pointwise or uniformly when the sequence of partial sums converges pointwise or uniformly.

Theorem 7.16 *(Weierstrass M-test) Let $f_k : E \subset I\!R \to I\!R$ and $M_k \geq 0$ with $\sum_k M_k < \infty$. If $|f_k(x)| \leq M_k$ for all k and $x \in E$ then $\sum_k f_k(x)$ converges absolutely and uniformly in x (over E).*

Proof: Let $\epsilon > 0$. Since $\sum_k M_k < \infty$, there is an $N \in I\!N$ so that for all $m > n \geq N$ we have $\sum_{k=n}^m M_k < \epsilon$. But then for all $x \in E$ we have

$$\left|\sum_{k=n}^m f_k(x)\right| \leq \sum_{k=n}^m |f_k(x)| \leq \sum_{k=n}^m M_k < \epsilon.$$

Thus the sequence of functions $s_n(x) = \sum_{k=1}^n f_k(x)$ is a uniformly Cauchy sequence which converges absolutely for all $x \in E$. ∎

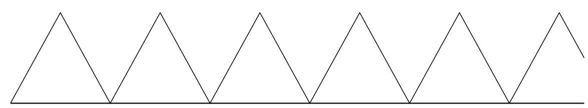
An everywhere continuous, nowhere differentiable function

We now use this theorem to give an example (due to Weierstrass) of a function $f : I\!R \to I\!R$ which is continuous on all R but has a derivative nowhere.

First we define $f_0(x)$ on $[0,1]$ by

$$f_0(x) = \begin{cases} x, & \text{if } 0 \le x \le 1/2 \\ 1-x, & \text{if } 1/2 \le x \le 1 \end{cases}$$

and extend f_0 to all of $\mathbb{R}$ by periodicity (that is, $f_0(x) = f_0(x+1)$).



Now for each $k \in \mathbb{N}$ define $f_k(x) = 2^{-k} f_0(2^k x)$ (that is, f_k is a sawtooth function that oscillates at a rate of 2^k faster than f_0).

Finally, we consider the series $\sum_{k=0}^{\infty} f_k(x)$. We know that $|f_k(x)| \le 2^{-k-1} = M_k$ and $\sum_k M_k = 1 < \infty$. Thus by the Weierstrass M-Test we know that the series converges uniformly and absolutely to a continuous function f on $\mathbb{R}$. To show that f is not differentiable anywhere it suffices to prove that f is not differentiable anywhere on $[0,1]$ (since it is periodic).

Now notice that $f_0'(y) = \pm 1$ as long as $2y \notin \mathbb{Z}$ (that is, if $y \ne n/2$ for any $n \in \mathbb{Z}$). Thus,

$$(*) \qquad f_k'(y) = \pm 1 \qquad \forall y \in \mathbb{R}, \quad 2^{k+1} y \notin \mathbb{Z}.$$

Suppose there is an $x \in [0,1)$ with f differentiable at x. For each n, choose $p_n \in \mathbb{Z}$ with $x \in [\alpha_n, \beta_n]$ where $\alpha_n = p_n/2^n$ and $\beta_n = (p_n+1)/2^n$. By the definition of the derivative, if f is differentiable at x then we have that

$$f'(x) = \lim_{n\to\infty} \frac{f(\beta_n) - f(\alpha_n)}{\beta_n - \alpha_n}.$$

Furthermore we see that for all $k < n$ we have

$$\frac{f_k(\beta_n) - f_k(\alpha_n)}{\beta_n - \alpha_n} = \pm 1.$$

On the other hand, since $f_0(y) = 0$ iff $y \in \mathbb{Z}$ we have that $f_k(\alpha_n) = f_k(\beta_n) = 0$ for all $k \ge n$. Thus

$$f(\beta_n) = \sum_{k=0}^{n-1} f_k(\beta_n) \quad \text{and} \quad f(\alpha_n) = \sum_{k=0}^{n-1} f_k(\alpha_n).$$

Since $f'(x)$ exists, we get that

$$f'(x) = \lim_{n\to\infty} \frac{f(\beta_n) - f(\alpha_n)}{\beta_n - \alpha_n} = \lim_{n\to\infty} \sum_{k=0}^{n-1} \frac{f_k(\beta_n) - f_k(\alpha_n)}{\beta_n - \alpha_n} = \sum_{k=0}^{\infty} f_k'(x)$$

so we must have that $\sum_k f_k'(x)$ exists. However, $f_k'(x) = \pm 1$, so the series cannot converge.

Power Series

Perhaps the simplest and thereby most important class of series of functions is the class of power series.

Definition 7.6 *A power series (centered at $x = x_0$) is a series of the form*

$$f(x) = \sum_{n=0}^{\infty} c_n (x - x_0)^n.$$

Proposition 7.17 *Let $\sum_n c_n x^n$ be a power series and assume that*

$$\alpha = \lim_n \sqrt[n]{|c_n|}$$

exists and define $R = 1/\alpha$. Then the power series converges if $|x| < R$ and diverges if $|x| > R$.

Proof: Let $a_n = |c_n x^n|$ and applying the root test we get

$$\sqrt[n]{a_n} = \sqrt[n]{|c_n x^n|} = |x| \sqrt[n]{|c_n|}$$

which is less than one if $|x| < R$ and greater than one if $|x| > R$. ∎

In this situation we say that the *radius of convergence* of the power series is R. The interval of convergence includes the interval $(-R, R)$ (it might include or exclude the endpoints).

Notice that the power series converges absolutely in the interval $(-R, R)$ and uniformly on any interval of the form $[-r, r]$ where $0 < r < R$.

Clearly if the power series is centered at x_0 (rather than 0) the interval of convergence is centered at x_0, so is of the form $(x_0 - R, x_0 + R)$.

Problems

1. Suppose that $a_k \geq 0$ and $\{a_k\}$ is a decreasing sequence. Further suppose that
$$\sum_{k=1}^{\infty} a_k$$
converges. Prove that $na_n \to 0$ as $n \to \infty$. (**Hint:** Consider the sum $\sum_{k=n/2}^{n} a_k$)

2. Show that if $\sum_n a_n x^n$ converges for $x = r$ then it converges for all $x \in (-r, r)$ and this convergence is uniform over the interval $[-r, r]$.

3. Suppose that the series $\sum_n a_n x^n$ converges in the interval $[-R, R]$ (and thus converges uniformly on this interval). Show that $f(x) = \sum_n a_n x^n$ is continuous on $[-R, R]$.

4. Suppose that the series $\sum_n a_n x^n$ converges in the interval $[-R, R]$ (and thus converges uniformly on this interval). Show that $f(x) = \sum_n a_n x^n$ is differentiable on $(-R, R)$ and calculate this derivative.

5. Suppose that $f : \mathbb{R} \to \mathbb{R}$ with $f(x) \geq 0$ for all $x \geq 0$ and $f(0) = 0$ and $f'(x)$ is continuous for all x. Finally, suppose that $0 \leq f'(0) < 1$. Prove that if $\{a_k\}$ is a nonnegative sequence with
$$\sum_k a_k$$
convergent then
$$\sum_k f(a_k)$$
is convergent as well.

6. Prove that $\sum_k a_k$ converges absolutely iff $\sum_k \epsilon_k a_k$ converges for any choice of $|\epsilon_k| \leq 1$.

7. Suppose that $\sum_k a_k$ converges with $a_k \geq 0$ for all k. Prove that
$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} k a_k = \lim_{n\to\infty} \sum_{k=1}^{n} \frac{k}{n} a_k = 0.$$

8. Suppose that $a_n > 0$, (a_n) is a decreasing sequence with
$$\frac{1}{a_{n+1}} - \frac{1}{a_n} \geq n.$$
Show that $\sum_n a_n$ converges.

9. Find an example of a real sequence (a_n) with $a_n > 0$ for all n, $\sum_n a_n < \infty$ but $n a_n \not\to 0$. (**Hint:** The sequence cannot be monotone decreasing.)

10. Suppose that $\sum_n a_n = S$ converges absolutely. Let $\pi : \mathbb{N} \to \mathbb{N}$ be a bijection and consider the sequence $b_n = a_{\pi(n)}$ (so that (b_n) is a rearrangement of (a_n)). Show that $\sum_n b_n$ also converges absolutely to S.

11. Suppose that $a_k > 0$ and that $\frac{a_{k+1}}{a_k} \geq 1 - \frac{1}{k}$. Prove that $\sum_k a_k$ diverges.

12. Let $f_n : [a, b] \to \mathbb{R}$ be a sequence of integrable functions. If $|f_n(x)| \leq M_n$ for all $x \in [a, b]$ and $\sum_n M_n < \infty$ show that
$$\int_a^b \sum_n f_n(x)\, dx = \sum_n \int_a^b f_n(x)\, dx$$
(that is, show that the function $\sum_n f_n(x)$ is integrable and that its integral is equal to the sum of the integrals of the individual f_n's).

13. Let $f_n : [a,b] \to I\!R$ be a sequence of integrable functions. Suppose that $f_n(x) \geq 0$ for all $x \in [a,b]$ and $f(x) = \sum_n f_n(x)$ exists as a pointwise limit. Suppose further that $f \in \mathcal{I}[a,b]$. Argue that

$$\int_a^b \sum_n f_n(x)\ dx = \sum_n \int_a^b f_n(x)\ dx.$$

14. Let $f_n : [0,1] \to I\!R$ be defined by

$$f_n(x) = \begin{cases} 0, & \text{if } x = 0 \\ 2^n, & \text{if } 0 < x \leq 1/3^n \\ 0, & \text{if otherwise.} \end{cases}$$

Show that $f(x) = \sum_{n=0}^{\infty} f_n(x)$ converges pointwise but not uniformly. If it exists, compute

$$\int_0^1 f(x)\ dx$$

(simplify your expression as much as you can).

15. Let $f_n : [0,1] \to I\!R$ be defined by

$$f_n(x) = \begin{cases} 0 < & \text{if } x = 0 \\ (-1)^n, & \text{if } 0 < x \leq 1/2^n \\ 0, & \text{if otherwise.} \end{cases}$$

Does $f(x) = \sum_{n=0}^{\infty} f_n(x)$ converge? In which sense does it converge (uniformly or pointwise)? If it makes sense, calculate

$$\int_0^1 f(x)\ dx$$

and justify the existence of this integral.

16. Let $f_n(x) = \alpha_n \sin(\beta_n x)$ where $\sum_n \alpha_n$ converges absolutely. Define $f(x) = \sum_n f_n(x)$. Show that f is defined and continuous on all of $I\!R$. Find conditions on α_n, β_n so that $f'(x)$ exists or does not exist.

Chapter 8

Convergence and Continuity in $I\!R^p$

In this chapter, we start the study of analysis in $I\!R^p$. First we must define convergence of sequences and extend all the results from real sequences to sequences in $I\!R^p$. Then we consider functions $f : I\!R^p \to I\!R^m$ and define limits and continuity for such functions.

This extension is rather straightforward, once the basic framework is set up. In fact, most of the proofs either carry over exactly or with very little change. Even though many of the proofs and results are so similar, we still include them. We do this both to be complete and also to allow one to study this chapter (mostly) independently from the chapter dealing with functions of one real variable.

The similarities between the proofs in the one-dimensional case and the multi-dimensional case hint that there might be some more general theory. This more general theory is examined in courses on topology.

First, we must establish a little notation. If we have a point $x \in I\!R^p$, we know that $x = (x_1, x_2, \ldots, x_p)$. In some instances we will write the ith component of x as x_i. However, when we have a sequence of points in $I\!R^p$, then this subscript notation is confusing as it is unclear if we mean the ith term in the sequence or the ith component. To avoid this confusion, in these cases we use the *projection functions* π_i. We define $\pi_i : I\!R^p \to I\!R$ by

$$\pi_i(x) = x_i \quad \text{where } x = (x_1, x_2, \ldots, x_i, \ldots, x_p).$$

Thus, $\pi_3(x) = 3$ when $x = (1, 2, 3, 4, 5, 6, 7, 8, 9)$.

We use the standard distance between points in $I\!R^p$. Conceptually, we first define the distance from the origin and then use this to define the distance between two arbitrary points. For a given $x \in I\!R^p$ its distance from the origin is given by

$$\|x\| = \left(\sum_{i=1}^{p} x_i^2 \right)^{1/2}.$$

This can be thought of as the "length" of the vector that starts at the origin and ends at x. For two points $x, y \in I\!R^p$ the distance between x and y is given by

$$\|x-y\| = \left(\sum_{i=1}^{p}(x_i-y_i)^2\right)^{1/2} = \left(\sum_{i=1}^{p}(\pi_i(x)-\pi_i(y))^2\right)^{1/2}$$

so we are simply transferring the distance as if one of the points was the origin.

We will use the following properties of this distance extensively.

Proposition 8.1 *Let $x, y \in I\!R^p$ and $\alpha \in I\!R$. Then*

1. $\|x\| \geq 0$ *and* $\|x\| = 0$ *iff* $x = 0$.
2. $\|\alpha x\| = |\alpha|\ \|x\|$.
3. $\|x+y\| \leq \|x\| + \|y\|$.

Limits of sequences in $I\!R^p$

Definition 8.1 *A sequence (x_n) in $I\!R^p$ is* bounded *if there is some $M > 0$ so that $\|x_n\| \leq M$ for all n.*

The proof of the following proposition is straightforward, but we include it to illustrate the technique of going from estimating distance in $I\!R^p$ to estimating the absolute value of the components.

Proposition 8.2 *A sequence (x_n) in $I\!R^p$ is bounded iff there is some $M > 0$ so that $|\pi_i(x_n)| \leq M$ for all i and all n.*

Proof: Suppose that (x_n) is bounded in $I\!R^p$. Then there is some $M > 0$ so that $\|x_n\| \leq M$ for all n. However, then for each i we see that

$$|\pi_i(x_n)| \leq \left(\sum_{i=1}^{p}\pi_i(x_n)^2\right)^{1/2} = \|x_n\| \leq M.$$

Conversely, suppose that there is some $M > 0$ so that $|\pi_i(x_n)| \leq M$ for all i and n. Then we see that

$$\|x_n\| = \left(\sum_{i=1}^{p}\pi_i(x_n)^2\right)^{1/2} \leq \left(\sum_{i=1}^{p}M^2\right)^{1/2} = \sqrt{p}M$$

so that (x_n) is bounded in $I\!R^p$. ∎

Definition 8.2 *Let (x_n) be a sequence in $I\!R^p$ and $L \in I\!R^p$. We say that x_n* converges to L *if for any $\epsilon > 0$ there is an $N \in I\!N$ so that whenever $n \geq N$ we have $\|x_n - L\| < \epsilon$.*

Notice that this is almost identical to the definition of sequential convergence for real sequences. The only difference is the use of the distance $\|\cdot\|$ rather than the absolute value. As an illustration, compare the proof of the following proposition to the proof of Proposition 3.1 in Chapter 3.

Proposition 8.3 *A sequence in $\mathbb{R}^p$ has at most one limit.*

Proof: Suppose that $x_n \to x$ and $x_n \to y$. Let $\epsilon > 0$ be given. Then there exist numbers $N_1, N_2 \in \mathbb{N}$ so that for all $n \geq N_1$ we have that $\|x_n - x\| < \epsilon/2$ and for all $n \geq N_2$ we have that $\|x_n - y\| < \epsilon/2$. Let $N = \max\{N_1, N_2\}$. Then for all $n \geq N$ we know

$$\|x - y\| = \|x - x_n + x_n - y\| \leq \|x - x_n\| + \|y - x_n\| < \epsilon/2 + \epsilon/2 = \epsilon.$$

Since $\|x - y\| < \epsilon$ for any $\epsilon > 0$, it must be the case that $\|x - y\| = 0$ so $x = y$. ∎

We also have the following easy proposition. Since the proof is exactly the same as the proof in $\mathbb{R}$, we leave it out.

Proposition 8.4 *Every convergent sequence is bounded.*

In order to make proofs involving limits in $\mathbb{R}^p$ simpler, we want to relate these limits to limits of sequences in $\mathbb{R}$. The following proposition does this.

Proposition 8.5 *Let (x_n) be a sequence in $\mathbb{R}^p$ and $L \in \mathbb{R}^p$. Then $x_n \to L$ iff for each coordinate i we have that the real sequence $(\pi_i(x_n))$ converges to $\pi_i(L)$.*

Proof: Suppose that $x_n \to L$. Fix some component i and let $\epsilon > 0$ be given. Then since $x_n \to L$ there is some $N \in \mathbb{N}$ so that whenever $n \geq N$ we have that $\|x_n - L\| < \epsilon$. Then for $n \geq N$ we have that

$$|\pi_i(x) - \pi_i(L)| = \sqrt{(\pi_i(x_n) - \pi_i(L))^2} \leq \left(\sum_{i=1}^{p} (\pi_i(x_n) - \pi_i(L))^2\right)^{1/2} < \epsilon$$

so $\pi_i(x) \to \pi_i(L)$.

Conversely, suppose that for all i we have that $\pi_i(x_n) \to \pi_i(L)$. Let $N \in \mathbb{N}$ be so that we have $|\pi_i(x_n) - \pi_i(L)| < \epsilon/\sqrt{p}$ for $n \geq N$ and for all i. Then we see that for $n \geq N$ we have

$$\|x_n - L\| = \left(\sum_{i=1}^{p} (\pi_i(x_n) - \pi_i(L))^2\right)^{1/2} < \left(\sum_{i=1}^{p} (\epsilon/\sqrt{p})^2\right)^{1/2} = \epsilon$$

so $x_n \to L$ as desired. ∎

Theorem 8.6 *Suppose that (x_n) and (y_n) are convergent sequences in $\mathbb{R}^p$ and $\alpha \in \mathbb{R}$. Then*

1. $\lim_{n\to\infty} x_n + y_n = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n$

2. $\lim_{n\to\infty} \alpha x_n = \alpha \lim_{n\to\infty} x_n$

Proof: We simply apply the previous theorem to the corresponding results for real sequences. ∎

Notice that since $\mathbb{R}^p$ has no natural total order we don't have natural extensions of the Squeeze Theorem or the Comparison Theorem. However, see the exercises for versions of them.

Cauchy Sequences in $\mathbb{R}^p$

We want to extend the notion of a Cauchy sequence to sequences in $\mathbb{R}^p$. We will end up proving (see Theorem 8.12) that all Cauchy sequences in $\mathbb{R}^p$ are convergent sequences. With this aim in view, we must first prove extensions of the Nested Cells Property and the Bolzano-Weierstrass Theorem to $\mathbb{R}^p$.

The proofs of most of these results are either simple modifications of the proofs in the one-dimensional case or exactly like the proofs of their one-dimensional cousins.

Definition 8.3 *A* closed cell *in $\mathbb{R}^p$ is a set of the form*

$$C = \{x \in \mathbb{R}^p : a_i \leq \pi_i(x) \leq b_i\}$$

where $a_i, b_i \in \mathbb{R}$ are fixed constants.

Theorem 8.7 *(Nested Cells) Let $C_n \subset \mathbb{R}^p$ be a closed cell for each $n \in \mathbb{N}$ with $C_{n+1} \subset C_n$. Then there is some $x \in \bigcap_n C_n$.*

Proof: By the definition of a closed cell, it is clear that if we fix an i, the sets $I_n = \pi_i(C_n)$ are closed intervals. Furthermore, it is easy to see that $I_{n+1} \subset I_n$ for each n. Thus, by the Nested Cells Theorem for $\mathbb{R}$, we see that there must be some $x_i \in \mathbb{R}$ so that $x_i \in \bigcap_n I_n = \bigcap_n \pi_i(C_n)$. Let $x \in \mathbb{R}^p$ be defined by these coordinates (that is $x = (x_1, x_2, \ldots, x_p)$). Then we see that $\pi_i(x) \in I_n = \pi_i(C_n)$ for all n and i so $x \in C_n$ for all n so $x \in \bigcap_n C_n$, as desired. ∎

We could use the Nested Cells Theorem in $\mathbb{R}^p$ to prove the Bolzano-Weierstrass Theorem (as was done in $\mathbb{R}$). However, it is simpler to appeal directly to the one-dimensional version of the Bolzano-Weierstrass Theorem.

Theorem 8.8 *(Bolzano-Weierstrass) Let (x_n) be a bounded sequence in $\mathbb{R}^p$. Then there is some subsequence (x_{n_k}) which converges.*

Proof: The real sequence $(\pi_1(x_n))$ is a bounded sequence so there must be some convergent subsequence of it. That is, there must be a subsequence (x_{n_k}) of (x_n) so that $\pi_1(x_{n_k}) \to a_1 I\!R$. Now, the sequence (x_{n_k}) is a bounded sequence in $I\!R^p$ so the real sequence $(\pi_2(x_{n_k}))$ is a bounded real sequence, so must have a convergent subsequence $\pi_2(x_{n_{k_l}}) \to a_2$. Notice that $\pi(x_{n_{k_l}}) \to a_1$ and $\pi_2(x_{n_{k_l}}) \to a_2$.

In order to avoid horrible notation, we relabel the subsequence (x_{n_k}) to be (z_k^1) and its subsequence $(x_{n_{k_l}})$ to be (z_l^2).

Continuing in this fashion, we arrive at a subsequence (z_j^p) which is a subsequence of the original sequence for which $\pi_i(z_j^p)$ converges for each i. However, then by Proposition 8.5 we see that this subsequence converges in $I\!R^p$ as well. Thus we have found our desired convergent subsequence of the original sequence (x_n). ▮

Definition 8.4 *A sequence (x_n) in $I\!R^p$ is a* Cauchy sequence *if for every $\epsilon > 0$ there is some $N \in I\!N$ so that whenever $n, m \geq N$ we have $\|x_n - x_m\| < \epsilon$.*

The next three results have proofs which are exactly like their one-dimensional versions, so the proofs are left as simple exercises.

Proposition 8.9 *Every convergent sequence in $I\!R^p$ is also a Cauchy sequence.*

Proposition 8.10 *Every Cauchy sequence in $I\!R^p$ is bounded.*

Proposition 8.11 *If (x_n) is a Cauchy sequence in $I\!R^p$ which has a convergent subsequence then (x_n) converges to the limit of this subsequence.*

We give a proof of the next theorem which uses the corresponding result for real sequences. It is also possible to prove it by the same argument used in the one-dimensional case, but we leave the details to the interested reader (and the problems).

Theorem 8.12 *(Cauchy) If (x_n) is a Cauchy sequence in $I\!R^p$ then (x_n) converges.*

Proof: Since (x_n) is a Cauchy sequence, it is easy to see that $(\pi_i(x_n))$ is Cauchy for each fixed i. However, then $\pi_i(x_n) \to z_i$ for some $z_i \in I\!R$. But then $x_n \to z$ where $z = (z_1, z_2, \ldots, z_p)$. ▮

Functions $f : I\!R^p \to I\!R^q$

Now we turn our attention to the study of continuity properties of functions $f : I\!R^p \to I\!R^q$. In order to do this we start with a couple of preliminary definitions.

Definition 8.5 *Let $x \in I\!R^p$ and $\epsilon > 0$ be given. The* open ball of radius ϵ centered at x *is the set*

$$B_\epsilon(x) = \{z \in I\!R^p : \|z - x\| < \epsilon\}.$$

Definition 8.6 *Let $E \subset I\!R^p$ and $x \in E$. We say that x is in the* interior *of E if there is some $\epsilon > 0$ so that $B_\epsilon(x) \subset E$.*

Definition 8.7 *Let $E \subset I\!R^p$ and a be in the interior of the set E. Furthermore, let f be a function defined on $E \setminus \{a\}$ with values in $I\!R^q$ and $L \in I\!R^q$. We say that L* is the limit of f as x approaches a, *written as $\lim_{x\to a} f(x) = L$ if for every $\epsilon > 0$ there is a $\delta > 0$ so that whenever $0 < \|x - a\| < \delta$ and $x \in E$ we have that $\|f(x) - L\| < \epsilon$.*

The following theorem allows us to reduce questions involving limits of functions to the corresponding question involving limits of sequences. Its proof is virtually identical to the proof of Theorem 4.1 in Chapter 4, but we include it to illustrate this.

Theorem 8.13 *Let $E \subset I\!R^p$ and a be in the interior of E and f be a function defined on $E\setminus\{a\}$ with values in $I\!R^q$. Then $\lim_{x\to a} f(x) = L$ iff for all sequences (x_n) in $E \setminus \{a\}$ with $x_n \to a$ we have $f(x_n) \to L$.*

Proof: Suppose that $\lim_{x\to a} f(x) = L$ and let $x_n \to a$ with $x_n \in I \setminus \{a\}$. Let $\epsilon > 0$ be given. Then since $\lim_{x\to a} f(x) = L$ there is a $\delta > 0$ so that whenever $0 < \|x - a\| < \delta$ and $x \in E$ we have $\|f(x) - L\| < \epsilon$. Now since $x_n \to a$ there is an $N \in I\!N$ so whenever $n \geq N$ we have $0 < \|x_n - a\| < \delta$ which implies that $\|f(x_n) - L\| < \epsilon$. Thus $f(x_n) \to L$ as desired.

Conversely, suppose that for any (x_n) with $x_n \to a$ and $x_n \in E\setminus\{a\}$ we know $f(x_n) \to L$. Suppose that $\lim_{x\to a} f(x) \neq L$. Then there is an $\epsilon > 0$ so that for all $\delta > 0$ there is some $x \in E \setminus \{a\}$ with $0 < \|x - a\| < \delta$ but $\|f(x) - L\| \geq \epsilon$. This means that for each $n \in I\!N$ we have an $x_n \in E$ with $0 < \|x_n - a\| < 1/n$ but $\|f(x_n) - L\| \geq \epsilon$. However, then $x_n \to a$ but $f(x_n) \not\to a$, which is a contradiction. ∎

Using this theorem, we can easily prove the relationship between basic arithmetic in $I\!R^p$ and limits. We leave the simple proof to the interested reader.

Theorem 8.14 *Let $E \subset I\!R^p$ and a be in the interior of E and f, g be functions defined on $E \setminus \{a\}$ with values in $I\!R^q$. Suppose further that $\lim_{x\to a} f(x)$ and $\lim_{x\to a} g(x)$ both exist. Then*

1. $\lim_{x\to a}(f(x) + g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$
2. $\lim_{x\to a} \alpha f(x) = \alpha \lim_{x\to a} f(x)$ *for all $\alpha \in I\!R$.*

Compare the following definition with the definition of continuity for a real-valued function of a real variable; it is (again) virtually identical.

Definition 8.8 *Let $\emptyset \neq E \subset I\!R^p$ and $f : E \to I\!R^q$.*

1. *f is said to be* continuous at the point $x_0 \in E$ *if for every $\epsilon > 0$ there is a $\delta > 0$ so whenever $x \in E$ and $\|x - x_0\| < \delta$ we have that $\|f(x) - f(x_0)\| < \epsilon$.*
2. *f is said to be* continuous on a set $A \subset E$ *if it is continuous at each point of A.*
3. *f is said to be* continuous *if it is continuous on its domain.*

Theorem 8.15 *Let $E \subset I\!R^p$ and $x_0 \in E$ and $f : E \to I\!R^q$. Then f is continuous at x_0 iff for every sequence $x_n \to x_0$ with $x_n \in E$ we have $f(x_n) \to f(x_0)$.*

Proof: Suppose that f is continuous at x_0 and $x_n \in E$ converges to x_0. Let $\epsilon > 0$ be given. Then since f is continuous at x_0, there is a $\delta > 0$ so that whenever $\|x - x_0\| < \delta$ and $x \in E$ we have $\|f(x) - f(x_0)\| < \epsilon$. However, since $x_n \to x_0$ there is an $N \in I\!N$ so that for all $n \geq N$ we have $\|x_n - x_0\| < \delta$ which implies that $\|f(x_n) - f(x_0)\| < \epsilon$. Thus $f(x_n) \to f(x_0)$, as desired.

Conversely, suppose $f(x_n) \to f(x_0)$ for all sequences $x_n \in E$ with $x_n \to x_0$. Suppose further that f is not continuous at x_0. Then there is some $\epsilon > 0$ so for all $\delta > 0$ there is some $x \in E$ with $\|x - x_0\| < \delta$ but $\|f(x) - f(x_0)\| \geq \epsilon$. Thus for each $n \in I\!N$ there is some $x_n \in E$ with $\|x_n - x_0\| < 1/n$ but $\|f(x_n) - f(x_0)\| \geq \epsilon$. However, this means that we have a sequence $x_n \to x_0$ with $f(x_n) \not\to f(x_0)$, which is a contradiction. ∎

These next two theorems have routine proofs, so we leave the proofs out. The proofs can easily be constructed by using the sequential characterization of continuity and the corresponding result for limits of sequences.

Theorem 8.16 *Let $E \subset I\!R^p$ and $f, g : E \to I\!R$ and $x_0 \in E$. Suppose that f and g are both continuous at x_0. Then so are $f + g$, fg, αf (for any $\alpha \in I\!R$) and f/g (as long as $g(x_0) \neq 0$).*

Theorem 8.17 *Let $E \subset I\!R^p$ and $f, g : E \to I\!R^q$ and $x_0 \in E$. If f and g are continuous at x_0 then so are $f + g$ and αf for any $\alpha \in I\!R$.*

We provide the proof of the following theorem to show how nice the sequential characterization of continuity is.

Theorem 8.18 *Let $E \subset I\!R^p$ and $x_0 \in E$ and $f : E \to I\!R^q$ be continuous at x_0. Let $f(E) \subset F \subset I\!R^q$ and $g : F \to I\!R^s$ be continuous at $f(x_0)$. Then the function $g \circ f : E \to I\!R^s$ is continuous at x_0.*

Proof: We must show that anytime $x_n \to x_0$ with $x_n \in E$ then $g(f(x_n)) \to g(f(x_0))$. To this end, let $\epsilon > 0$ be given. Then since g is continuous at $f(x_0)$ there is an $\alpha > 0$ so that whenever $y \in F$ and $\|y - f(x_0)\| < \alpha$ we have $\|g(y) - g(f(x_0))\| < \epsilon$. Now since f is continuous at x_0 we know that there is some $\delta > 0$ so that whenever $x \in E$ with $\|x - x_0\| < \delta$ we have $\|f(x) - f(x_0))\| < \alpha$. Finally, since $x_n \to x_0$ we know that there is an $N \in I\!N$ so whenever $n \geq N$ we have that $\|x_n - x_0\| < \delta$. Thus, for $n \geq N$ we have $\|x_n - x_0\| < \delta$ so $\|f(x_n) - f(x_0)\| < \alpha$ so $\|g(f(x_n)) - g(f(x_0))\| < \epsilon$ so $g(f(x_n)) \to g(f(x_0))$ as desired. ∎

The next theorem states that a function $f : I\!R^p \to I\!R^q$ is continuous iff each of its component functions is continuous. That is, if we write f as

$$f(x) = (f_1(x), f_2(x), \ldots, f_q(x))$$

then f is continuous iff f_i is continuous for each i.

Theorem 8.19 *Let $E \subset I\!R^p$ and $x_0 \in E$ and $f : E \to I\!R^q$. Then f is continuous if $\pi_i \circ f : E \to I\!R$ is continuous for each i.*

Proof: Suppose that f is continuous. Then we see that π_i is clearly continuous so the composition $\pi_i \circ f$ is continuous.

Conversely, suppose that $\pi_i \circ f$ is continuous for each i. Let $x_n \in E$ converge to x_0. Then we know that $\pi_i(f(x_n)) \to \pi_i(f(x_0))$ for each i so we see that $f(x_n) \to f(x_0)$ and thus f is continuous at x_0. ∎

The next theorem is a version of the Extreme Value Theorem for functions with domain in $I\!R^p$. The proof is identical to the proof of the one-dimensional version.

Theorem 8.20 *Let $C \subset I\!R^p$ be a non-empty closed cell and $f : C \to I\!R$ be continuous. Then there are points $x_*, x^* \in C$ with $f(x_*) \leq f(x) \leq f(x^*)$ for all $x \in C$.*

Proof: Suppose that f is not bounded on C. Then there is some sequence $x_n \in C$ with $f(x_n) \geq n$ for all $n \in I\!N$. Since $x_n \in C$, it is a bounded sequence so by the Bolzano-Weierstrass Theorem there is a subsequence $x_{n_k} \to x \in C$. Since f is continuous on C we have $f(x_{n_k}) \to f(x)$, however $f(x_{n_k}) \geq n_k$ so $f(x_{n_k}) \to \infty$, which contradicts the fact that $f(x_{n_k}) \to f(x)$. Thus f must be bounded on C.

This means that $M = \sup_{x \in C} f(x)$ and $m = \inf_{x \in C} f(x)$ both exist.

Suppose that there is no $x_* \in C$ with $f(x_*) = m$. Then $f(x) > m$ for all $x \in C$ so the function $g : C \to I\!R$ defined by $g(x) = 1/(f(x) - m)$ is well-defined (since $f(x) \neq m$) and continuous on C. But this means that g is bounded on C so $|g(x)| = g(x) \leq c$ for some $c > 0$. Rewriting this we get $f(x) \geq m + 1/c$ which contradicts $m = \inf_{x \in C} f(x)$. Thus there must be some $x_* \in C$ with $f(x_*) = m$.

A similar argument shows that there is an $x^* \in C$ with $f(x^*) = M$. ∎

Lemma 8.21 *(Sign Preserving Property) Suppose $E \subset I\!R^p$ and $f : E \to I\!R$ is continuous at $x_0 \in E$ and $f(x_0) > 0$. Then there are positive numbers $\delta, \epsilon > 0$ with $\|x - x_0\| < \delta$ implying $f(x) > \epsilon$.*

Proof: The proof is identical to the one-dimensional version, so is left out. ∎

Connectedness and the Intermediate Value Theorem

The Intermediate Value Theorem has, as its basis, the fact that a continuous function preserves connected sets. That is, if f is continuous and I is a connected set, then $f(I)$ should also be a connected set. We take this view here in extending the Intermediate Value Theorem to multi-dimensions.

Definition 8.9 *Let $E \subset I\!R^p$ and $x, y \in E$. A* path *in E from x to y is a continuous function $\sigma : [0, 1] \to E$ with $\sigma(0) = x$ and $\sigma(1) = y$.*

Definition 8.10 *A set $C \subset I\!R^p$ is a* path connected set *if for each $x, y \in C$ there is a path σ from x to y.*

Proposition 8.22 *Any closed cell $C \subset I\!R^p$ is path connected.*

Proof: See the problems.

Proposition 8.23 *The only path connected subsets of $I\!R$ are intervals.*

Proof: See the problems.

Theorem 8.24 *(Preservation of path connectedness) Let $E \subset I\!R^p$ and $f : E \to I\!R^q$ be continuous. If $A \subset E$ is path connected, then so is $f(A)$.*

Proof: Let $y_0, y_1 \in f(A)$. Then there are points $x_0, x_1 \in A$ so that $f(x_0) = y_0$ and $f(x_1) = y_1$. Since A is path connected, there is a path σ in A from x_0 to x_1. However, then $f \circ \sigma : [0, 1] \to f(A)$ is a path in $f(A)$ from y_0 to y_1. Thus, $f(A)$ is path connected. ∎

Corollary 8.25 *(Intermediate Value Theorem) Let C be a closed cell in $I\!R^p$ and $f : C \to I\!R$ be continuous. Let $a, b \in C$ with $f(a) \le f(b)$. Then for any y_0 with $f(a) \le y_o \le f(b)$ there is an $x_0 \in C$ with $f(x_0) = y_0$.*

Proof: Since C is path connected, $f(C) \subset I\!R$ must be path connected as well, so $f(C)$ must be an interval. Thus, $y_0 \in [f(a), f(b)] \subset f(C)$ which implies that there is some $x_0 \in C$ with $f(x_0) = y_0$. ∎

Uniform Continuity

Definition 8.11 *Let $E \subset I\!R^p$ and $f : E \to I\!R^q$. We say that f is* uniformly continuous on E *if for every $\epsilon > 0$ there is a $\delta > 0$ so that of $\|x - y\| < \delta$ and $x, y \in E$ then $\|f(x) - f(y)\| < \epsilon$.*

Theorem 8.26 *Let $C \subset I\!R^p$ be a closed cell and $f : C \to I\!R^q$ be continuous. Then f is uniformly continuous.*

Proof: Suppose that f is not uniformly continuous. Then there is an $\epsilon > 0$ so that for each $n \in I\!N$ there are two points $x_n, y_n \in C$ so that $\|f(x_n) - f(y_n)\| > \epsilon$ and $\|x_n - y_n\| < 1/n$. However, by the Bolzano-Weierstrass Theorem there are subsequences x_{n_k} and y_{n_k} which both converge. Since $\|x_{n_k} - y_{n_k}\| < 1/n_k$ we know that both of these subsequences converge to the same limit, call it z. Now, f is continuous so that $f(x_{n_k}) \to f(z)$ and $f(y_{n_k}) \to f(z)$. However, $\|f(x_{n_k}) - f(y_{n_k})\| > \epsilon$, which is a contradiction. Thus it must be the case that f is uniformly continuous.∎

Sequentially compact sets in $I\!R^p$

In this section we extend the notions and results of compact sets to $I\!R^p$. Since the underlying structure is the same, the proofs are simple modifications of the proofs in the one-dimensional case.

Definition 8.12 *A set $E \subset I\!R^p$ is said to be* sequentially compact *if all sequences (x_n) in E have a convergent subsequence which converges to an element of E.*

Usually we will simply call such sets *compact*.

By the Bolzano-Weierstrass Theorem all closed cells in $I\!R^p$ are sequentially compact. The following two properties of compact sets are easy to prove.

Proposition 8.27

1. *Every finite union of compact sets is compact.*
2. *The arbitrary intersection of compact sets is compact.*

Definition 8.13 *We say that a set $O \subset I\!R^p$ is* open *if for every $x \in O$ there is an $\epsilon > 0$ so that $B_\epsilon(x) \subset O$. We say that $K \subset I\!R^p$ is closed if $I\!R^p \setminus K$ is open.*

Theorem 8.28 *A subset $K \subset I\!R^p$ is closed iff for every convergent sequence (x_n) in K we have that the limit point of the sequence is also in K.*

Proof: Suppose that K is closed and that $x_n \in K$ with $x_n \to K$. Suppose that $x \notin K$. Then since K is closed, $I\!R^p \setminus K$ is open, so there is some $\epsilon > 0$ so that $B_\epsilon(x) \subset I\!R^p \setminus K$. Now since $x_n \to x$ there is some $N \in I\!N$ so that for all $n \geq N$ we have that $\|x - x_n\| < \epsilon$ which implies that $x_n \in B_\epsilon(x)$ so that $x_n \notin K$, which is a contradiction. Thus $x \in K$.

Conversely, suppose that for all sequences $x_n \in K$ with $x_n \to x$ we have that $x \in K$ as well. We wish to prove that K is closed. Suppose that K is not closed, then $\mathbb{R}^p \setminus K$ is not open so there is some $x \notin K$ so that for all $\epsilon > 0$ we have that $B_\epsilon(x) \cap K \neq \emptyset$. Thus for all $n \in \mathbb{N}$ there is some $x_n \in K$ with $\|x_n - x\| < 1/n$. But then we have a sequence (x_n) with $x_n \in K$ and $x_n \to x$ but $x \notin K$. ∎

Theorem 8.29 *(Heine-Borel) A subset $K \subset \mathbb{R}^p$ is compact iff it is closed and bounded.*

Proof: Suppose that K is compact. We wish to prove that K is closed and bounded. First suppose that K is not bounded. Then there is some sequence $x_n \in K$ with $\|x_n\| \geq n$ for each n. But then the sequence (x_n) cannot have a Cauchy subsequence, so it cannot have a convergent subsequence. Thus K must be bounded.

To show that K is closed, let (x_n) be a convergent sequence in K with limit $x \in \mathbb{R}^p$. Then since K is compact, there is a subseqence (x_{n_k}) of the sequence (x_n) which converges with $x_{n_k} \to y \in K$. However, since $x_n \to x$ we know that $x_{n_k} \to x$ as well, so $x = y \in K$, so K is closed.

Conversely, suppose that K is closed and bounded. Then there is some $M > 0$ so that $K \subset B_M(0) = \{x \in \mathbb{R}^p : \|x\| \leq M\}$. Let (x_n) be a sequence in K. Then (x_n) is a bounded sequence, so by the Bolzano-Weierstrass Theorem there is a subsequence (x_{n_k}) and a point $x \in \mathbb{R}^p$ so that $x_{n_k} \to x$. But since K is closed, we must have that $x \in K$. ∎

Theorem 8.30 *Suppose that $f : K \to \mathbb{R}^q$ is continuous and K is compact. Then f is uniformly continuous on K.*

Proof: Suppose that f is not uniformly continuous on K. Then there is an $\epsilon > 0$ so that for any $\delta > 0$ we have points $x, y \in K$ so that $\|x - y\| < \delta$ but $\|f(x) - f(y)\| \geq \epsilon$. Thus for all $n \in \mathbb{N}$ there are points $x_n, y_n \in K$ with $\|x_n - y_n\| < 1/n$ but $\|f(x_n) - f(y_n)\| \geq \epsilon$.

Now since K is compact there is some subsequence (x_{n_k}) with $x_{n_k} \to x$. However, then there is a subsequence of (y_{n_k}) which converges. Call this subsequence $(y_{n_{k_l}})$ so that $y_{n_{k_l}} \to y \in K$. Thus we have that $x_{n_{k_l}} \to x$ and $y_{n_{k_l}} \to y$. Now by our choice of x_n and y_n we know that

$$\|x_{n_{k_l}} - y_{n_{k_l}}\| < 1/n_{k_l} \qquad \Rightarrow \qquad x = y.$$

Finally, since f is continuous, we see that $f(x_{n_{k_l}}) \to f(x)$ and $f(y_{n_{k_l}}) \to f(y)$ with $f(x) = f(y)$. However, this contradicts the fact that $\|f(x_{n_{k_l}}) - f(y_{n_{k_l}})\| \geq \epsilon$. ∎

There is a version of the Extreme Value Theorem valid for compact subsets of $\mathbb{R}^p$ as well. We give this next.

Theorem 8.31 *Suppose that $f : K \to \mathbb{R}$ is continuous and $K \subset \mathbb{R}^p$ is compact. Then there are points $x_*, x^* \in K$ so that $m = f(x_*) \leq f(x) \leq f(x^*) = M$ for all $x \in K$.*

Proof: First we show that f is bounded on K. Suppose it is not bounded. Then for every $n \in \mathbb{N}$ there is a point $x_n \in K$ so that $\|f(x_n)\| \geq n$. However, since K is compact there is a subsequence (x_{n_k}) of (x_n) so that $x_{n_k} \to x \in K$. But then since f is continuous we know that $f(x_{n_k}) \to f(x)$, which contradicts the fact that $\|f(x_{n_k})\| \to \infty$. Thus f is bounded.

Since f is bounded both $m = \inf_{x\in K} f(x)$ and $M = \sup_{x\in K} f(x)$ exist. Suppose that there is no point $x^* \in K$ so that $f(x^*) = M$. Then we have that $f(x) < M$ for all $x \in K$ so the function $g : K \to \mathbb{R}$ defined by $g(x) = 1/(M - f(x))$ is well-defined on K and continuous. But then we know that g is bounded on K, so there is a number $C > 0$ so that $g(x) < C$ for all $x \in K$. Rearranging this, we see that this implies that $f(x) < M - 1/C$ for all $x \in K$, which contradicts the fact that $M = \sup_{x\in K} f(x)$. ∎

Theorem 8.32 *Let $f : K \to \mathbb{R}^q$ be continuous where $K \subset \mathbb{R}^p$ is a compact set. Suppose that f is a bijection. Then $f^{-1} : f(K) \to K$ is also continuous.*

Proof: It is possible to show that $f(K)$ is also compact (see the problems). To show that f^{-1} is continuous, let (y_n) be a convergent sequence in $f(K)$ with $f(x_n) = y_n \to y$. Then since $f(K)$ is compact, we know that there is some $x \in K$ with $f(x) = y$. We wish to prove that $x_n \to x$. Suppose that $x_n \not\to x$. Then there is an $\epsilon > 0$ and a subsequence (x_{n_k}) of (x_n) for which $\|x_{n_k} - x\| \geq \epsilon$. Since K is compact, there must be some subsequence $(x_{n_{k_l}})$ of (x_{n_k}) which converges, say $x_{n_{k_l}} \to z \in K$. But then since f is continuous we know that $y_{n_{k_l}} = f(x_{n_{k_l}}) \to f(z)$. However, $y_n \to y$, which means that $f(z) = y$ as well. Since f is a bijection, we have that $z = x$. But this contradicts the fact that $\|x_{n_{k_l}} - x\| \geq \epsilon$. Thus we must have that $x_n \to x$ so f^{-1} is continuous. ∎

Sequences of functions

We now turn to the discussion of the convergence of sequences of functions.

Definition 8.14 *Let $E \subset \mathbb{R}^p$ and $f_n : E \to \mathbb{R}^q$ be a sequence of functions. We say that f_n* converges pointwise *on E if for every $x \in E$ we have that $\lim_{n\to\infty} f_n(x)$ exists.*

Notice that if f_n converges pointwise on E, we can define a function $f : E \to \mathbb{R}^q$ by $f(x) = \lim_{n\to\infty} f_n(x)$. In this case, we say that $f_n \to f$ pointwise on E.

As is the case for real-valued functions of one variable, the functions f_n can all be continuous without the limit function being continuous.

Definition 8.15 *Let $f_n : A \to I\!R^q$ be a sequence of functions where $A \subset I\!R^p$. We say that $f_n \to f$ (a function on A)* uniformly *if for all $\epsilon > 0$ there is an $N \in I\!N$ so that for all $n \geq N$ and all $x \in A$ we have $\|f_n(x) - f(x)\| < \epsilon$.*

How is this different from pointwise convergence? We can write pointwise convergence as: *$f_n \to f$ pointwise if for each $x \in A$ and $\epsilon > 0$ there is an $N \in I\!N$ so that for all $n \geq N$ we have $\|f_n(x) - f(x)\| < \epsilon$.*

Thus, for pointwise convergence, the N can depend on both x and ϵ while for uniform convergence the N depends on only ϵ (and NOT on x).

The proof of the next proposition is left as an exercise.

Proposition 8.33 *Suppose that $f_n \to f$ uniformly. Then $f_n \to f$ pointwise as well.*

Proposition 8.34 *(Cauchy criterion for uniform convergence) The sequence of functions $f_n : A \to I\!R^q$ converges uniformly iff for all $\epsilon > 0$ there is an $N \in I\!N$ so for any two $n, m \geq N$ we have $\|f_n(x) - f_m(x)\| < \epsilon$ for all $x \in A$.*

Proof: Suppose that $f_n \to f$ uniformly on A and let $\epsilon > 0$ be given. Then there is an $N \in I\!N$ so that for all $n \geq N$ we have $\|f_n(x) - f(x)\| < \epsilon/2$ for all $x \in A$. Thus for all $n, m \geq N$ we have

$$\|f_n(x) - f_m(x)\| \leq \|f_n(x) - f(x)\| + \|f(x) - f_m(x)\| < \epsilon.$$

Conversely, suppose that for all $\epsilon > 0$ there is an $N \in I\!N$ so for all $n, m \geq N$ we have $\|f_n(x) - f_m(x)\| < \epsilon$ for all $x \in A$. This means that for each $x \in A$ the sequence $(f_n(x))$ is a Cauchy sequence and thus converges to some point in $I\!R^q$, call it $f(x)$. Now, given an $\epsilon > 0$ we know that there is an $N \in I\!N$ so whenever $n, m \geq N$ we have that $\|f_n(x) - f_m(x)\| < \epsilon/2$ for all $x \in A$. If we take the limit of this inequality as $m \to \infty$ we get $\|f_n(x) - f(x)\| \leq \epsilon/2$ for all $x \in A$ and thus $f_n \to f$ uniformly on A. ∎

Theorem 8.35 *Suppose $f_n \to f$ pointwise on A. Then $f_n \to f$ uniformly on A iff*

$$\sup_{x \in A} \|f_n(x) - f(x)\| \to 0 \quad \text{as} \quad n \to \infty$$

Continuity and uniform limits

Theorem 8.36 *Let $A \subset I\!R^p$ and $f_n : A \to I\!R^q$ converge uniformly to $f : A \to I\!R^q$. Suppose that $z \in A$ and for each n we have $\lim_{x \to z} f_n(x) = L_n$ exists. Then (L_n) converges and $\lim_{x \to z} f(x) = \lim_{n \to \infty} L_n$.*

Proof: Let $\epsilon > 0$ be given. Since (f_n) converges uniformly, there is an $N \in I\!N$ so for all $n, m \geq N$ we have $\|f_n(x) - f_m(x)\| < \epsilon$. Taking limits as $x \to z$ we have that $\|L_n - L_m\| < \epsilon$ for all $n, m \geq N$. Thus, (L_n) is a Cauchy sequence so converges, say to L.

Now,

$$\|f(x) - L\| \leq \|f(x) - f_n(x)\| + \|f_n(x) - L_n\| + \|L_n - L\|.$$

We first choose $N \in I\!N$ so that for any $n, m \geq N$ we have $\|f_n(x) - f(x)\| < \epsilon/3$ for all $x \in A$ and so that $\|L_n - L\| < \epsilon/3$. Now fix an $n \geq N$. Then choose $\delta > 0$ so that if $\|x - z\| < \delta$ we have $\|f_n(x) - L_n\| < \epsilon/3$. Thus we have $\|f(x) - L\| < \epsilon$ for $\|x - z\| < \delta$, so $\lim_{x \to z} f(x) = L = \lim_{n \to \infty} f_n(z)$. ▮

Corollary 8.37 *Let $f_n : A \to I\!R^q$ converges uniformly to $f : A \to I\!R^q$. If each f_n is continuous, then so is f.*

Problems

1. Let $1 < r < s < \infty$ and $x \in I\!R^n$. Show that
$$\|x\|_s \leq \|x\|_r \leq n^{\frac{s-r}{rs}} \|x\|_s$$
where
$$\|x\|_s = \left(\sum_{i=1}^{n} |x_i|^s \right)^{1/s}.$$

2. Let $g_i > 0$ for $i = 1, 2, \ldots, p$ and define the norm $\|x\|_g$ on $I\!R^p$ by
$$\|x\|_g = \sum_{i=1}^{p} g_i |\pi_i(x)|$$
(that is, the sum of the absolute values of the components of x multiplied by the corresponding g_i). Show that for any sequence $x_n \in I\!R^p$ we have that $\|x_n\|_g \to 0$ iff $\pi_i(x_n) \to 0$ for all i.

3. Suppose that $x_n \to x$ and $y_n \to y$ in $I\!R^p$. Suppose further that there is some $N \in I\!N$ so that for each i and $n \geq N$ we have $\pi_i(x_n) \leq \pi_i(y_n)$. Show that $\pi_i(x) \leq \pi_i(y)$ for each i.

4. Suppose that $x_n \to 0$ with $x_n \in I\!R$ and (y_n) is a bounded sequence in $I\!R^p$. Prove that $x_n y_n \to 0$.

5. Suppose that $x_n \to x$ and $z_n \to x$ in $I\!R^p$ and that (y_n) is a sequence in $I\!R^p$ with $\pi_i(x_n) \leq \pi_i(y_n) \leq \pi_i(z_n)$ for each i and $n \geq N$. Prove that $y_n \to x$ as well.

6. Suppose that $x_n \to x$ and $y_n \to y$ in $I\!R^p$. Prove that $x_n \cdot y_n \to x \cdot y$, where $x_n \cdot y_n$ is the dot product of x_n and y_n.

7. Prove Proposition 8.9.
8. Prove Proposition 8.10.
9. Prove Proposition 8.11.
10. Prove Theorem 8.12. Do this by mimicking the proof of the one-dimensional version.
11. Suppose that $f, g : E \to \mathbb{R}^q$ are continuous with $\emptyset \neq E \subset \mathbb{R}^p$. Define the function $h : E \to \mathbb{R}^q$ by $h(x) = f(x) \cdot g(x)$ (that is, the dot product of $f(x)$ and $g(x)$). Prove that h is continuous.
12. Prove that any closed cell in $\mathbb{R}^p$ is path connected.
13. Prove the Sign Preserving Property.
14. Prove that the only path connected subsets of $\mathbb{R}$ are intervals (either open, half-open or closed).
15. Prove that if $A \subset \mathbb{R}^p$ is path connected and $f : A \to \mathbb{R}$ is continuous then for all $a, b \in A$ and $y_0 \in \mathbb{R}$ with $f(a) \leq y_0 \leq f(b)$ there is some $x_0 \in A$ with $f(x_0) = y_0$.
16. Let $K \subset \mathbb{R}^p$ be compact and $\epsilon > 0$. Show that there are finitely many points x_i, $i = 1, 2, \ldots, N$ with

$$K \subset \bigcup_{i=1}^{N} B_\epsilon(x_i).$$

    (That is, that each point in K is within ϵ of one of the points x_i).
17. Let $K_n \subset \mathbb{R}^p$ be sequentially compact for each $n \in \mathbb{N}$. Furthermore, suppose that $K_{n+1} \subset K_n$. Show that $\cap_n K_n$ is not empty.
18. Suppose that $K \subset \mathbb{R}^p$ is sequentially compact and $f : K \to \mathbb{R}^q$ is continuous. Show that $f(K)$ is sequentially compact.
19. Prove that the function

$$f(x, y) = \sin(\frac{2xy}{x^2 + y^2})$$

    has no limit as $(x, y) \to (0, 0)$.
20. Suppose that $B \subset \mathbb{R}^p$ and $a \notin B$.
    (a) If B is open, is there necessarily a point $b \in B$ so that $\|a-b\| \leq \|a-x\|$ for all $x \in B$?
    (b) If B is closed, is there necessarily a point $b \in B$ so that $\|a - b\| \leq \|a - x\|$ for all $x \in B$?

(c) If B is compact, is there necessarily a point $b \in B$ so that $\|a-b\| \leq \|a-x\|$ for all $x \in B$?

21. In this problem, we provide an alternate proof of the extreme value theorem for functions $f : \mathbb{R}^p \to \mathbb{R}$.

    (a) Let $K \subset \mathbb{R}$ be compact. Show that there are points $x_*, x^* \in K$ with $x_* \leq x \leq x^*$ for all $x \in K$.

    (b) Let $f : \mathbb{R}^p \to \mathbb{R}^q$ be continuous and $K \subset \mathbb{R}^p$ be compact. Show that $f(K)$ is also compact.

    (c) Let $f : K \subset \mathbb{R}^p \to \mathbb{R}$ be continuous with K compact. Show that there are points $x_*, x^* \in K$ with $f(x_*) \leq f(x) \leq f(x^*)$ for all $x \in K$.

Chapter 9

Derivatives for multivariate functions

Now we turn to the study of differentiation of multi-variate functions.

We first need to have a brief review of the basic ideas of linear functions in $I\!R^p$.

Definition 9.1 *A function $L : I\!R^p \to I\!R^q$ is* linear *if for all $x, y \in I\!R^p$ and $\alpha \in I\!R$ we have*

1. $L(x_1 + y_1, x_2 + y_2, \ldots, x_p + y_p) = L(x_1, x_2, \ldots, x_p) + L(y_1, y_2, \ldots, y_p)$.
2. $L(\alpha x_1, \alpha x_2, \ldots, \alpha x_p) = \alpha L(x_1, x_2, \ldots, x_p)$.

Given a linear function $L : I\!R^p \to I\!R^q$, there is a unique $q \times p$ matrix A with the property that $L(x) = Ax$, where we think of x as being a column vector. That is,

$$L(x_1, x_2, \ldots, x_p) = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,p} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{q,1} & a_{q,2} & \cdots & a_{q,p} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_p \end{pmatrix}.$$

We denote the collection of all linear functions from $I\!R^p$ to $I\!R^q$ by $Lin(I\!R^p, I\!R^q)$. Notice that $Lin(I\!R^p, I\!R^q)$ is also a vector space, as we can add two linear functions to get another linear function and a multiple of a linear function is also a linear function.

Proposition 9.1 *Any linear function $L : I\!R^p \to I\!R^q$ is continuous.*

Proof: Let e_i for $i = 1, 2, \ldots, p$ be the standard basis vectors in $I\!R^p$ and let $M = \max_i \|L(e_i)\|$. If $u = (u_1, u_2, \ldots, u_p) \in I\!R^p$ has unit length then

$|u_i| \leq \|u\| = 1$ for all i. Thus $\|L(u)\| \leq pM$ for all unit vectors. However, this means that $\|L(x)\| \leq pM\|x\|$ for all vectors x.

Let $x \in I\!R^p$ and $\epsilon > 0$ be given. Then if $\|y - x\| < \epsilon/(pM)$ we have that

$$\|L(y) - L(x)\| = \|L(y - x)\| \leq pM\|y - x\| < pM(\epsilon/(pM)) = \epsilon.$$

∎

We will also need to consider functions whose output consists of linear functions. For example, consider the function $R : [0, 2\pi] \to Lin(I\!R^2, I\!R^2)$ given by

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Clearly for each θ we see that $R(\theta)$ is a linear function (a rotation by angle θ about the origin in $I\!R^2$).

Definition 9.2 *Let $E \subset I\!R^p$ with x a point in the interior of E and let $f : E \to I\!R^q$. We say that f* is differentiable *at x if there is a linear function $L \in Lin(I\!R^p, I\!R^q)$ so that*

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

The linear function L is the derivative of f at x *and is denoted by $Df(x)$.*

It is helpful to consider an example of $f : I\!R \to I\!R$ to understand this definition. Suppose that $f(x) = x^2$ and we are interested in the derivative at the point $x = 2$. We know that all linear functions $L : I\!R \to I\!R$ are of the form $L(v) = av$ for some $a \in I\!R$. So, we want to find an $a \in I\!R$ so that

$$0 = \lim_{\epsilon \to 0} \frac{|f(2 + \epsilon) - f(2) - a(\epsilon)|}{|\epsilon|} = \lim_{\epsilon \to 0} \frac{|4 + 4\epsilon + \epsilon^2 - 4 - a\epsilon|}{|\epsilon|}$$

so clearly $a = 4$ will work. That is, the slope of the tangent line (the one-variable way of describing the derivative) gives a description of the linear function. Notice that we are approximating $f(x + \epsilon)$ by $f(x) + a\epsilon$.

What we are doing is trying to find the linear function $L \in Lin(I\!R^p, I\!R^q)$ so that $f(x + h) \approx f(x) + L(h)$.

Another point to notice is that the derivative of f at x depends *both* on f and x. If we take the same function f and find its derivative at some other point y, we are likely to get a different linear function.

We will prove a theorem which shows how to compute the derivative of a large class of functions $f : I\!R^p \to I\!R^q$.

Theorem 9.2 *If f has a derivative at x, then it is continuous at x.*

Proof: Suppose that f has a derivative L at x. Then we know that

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

Thus

$$\|f(x+h)-f(x)\| = \|f(x+h)-f(x)-L(h)+L(h)\| \le \|f(x+h)-f(x)-L(h)\|+\|L(h)\| \to 0$$

as $h \to 0$. In this, we know that $\|L(h)\| \to 0$ since all linear functions are continuous. This means that $\lim_{z \to x} f(z) = f(x)$ so f is continuous at x. ∎

In fact, we can get the slightly stronger result that if f is differentiable at x then f is uniformly continuous "near" x.

Proposition 9.3 *Suppose that f has a derivative at x. Then there are constants $K, \delta > 0$ so that for all $\|h\| < \delta$ we have that $\|f(x+h) - f(x)\| \le K\|h\|$.*

Proof: By the definition of derivative, there is some $\delta > 0$ so that if $\|h\| < \delta$ we have $\|f(x+h) - f(x) - Df(x)(h)\| \le \|h\|$. Now this gives us that $\|f(x+h) - f(x)\| \le \|Df(x)(h)\| + \|h\|$. However, since $Df(x)$ is a linear function, we know that there is a constant $M > 0$ so that $\|Df(x)(h)\| \le M\|h\|$ for all h. Thus, we take $K = M + 1$. ∎

We now want to derive the usual arithmetical rules of derivatives. Notice that we don't include a product rule, since there isn't a well-defined product in $\mathbb{R}^q$ (in fact, there might be many ways of "multiplying" the vectors).

Theorem 9.4 *Let $E \subset \mathbb{R}^p$ and $f, g : E \to \mathbb{R}^q$ be differentiable at the point x and $\alpha \in \mathbb{R}$. Then*

1. *$f + g$ is differentiable at x with $D(f + g) = Df + Dg$.*
2. *αf is differentiable at x with $D(\alpha f) = \alpha Df$.*

Proof: 1) We see that for $h \neq 0$ we have

$$\frac{\|f(x+h) + g(x+h) - f(x) - g(x) - Df(x)(h) - Dg(x)(h)\|}{\|h\|} \le$$
$$\frac{\|f(x+h) - f(x) - Df(x)(h)\|}{\|h\|} + \frac{\|g(x+h) - g(x) - Dg(x)(h)\|}{\|h\|}$$

so that

$$\lim_{h \to 0} \frac{\|f(x+h) + g(x+h) - f(x) - g(x) - Df(x)(h) - Dg(x)(h)\|}{\|h\|} \le$$

$$\lim_{h\to 0}\frac{\|f(x+h)-f(x)-Df(x)(h)\|}{\|h\|}+\lim_{h\to 0}\frac{\|g(x+h)-g(x)-Dg(x)(h)\|}{\|h\|} \quad\to\quad 0.$$

For 2), we just notice that

$$\frac{\|\alpha f(x+h)-\alpha f(x)-\alpha Df(x)(h)\|}{\|h\|}=\alpha\frac{\|f(x+h)-f(x)-Df(x)(h)\|}{\|h\|}$$

from which the result follows. ∎

Before proving the chain rule, we first prove a result which gives us a formula for the derivative in many cases. However, we need to define *partial derivatives* first. Given an $f: E\subset I\!R^p\to I\!R$, we think of $f(x)=f(x_1,x_2,\ldots,x_p)$.

Definition 9.3 *Let $f: E\subset I\!R^p\to I\!R$ and $x\in E$ be given. Then the i*th partial derivative of f *(or the* partial derivative of f with respect to x_i*) is*

$$\lim_{\epsilon\to 0}\frac{f(x_1,x_2,\ldots,x_i+\epsilon,\ldots,x_p)-f(x_1,x_2,\ldots,x_i,\ldots,x_p)}{\epsilon},$$

if it exists.

This partial derivative is denoted by

$$\frac{\partial f}{\partial x_i}(x_1,x_2,\ldots,x_p)$$

or just $\partial f/\partial x_i$.

Proposition 9.5 *Let $f: E\subset I\!R^p\to I\!R$ be differentiable at $x\in E$. Then all the partial derivatives of f exist at the point x and $\partial f/\partial x_i = Df(x)(e_i)$.*

Proof: Let $e_i=(0,0,\ldots,1,\ldots,0)$, where the 1 is in the ith position. Then we see that

$$\lim_{\epsilon\to 0}\frac{f(x+\epsilon e_i)-f(x)-\epsilon Df(x)(e_i)}{\epsilon}=0,$$

since f is differentiable at x. However,

$$f(x+\epsilon e_i)-f(x)=f(x_1,x_2,\ldots,x_i+\epsilon,x_p)-f(x_1,x_2,\ldots,x_i,\ldots,x_p)$$

and $Df(x)(e_i)\in I\!R$ (since $Df\in Lin(I\!R^p,I\!R)$). Clearly

$$\lim_{\epsilon\to 0}\frac{\epsilon Df(x)(e_i)}{\epsilon}=Df(x)(e_i)$$

and so we must have that

$$\frac{\partial f}{\partial x_i}=\lim_{\epsilon\to 0}\frac{f(x+\epsilon e_i)-f(x)}{\epsilon}$$

exists as well. Furthermore, we see that $\partial f/\partial x_i = Df(x)(e_i)$. ∎

Let $f : E \subset I\!R^p \to I\!R^q$ and $x \in E$. Then we can write $f(x)$ as

$$f(x) = (f_1(x), f_2(x), \ldots, f_q(x))$$

so that $f_i(x) = \pi_i(f(x))$ for $i = 1, 2, \ldots, q$. Furthermore, in coordinates each f_i has the form $f_i(x) = f_i(x_1, x_2, x_3, \ldots, x_p)$ where $x = (x_1, x_2, \ldots, x_p) \in I\!R^p$.

We use the previous proposition to prove the more general theorem. This theorem tells us that the matrix that represents $Df(x)$ in the standard basis is the (Jacobian) matrix

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_p} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_p} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_q}{\partial x_1} & \frac{\partial f_q}{\partial x_2} & \cdots & \frac{\partial f_q}{\partial x_p} \end{pmatrix}.$$

Theorem 9.6 *Let $f : E \subset I\!R^p \to I\!R^q$ be differentiable at $x \in E$. Then all the partial derivatives of $f_i = \pi_i \circ f$ exist at the point x and*

$$\frac{\partial f_i}{\partial x_j} = e_i \cdot Df(x)(e_j),$$

where $u \cdot v$ denotes the dot product in $I\!R^q$.

Proof: From the previous proposition we know that for each i we have that $\partial f_i/\partial x_j = Df_i(e_j)$. Thus, we need only show that this is also equal to $e_i \cdot Df(e_j)$. To see this, we note that

$$\frac{\|f(x+h) - f(x) - Df(x)(h)\|}{\|h\|} \to 0$$

so we must have that

$$\frac{\|e_i \cdot (f(x+h) - f(x) - Df(x)(h))\|}{\|h\|} \to 0$$

as well. However,

$$e_i \cdot (f(x+h) - f(x) - Df(x)(h)) = f_i(x+h) - f_i(x) - e_i \cdot Df(x)(h)$$

which means that $e_i \cdot Df(x) = Df_i(x)$. ∎

We next want to prove a converse of the previous theorem. Unfortunately, the converse isn't true. It is possible for a function to be discontinuous and have all directional derivatives existing (see the problems for an example). A directional derivative is like the partial derivative, except in an arbitrary direction.

If $f : \mathbb{R}^p \to \mathbb{R}^q$ and $u \in \mathbb{R}^p$ with $\|u\| = 1$, the *directional derivative of f at the point x in the direction u* is the limit

$$\lim_{t \to 0} \frac{f(x+tu) - f(x)}{t},$$

if this limit exists. We denote this directional derivative by $D_u f(x)$. The proof of the next simple proposition is left as an exercise.

Proposition 9.7 *Let $f : \mathbb{R}^p \to \mathbb{R}^q$ be differentiable at the point x. Then all directional derivatives exist at x and*

$$D_u f(x) = Df(x)(u).$$

Finally, we get to the result describing the structure of $Df(x)$.

Theorem 9.8 *Let $f : E \subset \mathbb{R}^p \to \mathbb{R}^q$. Suppose that x is on the interior of E and that $\partial f_i / \partial x_j$ are all continuous in some open set V containing x. Then f is differentiable at x and the matrix representing $Df(x)$ is the Jacobian matrix.*

Proof: Let $L(h) = J(x) \cdot h$, where $J(x)$ is the Jacobian matrix at x. We want to show that

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

Since the limit of a vector quantity is zero iff the limit of each component is zero, we assume that $f : E \to \mathbb{R}$, so that

$$L(h) = \sum_i h_i \frac{\partial f}{\partial x_i}.$$

Fix h for the moment and consider the $p+1$ points $y_0 = x$ and

$$y_1 = (x_1+h_1, x_2, x_3, \ldots, x_p), y_2 = (x_1+h_1, x_2+h_2, x_3, \ldots, x_p), \ldots, y_{p-1} = (x_1+h_1, x_2+h_2, x_3+h_3, \ldots, x_{p-1}+h_{p-1}, x_p)$$

and $y_p = x + h$. Now, we note that

$$f(x+h) - f(x) = (f(y_p) - f(y_{p-1})) + (f(y_{p-1}) - f(y_{p-2})) + \cdots + (f(y_1) - f(y_0)).$$

Applying the Mean Value Theorem we get a point z_i, lying on the line segment between y_{i-1} and y_i and such that

$$f(y_i) - f(y_{i-1}) = \frac{\partial f}{\partial x_i}(z_i)\ (y_i - y_{i-1}) = h_i \frac{\partial f}{\partial x_i}(z_i)$$

which gives us

$$f(x+h) - f(x) = \sum_{i=1}^{p} f(y_i) - f(y_{i-1}) = \sum_{i=1}^{p} h_i \frac{\partial f}{\partial x_i}(z_i).$$

Thus,

$$f(x+h) - f(x) - L(h) = \sum_{i=1}^{p} h_i \left(\frac{\partial f}{\partial x_i}(z_i) - \frac{\partial f}{\partial x_i}(x) \right).$$

Applying the Cauchy-Schwartz inequality to the last sum (viewed as a dot product), we get

$$\|f(x+h) - f(x) - L(h)\| \leq \|h\|\|d\|$$

(where the vector d has components $d_i = \frac{\partial f}{\partial x_i}(z_i) - \frac{\partial f}{\partial x_i}(x)$) and thus

$$\frac{\|f(x+h) - f(x) - L(h)\|}{\|h\|} \leq \|d\| \to 0$$

as $h \to 0$ since each partial derivative is continuous at x. ∎

After this brief interlude, we will now prove the chain rule. Notice what a particularly nice form this chain rule has, as opposed to the many different "chain rules" that one sees in vector calculus. It simply says that the derivative of a composition of functions is the composition of the derivatives.

Theorem 9.9 *(The Chain Rule) Suppose that $f : E \subset \mathbb{R}^p \to \mathbb{R}^q$ with x in the interior of E and $Df(x)$ exists. Suppose further that $g : f(E) \to \mathbb{R}^s$ with $f(x)$ in the interior of $f(E)$ and $Dg(f(x))$ existing. Then $D(g \circ f)(x)$ exists and*

$$D(g \circ f)(x) = Dg(f(x))Df(x).$$

Proof: Let $\epsilon > 0$ be given. First, we see that since $Df(x)$ exists, there are constants $M, \delta_1 > 0$ so that if $\|h\| < \delta_1$ we have that $\|f(x+h) - f(x)\| < M\|h\|$. Furthermore, there is a constant $K > 0$ so that $\|Dg(f(x))(k)\| \leq K\|k\|$ for all k (since $Dg(f(x))$ is a linear function).

Since $Dg(f(x))$ exists, there is some $\delta_2 > 0$ so that if $\|k\| < \delta_2$ then

$$\|g(z+k) - g(z) - Dg(f(x))(k)\| \leq \epsilon\|k\|.$$

Similarly, since $Df(x)$ exists, there is some $\delta_3 > 0$ so that if $\|h\| < \delta_3$ then

$$\|f(x+h) - f(x) - Df(x)(h)\| \leq \epsilon\|h\|$$

Let $\|h\|$ be small enough so that $\|h\| < \min\{\delta_1, \delta_2/M, \delta_3\}$. Then $\|f(x+h) - f(x)\| < \delta_2$ so that

$$\|g[f(x)+f(x+h)-f(x)]-g(f(x))-Dg(f(x))(f(x+h)-f(x))\| \leq \epsilon\|f(x+h)-f(x)\| \leq \epsilon M\|h\|.$$

Now, we also see that

$$\|Dg(f(x))[f(x+h)-f(x)-Df(x)(h)]\| \leq K\|f(x+h)-f(x)-Df(x)(h)\| \leq \epsilon K\|h\|.$$

Putting these together we see that

$$\|g[f(x+h)] - g(f(x)) - Dg(f(x))(h)\| \leq \epsilon(K+M)\|h\|$$

so that

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

as desired. ∎

It is also possible to get a version of the Mean Value Theorem. Unfortunately, the most natural generalization of the Mean Value Theorem is not true.

Example Let $f : \mathbb{R} \to \mathbb{R}^2$ be given by $f(x) = (x - x^2, x - x^3)$. Then $Df(x)(h) = ((1-2x)h, (1-3x^2)h)$. Now, $f(0) = (0,0) = f(1)$. However, there is no value of x that we can choose that makes $Df(x)(h) = (0,0)$ for all h. Thus it is impossible for $f(1) - f(0) = Df(c)(1-0)$.

However, the following "one dimensional" version of the Mean Value Theorem is easy to prove. This theorem really is nothing other than a re-statment of the Mean Value Theorem for one variable, but in this context. This is because $Df(c)(b-a) = D_{b-a}f(c)$ (the directional derivative of f at the point c in the direction of the line segment from a to b).

Theorem 9.10 *(Mean Value Theorem) Let $f : E \to \mathbb{R}$ be differentiable with $E \subset \mathbb{R}^p$. Suppose that $a, b \in E$ along with the line segment from a to b. Then there is a point c on this line segment such that*

$$f(b) - f(a) = Df(c)(b-a).$$

Proof: The line segment from a to b is given by $(1-t)a+tb$ for $t \in [0,1]$. Thus, restricted to this line we have $g(t) = f((1-t)a+tb)$ is a function $g : [0,1] \to \mathbb{R}$. It is easy to check that this function is differentiable on $(0,1)$ and continuous on $[0,1]$ so there is some point $s \in [0,1]$ so that $g(1) - g(0) = g'(s)$. However, by the Chain Rule we see that

$$g'(s) = Df(g(s))(b-a).$$

Letting $c = g(s)$ we get the desired result. ∎

There is a nice version of the Mean Value Theorem for functions with values in $\mathbb{R}^q$. However, this version yeilds an inequality rather than an equality. See the exercises for an application of this result.

Theorem 9.11 *(Mean Value Theorem) Let $f : E \to \mathbb{R}^q$ where $E \subset \mathbb{R}^p$ is open. Suppose that f is differentiable on E and that $a, b \in E$ along with the line* *segment between a and b. Then there is some point c on this line segment so that*

$$\|f(b) - f(a)\| \leq \|Df(c)(b-a)\|.$$

Proof: Define $F : E \to I\!R$ by $F(x) = (f(b) - f(a)) \cdot f(x)$. Then F is differentiable on E with $DF(x)(h) = (f(b) - f(a)) \cdot Df(x)(h)$. By the Mean Value Theorem there is some point c on the line segment between a and b so that

$$(f(b)-f(a))\cdot(f(b)-f(a)) = F(b)-F(a) = DF(c)(b-a) = (f(b)-f(a))\cdot Df(x)(b-a).$$

The result follows now by the Cauchy-Schwartz Inequality. ▮

Higher Order Derivatives and Taylor's Theorem

It is certainly possible think about taking more derivatives, just like in the case of a single variable. What would the second derivative mean? Well, we see that for $f : I\!R^p \to I\!R^q$ we have $Df : I\!R^p \to Lin(I\!R^p, I\!R^q)$. The space $Lin(I\!R^p, I\!R^q)$ can be identified with $I\!R^{pq}$, so we can think about the derivative of Df. This would give a function $D^2f : I\!R^p \to Lin(I\!R^p, Lin(I\!R^p, I\!R^q))$. We can keep repeating this as many times as we like.

An example helps to see what is going on. Take the function $f(x_1, x_2) = (x_1^2, x_2^2, x_1x_2)$ so that $f : I\!R^2 \to I\!R^3$. Then we see that Df is naturally associated with the matrix

$$Df(x) = \begin{pmatrix} 2x_1 & 0 \\ 0 & 2x_2 \\ x_2 & x_1 \end{pmatrix}.$$

What about $D^2f(x)$? How would we represent this? It is an element of $Lin(I\!R^p, Lin(I\!R^p, I\!R^q))$. The most naturaly way to represent it is as a "3D matrix" – that is, a $3 \times 2 \times 2$ array of numbers. In this case, the first "slice" would be represented by the matrix

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}$$

and the second "slice" is represented by the matrix

$$\begin{pmatrix} 0 & 0 \\ 0 & 2 \\ 1 & 0 \end{pmatrix}.$$

That is, the entries are given by

$$(D^2f(x))_{i,j,k} = \frac{\partial^2 f_i}{\partial x_j \partial x_k}.$$

The set $Lin(I\!R^p, Lin(I\!R^p, I\!R^q))$ is the set of all *bilinear* functions $T : I\!R^p \times I\!R^p \to I\!R^q$. Such functions satisfy

1. $T(\lambda x, y) = T(x, \lambda y) = \lambda T(x, y)$
2. $T(x + y, z) = T(x, z) + T(y, z)$ and $T(x, y + z) = T(x, y) + T(y, z)$

for all appropriate x, y, z and $\lambda \in I\!\!R$. Since we are not going to delve deeply into higher-order derivatives, we won't get too involved in the theory of multi-linear functions. There are many good references on the subject.

In the special case that $f : E \subset I\!\!R^p \to I\!\!R$, the situation is a little simpler as $Df(x)$ can be represented by a row vector in $I\!\!R^p$ and $D^2 f(x)$ is represented by a $p \times p$ matrix, $D^3 f(x)$ is represented by a $p \times p \times p$ matrix, and so on.

Theorem 9.12 *(Taylor's Theorem) Let $f : E \subset I\!\!R^p \to I\!\!R$ with E open. Suppose that f has continuous nth order partial derivatives in E and $a, b = a+u \in E$ along with the line segment between a and b. Then there is a point c on this line segment such that*

$$\begin{aligned} f(a+u) &= f(a) + Df(a)(u) + \frac{1}{2!} D^2 f(a)(u, u) \\ &+ \frac{1}{3!} D^3 f(a)(u, u, u) + \cdots + \frac{1}{(n-1)!} D^{n-1} f(a)(u, u, \ldots, u) \\ &+ \frac{1}{n!} D^n f(c)(u, u, \ldots, u) \end{aligned}$$

.

Proof: ▮

Local Min/Max

Taylor polynomials describe the local behaviour of a function $f : I\!\!R^p \to I\!\!R$. If $Df(x) = 0$ (x is a *critical point* of f), then the first degree Taylor polynomial is a constant so we need to look at the second degree Taylor polynomial in order to understand the local behaviour. We see that

$$f(x + h) \approx f(x) + D^2 f(x)(h, h)$$

where we can write

$$D^2 f(x)(h, h) = \begin{pmatrix} h_1 & h_2 & \cdots & h_p \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_p \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_p \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_p} & \frac{\partial^2 f}{\partial x_2 \partial x_p} & \cdots & \frac{\partial^2 f}{\partial x_p^2} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_p \end{pmatrix}.$$

Generically we know that $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$, so the above matrix is symmetric. So, what does such a function look like locally?

To understand this, we start with the simple example of the function

$$g(h_1, h_2) = \begin{pmatrix} h_1 & h_2 \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \end{pmatrix} = \lambda_1 h_1^2 + \lambda_2 h_2^2.$$

If both $\lambda_1 > 0$ and $\lambda_2 > 0$, then clearly $(0,0)$ is a local minimum. If, on the other hand, $\lambda_1 > 0$ and $\lambda_2 < 0$ or $\lambda_1 < 0$ and $\lambda_2 > 0$, then $(0,0)$ is a saddle point. Finally, if both $\lambda_1 < 0$ and $\lambda_2 < 0$, then $(0,0)$ is a local maximum. If either $\lambda_1 = 0$ or $\lambda_2 = 0$, then we can't tell. This example is simple enough. Extending it to p variables, we get

$$\begin{pmatrix} h_1 & h_2 & h_3 & \cdots & h_p \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_p \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ \vdots \\ h_p \end{pmatrix} = \lambda_1 h_1^2 + \lambda_2 h_2^2 + \lambda_3 h_3^2 + \cdots + \lambda_p h_p^2.$$

If any $\lambda_i = 0$, then the function does not depend on h_i. If all $\lambda_i > 0$, then $(0,0,0,\ldots,0)$ is a local minimum, if all $\lambda_i < 0$, then it is a local maximum. If some λ_i are positve and some are negative, then $(0,0,0,\ldots,0)$ is a saddle point.

Returning to $D^2f(x)$, we see that the matrix which represents $D^2f(x)$ is symmetric, so it can be diagonalized. This means that there is a simple linear change of variables so that when we represent $D^2f(x)$ in the new variables we get a diagonal matrix. Furthermore, the entries on the diagonal are the eigenvalues of $D^2f(x)$ (these eigenvalues do not depend on the matrix representation). All this discussion leads to the following result.

Proposition 9.13 *(Second Derivative Test) Suppose that $f : E \to I\!\!R$ (with $E \subset I\!\!R^p$) is continuously differentiable in a neighborhood of $x \in E$ and that $Df(x) = 0$. Let $\lambda_1, \lambda_2, \ldots, \lambda_p$ be the eigenvalues of $D^2f(x)$. Then*

1. *If any $\lambda_i = 0$, the second derivative test fails and we cannot use it to determine the type of the critical point x.*
2. *If $\lambda_i < 0$ for all i, then x is a local maximum.*
3. *If $\lambda_i > 0$ for all i, then x is a local minumum.*
4. *If $\lambda_i > 0$ and $\lambda_j < 0$ for some i and j, then x is a saddle point.*

Mapping Theorems

The derivative Df gives a local linear approximation to the function f. In many cases, this information is enough to show quite strong properties of the function f. The following theorems are examples of this. We say that $f : E \to I\!\!R^q$ is continuously differentiable in the case where $Df : E \to Lin(I\!\!R^p, I\!\!R^q)$ is a continuous function.

Theorem 9.14 *(Injective Mapping Theorem) Suppose that $f : E \to I\!\!R^q$ is continuously differentiable on the open subset $EI\!\!R^p$. Suppose further that $x_0 \in E$ with $Df(x_0)$ is an injective (linear) map. Then there is a number $\delta > 0$ so the restriction of f to $B_\delta(x_0)$ is an injection. Furthermore, the inverse of this restriction is a continuous function on $f(B_\delta(x_0))$.*

Proof: ∎

Theorem 9.15 *(Surjective Mapping Theorem) Suppose that $f : E \to I\!R^q$ is continuously differentiable on the open subset $E \subset I\!R^p$. Suppose further that $x_0 \in E$ with $Df(x_0)$ a surjective (linear) map. Then there exist $\epsilon, \delta > 0$ so that for each $y \in I\!R^q$ with $\|y - f(x_0)\| < \epsilon$ there is an $x \in'$ with $\|x - x_0\| < \delta$ so that $f(x) = y$.*

Proof: ∎

The following theorem is obviously a combination of the previous two theorems. The main new feature is the formula for the derivative of the inverse. It is as nice as one could hope for in that $Df^{-1}(y) = (Df(x))^{-1}$ where $f(x) = y$.

Theorem 9.16 *(Inverse Mapping Theorem) Suppose that $f : E \to I\!R^p$ is continuously differentiable on the open subset $E \subset I\!R^p$. Suppose further that $x_0 \in E$ with $Df(x_0)$ a bijection. Then there exists an open neighborhood U of x_0 and an open neighborhood V of $y_0 = f(x_0)$ such that $V = f(U)$ and the restriction of f to U is a bijection. Furthermore, f^{-1} on V is differentiable with*

$$Df^{-1}(y_0) = [Df(x)]^{-1}.$$

Proof: ∎

The next theorem (the Implicit Function Theorem) is a very important theoretical tool. It gives some sufficient conditions in which one can "solve" an equation for some variables in terms of the other variables. For example, suppose we wish to solve the equations

$$x^2 + y^2 + z^2 - 1 = 0 \quad x^3y^3 + x^3z^3 + y^3z^3 = 0$$

for z in terms of x and y. How do we know that we can do that?

In general, there might be more than one solution, as in solving $x^2 + y^2 = 1$ for y in terms of x. However, "locally" we think that we can write y as a function of x. In fact, in this case clearly $y = \sqrt{1 - x^2}$, which works except at the points $x = \pm 1$. Compare this to the hypothesis of the theorem (especially the hypothesis that the linear map is a bijection − in this simple example this linear map fails to be a bijection precisely at the points $x = \pm 1$).

The first conclusion of the theorem tells us that we can find a local parameterization (solution) while the second conclusion tells us that this local solution is unique.

Theorem 9.17 *(Implicit Function Theorem) Let $E \subset \mathbb{R}^p \times \mathbb{R}^q$ be open and $f : E \to \mathbb{R}^q$ be continuously differentiable on E. Suppose that $(a, b) \in E$ with $f(a, b) = 0$ and that the linear map*

$$L(v) = Df(a, b)(0, v)$$

is a bijection.

- *There exists an open neighborhood U of $a \in \mathbb{R}^p$ and a continuously differentiable function $\phi : U \to \mathbb{R}^q$ such that $\phi(a) = b$ and $f(x, \phi(x)) = 0$ for all $x \in U$.*
- *There exists an open neighborhood V of (a, b) in $\mathbb{R}^p \times \mathbb{R}^q$ such that $(x, y) \in V$ with $f(x, y) = 0$ iff $y = \phi(x)$.*

Proof: ∎

Problems

1. Let $U \subset \mathbb{R}^2$ and $f : U \to \mathbb{R}^3$ and $M : U \to Lin(\mathbb{R}^3, \mathbb{R}^3)$ both be differentiable over U. Derive a "product rule" for the derivative of the function $g : U \to \mathbb{R}^q$ given by
$$g(x) = M(x) \cdot f(x).$$
2. Let $U \subset \mathbb{R}^2$ and $f : U \to Lin(\mathbb{R}^2, \mathbb{R}^2)$ be differentiable over U. Derive a formula for the derivative of $(f(x))^{-1}$.
3. Suppose that $f : E \to \mathbb{R}$ is differentiable with $E \subset \mathbb{R}^p$. Prove that for any relative extrema $x \in E$ we have that $Df(x) = 0$ (the zero linear function).
4. Let $g(x)$ be any function that satisfies $g(x) > 0$ for $x > 0$ and $g(x) = 0$ for $x \leq 0$. Define
$$f(x, y) = \begin{cases} g(x/y^2 - 1)g(2 - x/y^2), & \text{if } y \neq 0 \\ 0, & \text{if } y = 0 \end{cases}.$$
Show that for any $u = (u_1, u_2) \in \mathbb{R}^2$ with $\|u\| = 1$ we have that $D_u f(0, 0) = 0$. Show that f is not continuous at 0 (the limit $\lim_{(x,y)\to(0,0)} f(x, y)$ doesn't exist).
5. Find any example of a continuous function $f : \mathbb{R}^2 \to \mathbb{R}$ so that $f(x, y)$ is not differentiable at $(0, 0)$ but
$$\frac{\partial f}{\partial x}(0, 0) \quad \text{and} \quad \frac{\partial f}{\partial y}(0, 0)$$
both exist.

6. Prove Proposition 9.7.

7. Suppose that $f : E \to \mathbb{R}^q$ where E is an open set and f is differentiable on E. Suppose further that $Df(x) = 0$ for all $x \in E$. Prove that f is constant on E.

8. Compute the nth order Taylor expansion for $f(x, y) = (e^{x+y}, \sin(x + y))$.

9. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be differentiable. Derive a version of Newton's method for f. Use this version of Newton's method to find a zero of the function $f(x, y) = (1-x^2-y^2, x^2/2+2y^2-1)$ starting at the initial point $(x_0, y_0) = (1, 1)$ (say to 20 decimal places of accuracy).

10. Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable with $f(x^*) = x^*$. Show that if $\|Df(x^*)\| = s < 1$ then there is some $\epsilon > 0$ so that for any $x \in B_\epsilon(x^*)$ we have that
$$\|f(x) - x^*\| \le \left(\frac{1+s}{2}\right) \|x - x^*\|.$$
Thus, if we start with $x_0 \in B_\epsilon(x^*)$, the iteration $x_{n+1} = f(x_n)$ will converge to x^*.

11. Suppose that $f : \mathbb{R} \to \mathbb{R}$ is continuously differentiable with $f(0) = 0$ but $f'(0) \neq 0$. Show that there is some $\delta > 0$ so that $f(x) = 0$ for $|x| < \delta$ implies $x = 0$.

12. We say that $x \in S \subset \mathbb{R}^q$ is on the boundary of S if for every $\epsilon > 0$ there is an $t \notin S$ with $\|x - t\| < \epsilon$ (that is, $B_\epsilon(x) \cap (\mathbb{R}^q \setminus S) \neq \emptyset$). Let $f : E \to \mathbb{R}^q$ be continuously differentiable with $E \subset \mathbb{R}^p$ an open set. Show that $f(x)$ is a boundary point of $f(E)$ only if $rank(Df(x)) < q$.